EVIDENCE-BASED PRACTICE Step by Step

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

Critical Appraisal of the Evidence: Part II

Digging deeper—examining the "keeper" studies.

This is the sixth 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 November's *Evidence-Based Practice, Step by Step*.

In July's evidence-based practice (EBP) article, Rebecca R., our hypothetical staff nurse, Carlos A., her hospital's expert EBP mentor, and Chen M., Rebecca's nurse colleague, collected the evidence to answer their clinical question: "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)?" As part of their rapid critical appraisal (RCA) of the 15 potential "keeper" studies, the EBP team found and placed the essential elements of each study (such as its population, study design, and setting) into an evaluation table. In so doing, they began to see similarities and differences between the studies, which Carlos told them is the beginning of synthesis. We now join the team as they continue with their RCA of these studies to determine their worth to practice.

RAPID CRITICAL APPRAISAL

Carlos explains that typically an RCA is conducted along with an RCA checklist that's specific to the research design of the study being evaluated—and before any data are entered into an evaluation table. However, since Rebecca and Chen are new to appraising studies, he felt it would be easier for them to first enter the essentials into the table and then evaluate each study. Carlos shows Rebecca several RCA checklists and explains that all checklists have three major questions in common, each of which contains other more specific subquestions about what constitutes a well-conducted study for the research design under review (see *Example of a Rapid Critical Appraisal Checklist*).

Although the EBP team will be looking at how well the researchers conducted their studies and discussing what makes a "good" research study, Carlos reminds them that the goal of critical appraisal is to determine the worth of a study to practice, not solely to find flaws. He also suggests that they consult their glossary when they see an unfamiliar word. For example, the term *randomization*, or *random assignment*, is a relevant feature of research methodology for intervention studies that may be unfamiliar. Using the glossary, he explains that random assignment and *random sampling* are often confused with one another, but that they're very different. When researchers select subjects from within a certain population to participate in a study by using a random strategy, such as tossing a coin, this is random sampling. It allows the entire population to be fairly represented. But because it requires access to a particular population, random sampling is not always feasible. Carlos adds that many health care studies are based on a *convenience sample*—participants recruited from a readily available population, such as a researcher's affiliated hospital, which may or may not represent the desired population. Random assignment, on the other hand, is the use of a random strategy to assign study

participants to the intervention or control group. Random assignment is an important feature of higher-level studies in the hierarchy of evidence.

Carlos also reminds the team that it's important to begin the RCA with the studies at the highest level of evidence in order to see the most reliable evidence first. In their pile of studies, these are the three systematic reviews, including the meta-analysis and the Cochrane review, they retrieved from their database search (see "Searching for the Evidence," and "Critical Appraisal of the Evidence: Part I," *Evidence-Based Practice, Step by Step*, May and July). Among the RCA checklists Carlos has brought with him, Rebecca and Chen find the checklist for systematic reviews.

As they start to rapidly critically appraise the meta-analysis, they discuss that it seems to be biased since the authors included only studies with a control group. Carlos explains that while having a control group in a study is ideal, in the real world most studies are lower-level evidence and don't have control or comparison groups. He emphasizes that, in eliminating lower-level studies, the meta-analysis lacks evidence that may be informative to the question. Rebecca and Chen—who are clearly growing in their appraisal skills—also realize that three studies in the meta-analysis are the same as three of their potential "keeper" studies. They wonder whether they should keep those studies in the pile, or if, as duplicates, they're unnecessary. Carlos says that because the meta-analysis only included studies with control groups, it's important to keep these three studies so that they can be compared with other studies in the pile that don't have control groups. Rebecca notes that more than half of their 15 studies don't have control or comparison groups. They agree as a team to include all 15 studies at all levels of evidence and go on to appraise the two remaining systematic reviews.

The MERIT trial[1] is next in the EBP team's stack of studies.

Example of a Rapid Critical Appraisal Checklist

Rapid Critical Appraisal of Systematic Reviews of Clinical Interventions or Treatments

1. Are the results of the review valid?		
A. Are the studies in the review randomized controlled trials?	Yes	No
B. Does the review include a detailed description of the search strategy used to find the relevant studies?	Yes	No
C. Does the review describe how the validity of the individual studies was assessed (such as, methodological quality, including the use of random assignment to study groups and complete follow-up of subjects)?	Yes	No
D. Are the results consistent across studies?	Yes	No
E. Did the analysis use individual patient data or aggregate data?	Patient	Aggregate
2. What are the results?		
A. How large is the intervention or treatment effect (odds ratio, relative risk, effect size, level of significance)?		
B. How precise is the intervention or treatment (confidence interval)?		
3. Will the results assist me in caring for my patients?		
A. Are my patients similar to those in the review?	Yes	No
B. Is it feasible to implement the findings in my practice setting?	Yes	No
C. Were all clinically important outcomes considered, including both risks and benefits of the treatment?	Yes	No
D. What is my clinical assessment of the patient, and are there any contraindications or circumstances that would keep me from implementing the treatment?	Yes	No
E. What are my patients' and their families' preferences and values concerning the treatment?	Yes	No

As we noted in the last installment of this series, MERIT is a good study to use to illustrate the different steps of the critical appraisal process. (Readers may want to retrieve the article, if possible, and follow along with the RCA.) Set in Australia, the MERIT trial examined whether the introduction of a rapid response team (RRT; called a medical emergency team or MET in the study) would reduce the incidence of cardiac arrest, death, and unplanned admissions to the ICU in the hospitals studied. To follow along as the EBP team addresses each of the essential elements of a well-conducted randomized controlled trial (RCT) and how they apply to the MERIT study, see their notes in *Rapid Critical Appraisal of the MERIT Study*.

ARE THE RESULTS OF THE STUDY VALID?

The first section of every RCA checklist addresses the validity of the study at hand—did the researchers use sound scientific methods to obtain their study results? Rebecca asks why validity is so important. Carlos replies that if the study's conclusion can be trusted—that is, relied upon to inform practice—the study must be conducted in a way that reduces bias or eliminates confounding variables (factors that influence how the intervention affects the outcome). Researchers typically use rigorous research methods to reduce the risk of bias. The purpose of the RCA checklist is to help the user determine whether or not rigorous methods have been used in the study under review, with most questions offering the option of a quick answer of "yes," "no," or "unknown."

Were the subjects randomly assigned to the intervention and control groups? Carlos explains that this is an important question when appraising RCTs. If a study calls itself an RCT but didn't randomly assign participants, then bias could be present. In appraising the MERIT study, the team discusses how the researchers randomly assigned entire hospitals, not individual patients, to the RRT intervention and control groups using a technique called *cluster randomization*. To better understand this method, the EBP team looks it up on the Internet and finds a PowerPoint presentation by a World Health Organization researcher that explains it in simplified terms: "Cluster randomized trials are experiments in which social units or clusters [in our case, hospitals] rather than individuals are randomly allocated to intervention groups."[2]

Was random assignment concealed from the individuals enrolling the subjects? Concealment helps researchers reduce potential bias, preventing the person(s) enrolling participants from recruiting them into a study with enthusiasm if they're destined for the intervention group or with obvious indifference if they're intended for the control or comparison group. The EBP team sees that the MERIT trial used an independent statistician to conduct the random assignment after participants had already been enrolled in the study, which Carlos says meets the criteria for concealment.

Were the subjects and providers blind to the study group? Carlos notes that it would be difficult to blind participants or researchers to the intervention group in the MERIT study because the hospitals that were to initiate an RRT had to know it was happening. Rebecca and Chen wonder whether their "no" answer to this question makes the study findings invalid. Carlos says that a single "no" may or may not mean that the study findings are invalid. It's their job as clinicians interpreting the data to weigh each aspect of the study design. Therefore, if the answer to any validity question isn't affirmative, they must each ask themselves: does this "no" make the study findings untrustworthy to the extent that I don't feel comfortable using them in my practice?

Were reasons given to explain why subjects didn't complete the study? Carlos explains that sometimes participants leave a study before the end (something about the study or the participants themselves may prompt them to leave). If all or many of the participants leave for the same reason, this may lead to biased findings. Therefore, it's important to look for an explanation for why any subjects didn't complete a study. Since no hospitals dropped out of the MERIT study, this question is determined to be not applicable.

Were the follow-up assessments long enough to fully study the effects of the intervention? Chen asks Carlos why a time frame would be important in studying validity. He explains that researchers must ensure that the outcome is evaluated for a long enough period of time to show that the intervention indeed caused it. The researchers in the MERIT study conducted the RRT intervention for six months before evaluating the outcomes. The team discusses how six months was likely adequate to determine how the RRT affected cardiopulmonary arrest rates (CR) but might have been too short to establish the relationship between the RRT and hospital-wide mortality rates (HMR).

Rapid Critical Appraisal of the MERIT Study

1. Are the results of the study valid?

A. Were the subjects randomly assigned to the intervention and control groups? **Yes** No Unknown

Random assignment of hospitals was made to either a rapid response team (RRT; intervention) group or no RRT (control) group. To protect against introducing further bias into the study, hospitals, not individual patients, were randomly assigned to the intervention. If patients were the study subjects, word of the RRT might have gotten around, potentially influencing the outcome.

B. Was random assignment concealed from the individuals enrolling the subjects? **Yes** No Unknown

An independent statistician randomly assigned hospitals to the RRT or no RRT group after baseline data had been collected; thus the assignments were concealed from both researchers and participants.

C. Were the subjects and providers blind to the study group? Yes **No** Unknown

Hospitals knew to which group they'd been assigned, as the intervention hospitals had to put the RRTs into practice. Management, ethics review boards, and code committees in both hospitals knew about the intervention. The control hospitals had code teams and some already had systems in place to manage unstable patients. But control hospitals didn't have a placebo strategy to match the intervention hospitals' educational strategy for how to implement an RRT (a red flag for confounding!). If you worked in one of the control hospitals, unless you were a member of one of the groups that gave approval, you wouldn't have known your hospital was participating in a study on RRTs; this lessens the chance of confounding variables influencing the outcomes.

D. Were reasons given to explain why subjects didn't complete the study? Yes No **Not Applicable**

This question is not applicable as no hospitals dropped out of the study.

E. Were the follow-up assessments long enough to fully study the effects of the intervention? **Yes** No Unknown

The intervention was conducted for six months, which should be adequate time to have an impact on the outcomes of cardiopulmonary arrest rates (CR), hospital-wide mortality rates (HMR), and unplanned ICU admissions (UICUA). However, the authors remark that it can take longer for an RRT to affect mortality, and cite trauma protocols that took up to 10 years.

F. Were the subjects analyzed in the group to which they were randomly assigned? **Yes** No Unknown

All 23 (12 intervention and 11 control) hospitals remained in their groups, and analysis was conducted on an intention-to-treat basis. However, in their discussion, the authors attempt to provide a reason for the disappointing study results; they suggest that because the intervention group was "inadequately implemented," the fidelity of the intervention was compromised, leading to less than reliable results. Another possible explanation involves the baseline quality of care; if high, the improvement after an RRT may have been less than remarkable. The authors also note a historical confounder: in Australia, where the study took place, there was a nationwide increase in awareness of patient safety issues.

G. Was the control group appropriate? Yes No **Unknown**

See notes to question C. Controls had no time built in for education and training as the intervention hospitals did, so this time wasn't controlled for, nor was there any known attempt to control the organizational "buzz" that something was going on. The study also didn't account for the variance in how RRTs were implemented across hospitals. The researchers indicate that the existing code teams in control hospitals "did operate as [RRTs] to some extent." Because of these factors, the appropriateness of the control group is questionable.

H. Were the instruments used to measure the outcomes valid and reliable? **Yes** No Unknown

The primary outcome was the composite of HMR (that is, unexpected deaths, excluding do not resuscitates [DNRs]), CR (that is, no palpable pulse, excluding DNRs), and UICUA (any unscheduled admissions to the ICU).

I. Were the demographics and baseline clinical variables of the subjects in each of the groups similar? Yes No **Unknown**

The researchers provided a table showing how the RRT and control hospitals compared on several variables. Some variability existed, but there were no statistical differences between groups.

2. What are the results?

A. How large is the intervention or treatment effect?

The researchers reported outcome data in various ways, but the bottom line is that the control group did better than the intervention group. For example, RRT calling criteria were documented more than 15 minutes before an event by more hospitals in the control group than in the intervention group, which is contrary to expectation. Half the HMR cases in the intervention group met the criteria compared with 55% in the control group (not statistically significant). But only 30% of CR cases in the intervention group met the criteria compared with 44% in the control group, which was statistically significant ($P = 0.031$). Finally, regarding UICUA, 51% in the intervention group compared with 55% in the control group met the criteria (not significant). This indicates that the control hospitals were doing a better job of documenting unstable patients before events occurred than the intervention hospitals.

B. How precise is the intervention or treatment?

The odds ratio (OR) for each of the outcomes was close to 1.0, which indicates that the RRT had no effect in the intervention hospitals compared with the control hospitals. Each confidence interval (CI) also included the number 1.0, which indicates that each OR wasn't statistically significant (HMR OR = 1.03 (0.84 – 1.28); CR OR = 0.94 (0.79 – 1.13); UICUA OR = 1.04 (0.89 – 1.21). From a clinical point of view, the results aren't straightforward. It would have been much simpler had the intervention hospitals and the control hospitals done equally badly; but the fact that the control hospitals did better than the intervention hospitals raises many questions about the results.

3. Will the results help me in caring for my patients?

A. Were all clinically important outcomes measured? Yes **No** Unknown

It would have been helpful to measure cost, since participating hospitals that initiated an RRT didn't eliminate their code team. If a hospital has two teams, is the cost doubled? And what's the return on investment? There's also no mention of the benefits of the code team. This is a curious question . . . maybe another PICOT question?

B. What are the risks and benefits of the treatment?

This is the wrong question for an RRT. The appropriate question would be: What is the risk of not adequately introducing, monitoring, and evaluating the impact of an RRT?

C. Is the treatment feasible in my clinical setting? **Yes** No Unknown

We have administrative support, once we know what the evidence tells us. Based on this study, we don't know much more than we did before, except to be careful about how we approach and evaluate the issue. We need to keep the following issues, which the MERIT researchers raised in their discussion, in mind: 1) allow adequate time to measure outcomes; 2) some outcomes may be reliably measured sooner than others; 3) the process of implementing an RRT is very important to its success.

D. What are my patients' and their families' values and expectations for the outcome and the treatment itself?

We will keep this in mind as we consider the body of evidence.

Were the subjects analyzed in the group to which they were randomly assigned? Rebecca sees the term *intention-to-treat analysis* in the study and says that it sounds like statistical language. Carlos confirms that it is; it means that the researchers kept the hospitals in their assigned groups when they conducted the analysis, a technique intended to reduce possible bias. Even though the MERIT study used this technique, Carlos notes that in the discussion section the authors offer some important caveats about how the study was conducted, including poor intervention implementation, which may have contributed to MERIT's unexpected findings.[1]

Was the control group appropriate? Carlos explains that it's challenging to establish an appropriate comparison or control group without an understanding of how the intervention will be implemented. In this case, it may be problematic that the intervention group received education and training in implementing the RRT and the control group received no comparable placebo (meaning education and training about something else). But Carlos reminds the team that the researchers attempted to control for known confounding variables by stratifying the sample on characteristics such as academic versus nonacademic hospitals, bed size, and other important parameters. This method helps to ensure equal representation of these parameters in both the intervention and control groups. However, a major concern for clinicians considering whether to use the MERIT findings in their decision making involves the control hospitals' code teams and how they may have functioned as RRTs, which introduces a potential confounder into the study that could possibly invalidate the findings.

Were the instruments used to measure the outcomes valid and reliable? The overall measure in the MERIT study is the composite of the individual outcomes: CR, HMR, and unplanned admissions to the ICU (UICUA). These parameters were defined reasonably and didn't include do not resuscitate (DNR) cases. Carlos explains that since DNR cases are more likely to code or die, including them in the HMR and CR would artificially increase these outcomes and introduce bias into the findings.

As the team moves through the questions in the RCA checklist, Rebecca wonders how she and Chen would manage this kind of appraisal on their own. Carlos assures them that they'll get better at recognizing well-conducted research the more RCAs they do. Though Rebecca feels less than confident, she appreciates his encouragement nonetheless, and chooses to lead the team in discussion of the next question.

Were the demographics and baseline clinical variables of the subjects in each of the groups similar? Rebecca says that the intervention group and the control or comparison group need to be similar at the beginning of any intervention study because any differences in the groups could influence the outcome, potentially increasing the risk that the outcome might be unrelated to the intervention. She refers the team to their earlier discussion about confounding variables. Carlos tells Rebecca that her explanation was excellent. Chen remarks that Rebecca's focus on learning appears to be paying off.

WHAT ARE THE RESULTS?

As the team moves on to the second major question, Carlos tells them that many clinicians are apprehensive about interpreting statistics. He says that he didn't take courses in graduate school on conducting statistical analysis; rather, he learned about different statistical tests in courses that required students to look up how to interpret a statistic whenever they encountered it in the articles they were reading. Thus he had a context for how the statistic was being used and interpreted, what question the statistical analysis was answering, and what kind of data were being analyzed. He also learned to use a search engine, such as Google.com, to find an explanation for any statistical tests with which he was unfamiliar. Because his goal was to understand what the statistic meant clinically, he looked for simple Web sites with that same focus and avoided those with Greek symbols or extensive formulas that were mostly concerned with conducting statistical analysis.

How large is the intervention or treatment effect? As the team goes through the studies in their RCA, they decide to construct a list of statistics terminology for quick reference (see *A Sampling of Statistics*). The major statistic used in the MERIT study is the odds ratio (OR). The OR is used to provide insight into the measure of association between an intervention and an outcome. In the MERIT study, the control group did better than the intervention group, which is contrary to what was expected. Rebecca notes that the researchers discussed the possible reasons for this finding in the final section of the study. Carlos says that the authors' discussion about why their findings occurred is as important as the findings themselves. In this study, the discussion communicates to any clinicians considering initiating an RRT in their hospital that they should assess whether the current code team is already functioning

A Sampling of Statistics

Statistic	Simple Definition	Important Parameters	Understanding the Statistic	Clinical Implications
Odds Ratio (OR)	The odds of an outcome occurring in the intervention group compared with the odds of it occurring in the comparison or control group.	• If an OR is equal to 1, then the intervention didn't make a difference. • Interpretation depends on the outcome. • If the outcome is good (for example, fall prevention), the OR is preferred to be above 1. • If the outcome is bad (for example, mortality rate), the OR is preferred to be below 1.	The OR for hospital-wide mortality rates (HMR) in the MERIT study was 1.03 (95% CI, 0.84 – 1.28). The odds of HMR in the intervention group were about the same as HMR in the comparison group.	From the HMR OR data alone, a clinician may not feel confident that a rapid response team (RRT) is the best intervention to reduce HMR but may seek out other evidence before making a decision.
Relative Risk (RR)	The risk of an outcome occurring in the intervention group compared with the risk of it occurring in the comparison or control group.	• If an RR is equal to 1, then the intervention didn't make a difference. • Interpretation depends on the outcome. • If the outcome is good (for example fall prevention), the RR is preferred to be above 1. • If the outcome is bad (for example, mortality rate), the RR is preferred to be below 1.	The RR of cardiopulmonary arrest in adults was reported in the Chan PS, et al., 2010 systematic review[a] as 0.66 (95% CI, 0.54 – 0.80), which is statistically significant because there's no 1.0 in the CI. Thus, the RR of cardiopulmonary arrest occurring in the intervention group compared with the RR of it occurring in the control group is 0.66, or less than 1. Since cardiopulmonary arrest is not a good outcome, this is a desirable finding.	The RRT significantly reduced the RR of cardiopulmonary arrest in this study. From these data, clinicians can be reasonably confident that initiating an RRT will reduce CR in hospitalized adults.
Confidence Interval (CI)	The range in which clinicians can expect to get results if they present the intervention as it was in the study.	• CI provides the precision of the study finding: a 95% CI indicates that clinicians can be 95% confident that their findings will be within the range given in the study. • CI should be narrow around the study finding, not wide. • If a CI contains the number that indicates no effect (for OR it's 1; for effect size it's 0), the study finding is not statistically significant.	See the two previous examples.	In the Chan PS, et al., 2010 systematic review,[a] the CI is a close range around the study finding and is statistically significant. Clinicians can be 95% confident that if they conduct the same intervention, they'll have a result similar to that of the study (that is, a reduction in risk of cardiopulmonary arrest) within the range of the CI, 0.54 – 0.80. The narrower the CI range, the more confident clinicians can be that, using the same intervention, their results will be close to the study findings.
Mean (X)	Average	• Caveat: Averaging captures only those subjects who surround a central tendency, missing those who may be unique. For example, the mean (average) hair color in a classroom of schoolchildren captures those with the predominant hair color. Children with hair color different from the predominant hair color aren't captured and are considered outliers (those who don't converge around the mean).	In the Dacey MJ, et al., 2007 study,[a] before the RRT the average (mean) CR was 7.6 per 1,000 discharges per month; after the RRT, it decreased to 3 per 1,000 discharges per month.	Introducing an RRT decreased the average CR by more than 50% (7.6 to 3 per 1,000 discharges per month).

[a] For study details on Chan PS, et al., and Dacey MJ, et al., go to http://links.lww.com/AJN/A11.

as an RRT prior to RRT implementation.

How precise is the intervention or treatment? Chen wants to tackle the precision of the findings and starts with the OR for HMR, CR, and UICUA, each of which has a confidence interval (CI) that includes the number 1.0. In an EBP workshop, she learned that a 1.0 in a CI for OR means that the results aren't statistically significant, but she isn't sure what statistically significant means. Carlos explains that since the CIs for the OR of each of the three outcomes contains the number 1.0, these results could have been obtained by chance and therefore aren't statistically significant. For clinicians, chance findings aren't reliable findings, so they can't confidently be put into practice. Study findings that aren't statistically significant have a probability value (*P* value) of greater than 0.5. Statistically significant findings are those that aren't likely to be obtained by chance and have a *P* value of less than 0.5.

WILL THE RESULTS HELP ME IN CARING FOR MY PATIENTS?

The team is nearly finished with their checklist for RCTs. The third and last major question addresses the applicability of the study—how the findings can be used to help the patients the team cares for. Rebecca observes that it's easy to get caught up in the details of the research methods and findings and to forget about how they apply to real patients.

Were all clinically important outcomes measured? Chen says that she didn't see anything in the study about how much an RRT costs to initiate and how to compare that cost with the cost of one code or ICU admission. Carlos agrees that providing costs would have lent further insight into the results.

What are the risks and benefits of the treatment? Chen wonders how to answer this since the findings seem to be confounded by the fact that the control hospital had code teams that functioned as RRTs. She wonders if there was any consideration of the risks and benefits of initiating an RRT prior to beginning the study. Carlos says that the study doesn't directly mention it, but the consideration of the risks and benefits of an RRT is most likely what prompted the researchers to conduct the study. It's helpful to remember, he tells the team, that often the answer to these questions is more than just "yes" or "no."

Is the treatment feasible in my clinical setting? Carlos acknowledges that because the nursing administration is open to their project and supports it by providing time for the team to conduct its work, an RRT seems feasible in their clinical setting. The team discusses that nursing can't be the sole discipline involved in the project. They must consider how to include other disciplines as part of their next step (that is, the implementation plan). The team considers the feasibility of getting all disciplines on board and how to address several issues raised by the researchers in the discussion section (see *Rapid Critical Appraisal of the MERIT Study*), particularly if they find that the body of evidence indicates that an RRT does indeed reduce their chosen outcomes of CR, HMR, and UICUA.

What are my patients' and their families' values and expectations for the outcome and the treatment itself? Carlos asks Rebecca and Chen to discuss with their patients and their patients' families their opinion of an RRT and if they have any objections to the intervention. If there are objections, the patients or families will be asked to reveal them.

The EBP team finally completes the RCA checklists for the 15 studies and finds them all to be "keepers." There are some studies in which the findings are less than reliable; in the case of MERIT, the team decides to include it anyway because it's considered a landmark study. All the studies they've retained have something to add to their understanding of the impact of an RRT on CR, HMR, and UICUA. Carlos says that now that they've determined the 15 studies to be somewhat valid and reliable, they can add the rest of the data to the evaluation table.

Be sure to join the EBP team for "Critical Appraisal of the Evidence: Part III" in the next installment in the series, when Rebecca, Chen, and Carlos complete their synthesis of the 15 studies and determine what the body of evidence says about implementing an RRT in an acute care setting. ▼

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.

REFERENCES

1. Hillman K, et al. Introduction of the medical emergency team (MET) system: a cluster-randomised controlled trial. *Lancet* 2005;365, 2091-7.
2. Wojdyla D. Cluster randomized trials and equivalence trials [PowerPoint presentation]. Geneva, Switzerland: Geneva Foundation for Medical Education and Research; 2005. http://www.gfmer.ch/PGC_RH_2005/pdf/Cluster_Randomized_Trials.pdf.