# Essential Skills for Evidence-based Practice: Appraising Evidence for Therapy Questions

**Jeanne Grace, RN, PhD**[1]

**Abstract**

Evidence to support the effectiveness of therapies commonly compares the outcomes between a group of individuals who received the therapy and a group of individuals who did not. Nurses must be able to determine how well the study design supports claims that the therapy caused the difference in group outcome (validity). When findings are valid, nurses must also consider whether those findings apply to the patient population of interest (generalizability) and, if so, how the findings can be applied to practice.

**Keywords:** critical appraisal, evidence-based practice, therapy, validity

*J Nurs Sci. 2012;30(1):10-15*

*Corresponding Author: Jeanne Grace, Box SON, 601 Elmwood Avenue, Rochester, NY, USA 14642;*
*e-mail: Jeanne_Grace@urmc.rochester.edu*
[1] *Emeritus Clinical Professor of Nursing, University of Rochester, Rochester, New York, USA*

The purpose of any therapy is to change some outcome for the patient who receives it. Nurses use many different therapies to achieve a variety of health care goals. We give medications to relieve pain. We modify the patient's physical environment to avoid falls. We apply wound dressings to promote healing and prevent infection. We provide anticipatory guidance to empower patients in their own care.

Because therapies are such a large part of nursing practice, nurses need to be able to appraise the evidence for therapy questions: does this action achieve the desired outcome? Is it more effective or efficient than some other action for achieving that outcome? What kinds of patients are likely to benefit from this action?

When collected and synthesized research studies are available, nurses should access these sources to guide their choices of effective therapies[1]. In some situations, however, we must critically appraise reports of single research studies for application to practice. There are many strategies for accomplishing this appraisal. The evidence-based practice approach[2] is particularly attractive because it focuses on study findings and the practical design issues that are most likely to introduce bias (allow influences other than the therapy being studied to affect the outcomes). The critical appraisal of therapy evidence is guided by a short set of questions about the validity of the results, the nature of the results, and the ways the results can be applied to patient care. Outlines of the critical appraisal questions for all types of evidence are organized into these three topics, although the specific questions asked depend on the type of evidence being appraised (e.g. therapy, harm, diagnosis, prognosis).

## Are the results valid?

When we say that study results are valid, we are saying we can trust that the findings are really answers to the questions being asked. Therapy questions are questions about cause and effect. Evidence for cause and effect must meet a higher set of standards than a simple demonstration of relationship:

- A change in the cause is associated with a change in the effect (covariation). For example, persons who start a new exercise program (change in cause) report more energy for daily activities (change in effect).
- The proposed cause must occur before the proposed effect (temporal precedence). For example, we cannot claim that a certain medication (cause) cured a skin rash when the rash disappeared (effect) before the patient started using the medication.
- Competing explanations for the effect other than the proposed cause must be ruled out (non-spuriousness). For example, patients receiving hospice care are more likely to die than patients receiving obstetric care. Hospice care, however, does not cause the deaths. Instead, the difference in outcomes can be explained by the difference in diagnoses between the two groups. The terminal illness of hospice patients is a competing explanation for their deaths that cannot be ruled out.

For therapy evidence, the strongest research designs to meet these standards are clinical trials where two or more groups are formed and the researcher has control over the change in the cause / therapy for each group. Comparison of differences (changes in the effect) on the outcomes of the groups after they experience that change in cause demonstrates that the cause and effect covary and the cause has temporal precedence over the effect. The crucial question, to demonstrate non-spuriousness, is whether the therapy and comparison groups were different from each other in some way other than their exposure to the therapy at any time during the study. Any group difference becomes a competing explanation of the outcome.

### Were the groups being compared similar to each other at the start of the study?

***Were individuals randomly assigned to the groups?***

The first question to answer about the validity of therapy study findings is whether the groups being compared were really similar to each other at the beginning of the study. The strongest evidence to support decisions about using therapies comes from randomized clinical trials (RCTs). In RCTs, good and bad outcomes for patients who received the therapy are compared to the same outcomes for patients who did not receive the therapy (or received a different therapy). This research design is particularly desirable because patients are assigned to the treatment groups by chance (random assignment) and the researcher controls the treatment that the patients receive

(manipulation of the independent variable). These design characteristics support the argument that the groups being compared were really similar to each other before treatment, and that it was the treatment, not some other influence, that caused any difference in outcomes.

***Was random assignment really random?***

Random assignment increases the likelihood that, as a whole, the groups being compared match each other on both recognized and unrecognized factors that could influence their outcomes. Some random assignment strategies, however, are more open to outside influence than others. If investigators already know the order of random assignment when they recruit subjects for the study, it is possible that knowledge could influence who is recruited or in what order. That, in turn, could result in groups that are not similar to each other at the start of the study. For example, a researcher who knows the next study subject will be assigned to a group with weekly study visits may recruit someone who lives nearby for that spot, and wait to recruit someone who lives far away for the group that will only complete questionnaires. Similarly, a potential study subject who already knows to which group he or she will be assigned may decide whether to take part based on personal opinions about the treatment being offered in that group. In either circumstance, group differences that could affect the outcomes being studied have been introduced before the treatment even begins.

Random assignment strategies with the least potential for this sort of biased selection are those that are concealed from both the researcher and the potential study subject until that person has agreed to take part in the study, no matter which group assignment he or she receives. Examples include random assignments made by opening opaque envelopes in strict order and assignments made by a central study office, over which neither researcher nor subject has any control.

***Were all assigned group members included in the final comparisons?***

Recall the example of a study where one group had weekly treatment visits and the other group simply completed questionnaires. When the burden of participation in one group is greater than the burden of participation in the other group, persons with greater health challenges or less motivation will be more likely to drop out of the more burdensome group over the course of the study. As a result, the persons remaining in the groups at the end of the study may no longer be similar to each other. Comparing only their outcomes provides biased information about the effectiveness of the therapy. The remedy for this problem, when not every assigned subject completes the study, is intention-to-treat analysis. In this analysis, the groups are compared using the outcomes of all the persons originally assigned, even if some of those persons did not receive any or all of the therapy being tested. This analysis is a fairer group comparison and more accurately reflects the likely application and effectiveness of the therapy in clinical practice.

***Were the groups really similar to each other in ways that affect the outcome?***

Although random assignment is the best strategy for creating groups that are similar to each other, it is not perfect. This is particularly true when the study involves small numbers of subjects. It is reassuring when the researchers compare their groups on the basis of other factors that could affect the outcome – for example age or disease severity – and demonstrate that there are no significant differences between the groups. In essence, this is evidence that randomization "worked". If the groups are similar on the known factors that affect outcome, it is reasonable to think they are also similar on unrecognized factors that might affect outcome.

In the extremely rare event that nurses are considering therapies based on evidence obtained from a single study without random assignment, the answer to this appraisal question becomes vitally important. In the absence of random assignment, we must consider whether the way the groups were formed would introduce differences in factors affecting outcome. For example, when therapy comparisons involve groups of patients on different care units, it is essential to know how patients are assigned to those units. Are there differences in diagnosis or level of acuity? Are there social or environmental differences between the units that could affect outcomes? The evidence should include an especially thorough comparison of the groups on all factors that could compete with the therapy as explanations of the outcomes.

**Did the groups being compared remain similar to each other, except for the therapy, during the study?**

After deciding the study groups were formed in a way that supports fair comparisons, nurses need to determine whether anything happened during the course of the study to alter the fairness of that comparison. Students of research design may be familiar with Campbell and Stanley's [3] list of threats to validity, or things that can "go wrong" in the research. When groups of randomly assigned subjects are being studied during the same time period in the same places and outcomes are being measured in the same way, most of these threat categories raise little concern. Any effect of history, maturation, testing or instrumentation probably affects the outcomes of all groups in the same way and would, in effect, cancel out of the group comparisons. The threats to validity that remain are related to the possible impact of participants' knowledge of group assignment on outcomes.

***Did the group members know which group they were in?***

Therapy evidence often involves comparisons between groups receiving some new therapy and groups receiving usual care. Drug studies often include a comparison group that is receiving placebo – a look-alike pill that has no active ingredients. Some therapy outcomes are measured by participant self-report. In these situations, it could be argued that the participants' knowledge of what treatment they received, rather than the treatment itself, resulted in group differences in the outcomes. For example, hospitalized adults who are assigned to receive an experimental therapy for chronic pain may ignore or under-report side effects because of the excitement of trying something new. When subjects do not know and cannot guess whether they are receiving the new therapy or usual care, their self-eports cannot be biased by that knowledge. This is, of course, easier to accomplish with some therapies (e.g. drugs and topical therapies) than it is with others (e.g. group vs. individual diabetes education).

***Were the persons providing clinical care to the group members aware of which group they were in?***

Whether or not the group members know which study group they are in, their experience during the course of the clinical trial can be different if the persons providing their care are aware of group assignment. For example, a study of therapies for regaining mobility might compare the effect of daily intensive physical therapy to usual care for elderly patients recovering from a hip fracture. Nurses caring for patients in both groups might neglect their usual ambulation efforts for the patients they knew were also walking in physical therapy or increase their efforts for those patients who were not. Either way, the therapy is no longer the only experience that is different between the groups: a competing explanation for differing outcomes now exists.

Where possible, the remedy for these situations is to keep study participants and their caregivers from knowing which participant is assigned to which group. This is described as "blinding" or "masking" the study. "Single-blind" means that either the study subjects or their caregivers (but not both) do not know group assignments. "Double-blind" means that neither the study subjects nor their caregivers know the group assignments.

***Did outcome assessors know which participants belonged to each group?***

Some outcomes are objective and we do not have to worry about whether the results are biased by the opinions of the person assessing them. Clinical lab results, weight gain or loss measured on an accurate scale and scores on standardized tests are examples. Many outcomes, however, are subject to interpretation. Does this mother's behavior indicate she is more responsive to her infant than she was previously? Has this patient experienced clinical improvement in his mood? Is this adolescent more motivated to practice safe sex? When outcomes are subject to interpretation and the researcher assessing those outcomes knows which treatment the participant received, there is a potential for bias. A researcher who has preconceived opinions about the benefit of the new therapy could overrate the outcomes of participants who received it and underrate the outcomes of those who did not.
The remedy here is again blinding or masking. When the person making the assessments is unaware of the group assignment of the study participant being assessed, the fairness of the group comparisons is supported.

Note that the assessor may have no other role in the research than determining outcomes. This is a particularly desirable arrangement in those situations

where group assignment cannot be hidden from the subject and / or caregiver, but could be from an outcomes assessor. When the assessor is also the caregiver, concealment of participant study group supports the assumption that care experiences and outcome assessments are comparable across study groups.

***Was follow-up complete?***

Ideally, every participant who enters a clinical trial would continue to take part throughout the entire trial and would be available for all outcome assessments. In practice, this is rarely the case. Subjects drop out of clinical trials, perhaps because of the burdens of the trial, perhaps for other reasons. This raises the possibility that the participants remaining in the study groups are no longer similar to each other. The greater the proportion of subjects lost, the greater the risk that group comparisons will not be fair tests of the effect of the therapy. (Campbell and Stanley[3] named this threat to validity "mortality", whether or not lost participants actually died). Intention- to-treat analysis, discussed previously, addresses the difficulties of comparing groups when some members do not complete all study activities, but it requires outcome measures for all participants. When outcomes are not available for all study participants, the minimum research response is a comparison of the lost subjects to the remaining subjects on the basis of available information (e.g. demographics, interim measures obtained before the participant was lost).

The strongest evidence to support the effectiveness of a therapy when some outcomes are missing is a group comparison where "worst case" outcomes are included for participants lost to follow-up. Missing group members from the therapy group are assigned values reflecting undesirable outcomes and missing group members from the comparison group are assigned values reflecting desirable outcomes. If group comparison with these assumed outcomes included does not change the conclusion about the effectiveness of the therapy, the missing participants are not a threat to the validity of the study findings.

## What are the results?

**How large was the treatment effect?**

Nurses need to make two judgments about the size of group differences in outcomes associated with a therapy:

- Is the difference between groups "real"?
- If the difference between groups is "real", is it clinically meaningful?

Treatment effects for outcomes measured as categorical variables are expressed as some comparison of the relative rates of occurrence of the outcome among the study groups. Treatment effects for outcomes measured as continuous variables are expressed as group mean differences. For details of these statistics, see my previous article[4] on statistics for therapy. The test for a "real" group difference is whether or not the difference is statistically significant, that is, unlikely to have occurred through sampling chance alone. The test for a clinically meaningful difference is the nurse's informed clinical judgment.

**How precise was the estimate of the treatment effect?**

In essence, precision is an estimate of the range of group differences in outcomes that might occur if the study were repeated with similar participants. For nurses considering the clinical use of a new therapy, precision provides an estimate of the smallest and greatest treatment effect the therapy is likely to produce. This range allows nurses to decide how confident they are that the new therapy is worthwhile to employ. See my previous article[4] for additional discussion.

## How can I apply the results to patient care?

The first two major sections of the evidence-based appraisal format for therapy evidence are focused on the content of the single study being reviewed: whether the results are likely to be trustworthy, based on the study design, and, if so, what those findings are. The third section guides nurses to consider whether the findings are relevant to their own practice situations.

**Were the study patients similar to the patients in my care?**

Students of research design will recognize this question as related to generalizability. Under what circumstances can we expect results from one set of participants will apply to another group?

***Does my patient match the study inclusion criteria?***

The statistical approach to generalizability sets sampling standards that are almost impossible to meet in any clinical trial with human subjects. Accepting the statistical approach makes nurses hesitant to apply any research findings to their practice. The evidence-based practice approach is more permissive and simply asks whether the patients who will receive the new therapy would have qualified to be in the research study. If so, it is reasonable to assume that the evidence is applicable.

***If my patient does not match the study inclusion criteria, are there compelling reasons why the results should not apply?***

Nurses are in an ideal position to understand the cultural, social, family and personal circumstances of their patients and to consider whether some crucial factor would alter the expected effects of an evidence-based therapy. Nurses are also in an ideal position to observe commonalities across diverse patient groups that would suggest a therapy might be effective for patients outwardly quite different from those studied in the clinical trial. The evidence-based practice approach to appraisal encourages application, asking appraisers to justify a decision not to do so.

**Were all the clinically important outcomes considered?**

***What were the primary and secondary outcomes of the study?***

Therapies often have effects other than the intended outcome, and those other effects may be important considerations in practice. Patients may be hesitant to pursue an effective smoking cessation therapy for fear of weight gain. Pediatric anticipatory guidance could increase anxiety for some parents, instead of confidence. Before deciding whether to employ the evidence in practice, nurses need to determine whether the evidence is complete enough to support that decision. Strong therapy evidence includes results for all the clinically important desirable and undesirable outcomes associated with the therapy.

***Were surrogate endpoints, as opposed to outcomes important to patients, used?***

Some therapy outcomes are obvious within minutes, while others take a lifetime to appear. As a consequence, clinical trials may measure outcomes in terms of markers thought to be related to the eventual clinical outcome of interest. For example, changes in knowledge or expressed intention may be measured instead of actual changes in risk behaviors. Short-term changes in cholesterol levels may be measured to represent long-term risk of heart attack or stroke. When evidence relies on surrogate outcomes, nurses need to consider how well the surrogate represents one or more clinically important outcomes for patients.

**Are the likely treatment benefits worth the potential harm and costs?**

This final question is at the core of clinical decision-making. Is this evidence worth applying in the care of my patients? This decision depends on the appraisals already made:

- whether I can trust the results (validity)
- whether the results apply to my patients (generalizability)
- whether the results are important enough to consider applying (magnitude of effect, use of surrogate outcomes)
- whether the results are complete enough to allow me to make a decision.

When the answer to all these questions is yes, nurses can use their clinical expertise, their knowledge of available resources and constraints, their knowledge of their patients' values and the findings from the evidence to chart an effective and efficient clinical course.

## References

1. Grace J. Essential skills for evidence-based practice: Understanding and using systematic reviews. J Nurs Sci. 2011;28(4):20-5.
2. Guyatt G, Rennie D, editors. (eds.) Users' guides to the medical literature. Chicago IL: AMA Press; 2002.
3. Campbell, DT, Stanley, JC. Experimental and quasi-experimental designs for research. Boston: Houghton Mifflin; 1963.
4. Grace J. Essential skills for evidence-based practice: Statistics for therapy questions. J Nurs Sci. 2010; 28(1):8-12.