

# Essential Skills for Evidence-based Practice: Strength of Evidence

**Jeanne Grace**

*Corresponding Author:*

*J. Grace*

*E-mail: Jeanne\_Grace@urmc.rochester.edu*

*Jeanne Grace RN PhD*

*Emeritus Clinical Professor of Nursing*

*University of Rochester*

*Rochester, New York, USA*

## **Abstract**

The strength of the evidence to answer any clinical question depends on the quantity and quality of the evidence and the consistency of findings across studies. The nature of the clinical question determines what study designs provide the strongest evidence. Randomized clinical trials provide the strongest evidence for therapy and harm questions, while various descriptive study designs provide strongest evidence for prognosis, diagnosis and human response / meaning questions. A systematic review of individual studies is a particularly efficient way for practitioners to determine the quantity of evidence as well as the consistency of findings. For evidence-based practice guidelines, explicit scales rate the strength of evidence underlying recommendations.

**Keywords:** Evidence - based practice (EBP), strength of evidence

*J Nurs Sci 2009;27(2): 8 - 13*

All evidence is not created equal. Nurses who wish to apply evidence to their clinical practice must be able to recognize what forms of evidence offer the strongest support for addressing clinical questions. Strong evidence has three general characteristics: quantity, quality and consistency. We are most confident about applying evidence when there are several (quantity) relevant well-designed (quality) studies and when the results of those studies agree with each other (consistency). As a result, a rigorous and unbiased summary of several studies, in the form of a systematic review or meta-analysis, is generally preferable to evidence from any single study.

A systematic review is a comprehensive review of the available evidence for a specifically defined clinical question. It is

conducted according to a pre-planned strategy intended to reduce sources of bias. Specifically the plan is intended to assure that all available evidence is considered for inclusion and that the decisions about what evidence to include are based on objective measures of design quality. Whenever possible, the results of similar studies are presented together and the combined results are summarized in a meta-analysis. This meta-analysis often includes statistics that estimate the consistency of the findings among studies. Systematic reviews are most frequently available for clinical questions in the therapy domain, although systematic reviews of some harm and diagnosis questions are also available.

We can think of all the evidence that supports our response to a clinical problem as

forming a pyramid. At the base of the pyramid, serving as a foundation for knowledge development, is evidence that was collected less systematically (for example, clinical observations) or from sources that are not directly applicable to clinical care (for example, animal studies or physiology lab experiments). At each step up the pyramid, the research designs can build on the evidence base generated from study designs on the lower steps, until finally, at the top, are designs capable of providing the strongest evidence – evidence we can confidently apply to clinical practice. Systematic reviews of multiple studies with the strongest design generally occupy the highest step of the pyramid.

The nature of the clinical question determines what sorts of research designs for individual studies have the highest quality as evidence, whether considered individually or combined in a systematic review. Therefore, there is not a single strength of evidence pyramid, there are several.

Many of the clinical questions that interest nurses fall into either the therapy or harm domain. Therapy questions ask “How can I make something desirable happen for my patient?” Harm questions ask “Why does something undesirable happen to my patient?” At the most basic level, these are questions about cause and effect or outcome. The scientific standards for demonstrating that some cause results in some outcome include

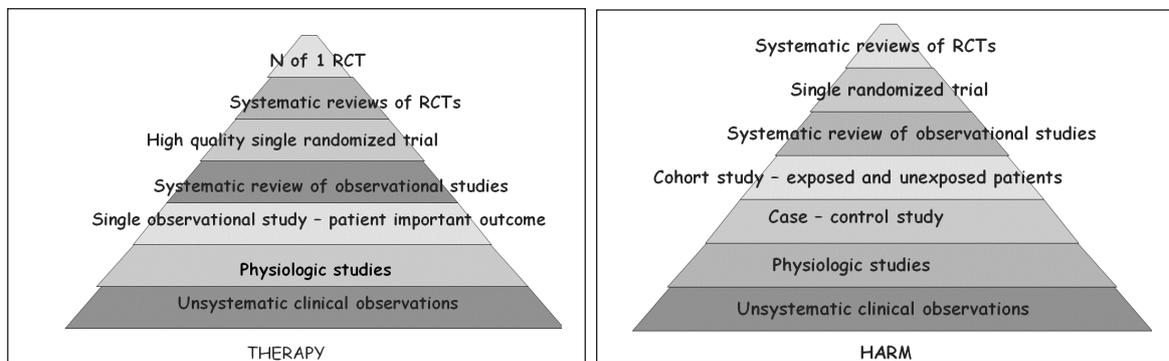
- establishing a relationship between variations in the exposure to the cause and variations in the outcome
- determining that the cause happened before the outcome
- ruling out other possible explanations of the outcome (competing causes).

Observational study designs are satisfactory for establishing a relationship and sometimes adequate for determining that the cause happens before the outcome. Ruling out competing causes generally involves including some way to control the effects of those competing causes in the study design.

The higher levels of strength of evidence pyramids for therapy and harm questions represent designs with increasingly effective ways to achieve control over the impacts of competing causes. The strongest research design to accomplish this control is a true experiment, implemented in health care as a randomized clinical trial (RCT). The hallmarks of this study design are that the researcher can control which subjects are exposed to the cause and which ones aren't (manipulation of the independent variable) and that subjects are assigned to the exposure and comparison groups by chance alone (randomization). Randomization is intended to assure that the groups being compared each contain the same mix of persons exposed to possible competing causes, whether those competing causes are recognized by the researcher or not. When study subjects are randomly assigned to groups, any effects of those causes would be the same on the outcomes for each group. These design features provide the strongest support for the claim that it is what the researcher did, and not something else, that accounts for differences in outcomes among the groups being compared.

Figure 1 presents strength of evidence pyramids for therapy and harm domain questions, based on evidence strength rankings published by Guyette and Rennie<sup>1</sup> and Ebell and associates<sup>2</sup>. The top of the therapy pyramid, considered even stronger evidence than a systematic review, is the “N of 1 randomized clinical trial (RCT)”. This is a very unusual within-subject study design, in which a single subject is exposed to all the therapies being studied in random order, with the response to each therapy compared to the other therapy responses for that same individual. It is considered the very strongest study design when the person being studied is the person for whom the answer is being sought. Thus, this design not only has controls for competing explanations, but also guarantees that the results of the study will apply to the patient of interest, because the study subject and patient are the same person.

**Figure 1** *Strength of Evidence Pyramids for Therapy and Harm*



(Therapy pyramid adapted from rankings of Guyette and Rennie<sup>1</sup> and reprinted from Levin, RF, Feldman, HR, editors. Teaching evidence-based practice in nursing. New York: Springer Publishing; 2006. Reproduced with the permission of Springer Publishing Company, LLC, New York, NY 10036. Harm pyramid adapted from SORT rankings<sup>2</sup> and reprinted from Grace JT, Powers BA. Claiming our core: appraising qualitative evidence for nursing questions about human response and meaning. Nurs Outlook. 2009;57(1):27-34, with permission from Elsevier.)

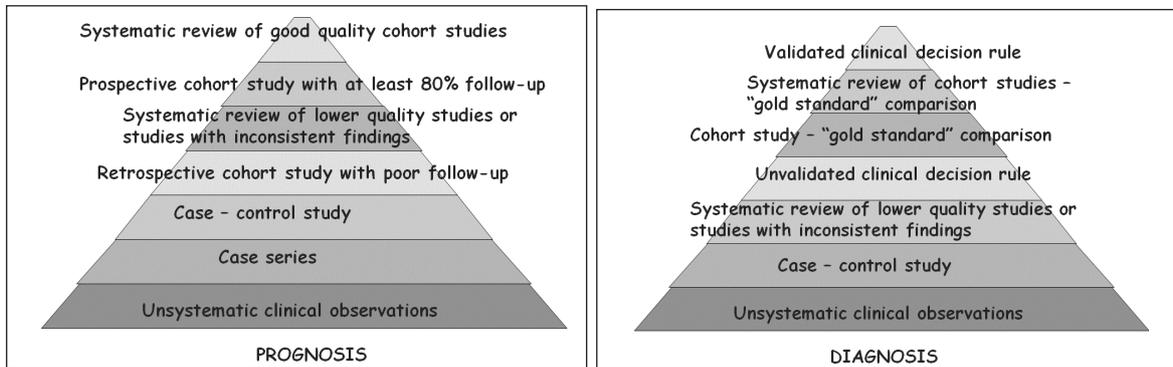
The strength of evidence pyramid for harm resembles the single strength of evidence pyramid widely published in early works on evidence-based practice. Although the randomized clinical trial is acknowledged as the strongest study design for establishing the cause of some undesirable outcome, there are many circumstances where that type of evidence cannot be obtained. We cannot ethically randomize study subjects to be exposed to known harms when there is no balancing benefit to be gained. Sometimes, the cause takes so long to result in the outcome, or the outcome is so uncommon, that it is simply not practical to design a randomized clinical trial to obtain evidence. In these circumstances, it is helpful to understand which observational study designs provide relatively strong evidence to guide our practice.

Cohort studies are observational studies in which a group of people with varying amounts of exposure to some potential harm are followed from the time the exposure is measured until their outcomes are known. For example, we could measure cigarette use among a group of apparently healthy adults, then follow them over time to see who develops cardiac and respiratory problems. The results would allow us to establish whether amount of smoking was related to likelihood of developing health problems and to determine that our proposed cause (smoking) happened before our

proposed outcome (health problems). We could only control for competing causes by restricting the study to subjects who all had the same exposure to those competing causes (for example, excluding subjects with known risk factors for health problems) or by statistical means. Both of these control strategies can only be applied for competing causes the researcher knows about, so this is a weaker design than random assignment in a clinical trial, which also controls for unrecognized competing causes.

Whereas the cohort study design looks forward in time, the case-control study design looks backward. Subjects are recruited after their outcomes are known (cases have the outcome of interest and controls do not), then researchers determine whether cases had more exposure to the proposed cause in the past than controls did. For example, we could identify a group of persons with (cases) and without (controls) lung cancer and compare their histories of smoking for the past twenty years. While it is possible to establish relationship between cause and outcome in this design and to determine that the cause happened before the outcome, the strategies for controlling for competing causes are even more limited than for a cohort study design. When outcomes are rare, however, or the cause takes a long time to result in the outcome, case-control studies may provide the strongest available evidence for practice.

**Figure 2** Strength of Evidence Pyramids for Diagnosis and Therapy



(Diagnosis and prognosis pyramids adapted from SORT rankings.<sup>2</sup> Diagnosis pyramid reprinted from Grace JT, Powers BA. Claiming our core: appraising qualitative evidence for nursing questions about human response and meaning. *Nurs Outlook*. 2009;57(1):27-34, with permission from Elsevier. Prognosis pyramid reprinted from Levin, RF, Feldman, HR, editors. *Teaching evidence-based practice in nursing*. New York: Spring Publishing; 2006. Reproduced with the permission of Springer Publishing Company, LLE, New York, NY 10036.)

The clinical question domains of diagnosis and prognosis do not address cause and effect. Diagnosis questions ask about the accuracy of signs, symptoms and tests to determine the nature of a patient's current problem. Prognosis questions ask what can be expected in the future, given the patient's current problem. The strongest evidence for these types of questions comes from observational (also called descriptive) studies, not experiments. The hallmark of these designs is that the researcher conducts structured observations, but does not attempt to alter the outcomes being observed.

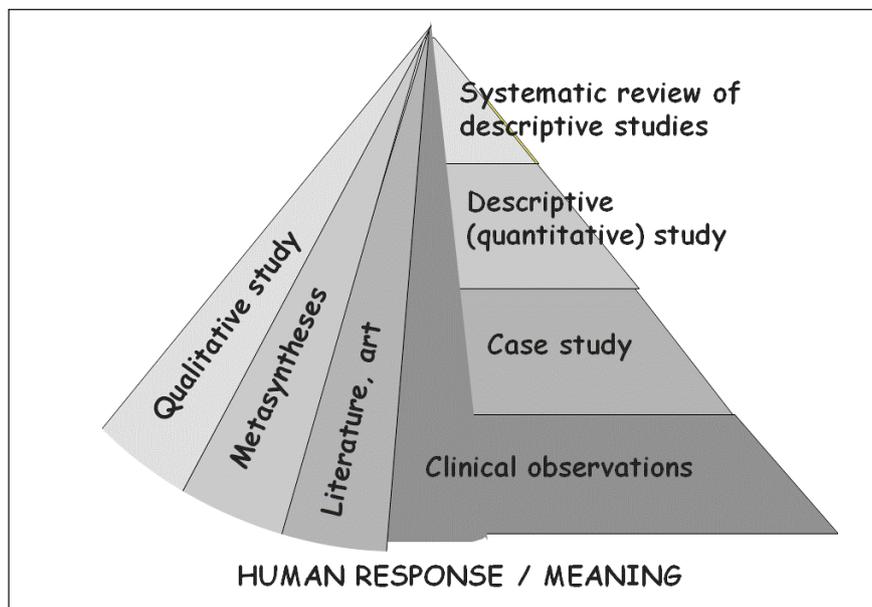
Figure 2 presents strength of evidence pyramids for diagnosis and prognosis questions, based on rankings published by Ebell and associates.<sup>2</sup> For diagnosis questions, the evidence is stronger when the means of diagnosis being studied can be compared to some established way to obtain an accurate diagnosis (the "gold standard") and both are used independently on study subjects. The evidence is also stronger when it comes from a study conducted at the point the subjects were in need of diagnosis (cohort study) rather than after the actual diagnosis was known (case control study). The very strongest evidence, the validated clinical decision rule, involves demonstrating that the diagnostic strategies of interest, applied in a systematic manner, result

in better outcomes for patients than the standard way of diagnosing their condition. In other words, the validated clinical decision rule provides evidence that the diagnostic test(s), signs or symptoms are not only accurate, but also useful.

Because prognosis questions ask about what is likely to happen in the future, the strongest research designs are those that follow subjects from a current known status (for example, current stage of some disease) into the future to see what happens to them. Studies that involve data collection at more than one time point are described as "longitudinal". Prospective (looking forward in time) designs are stronger than retrospective (looking back in time) because of the possibility that someone's recall of the past might be biased by their knowledge of what the outcome has been.

One potential problem with a longitudinal study is subject dropout over the course of the study. When a large number of subjects fail to complete the study, it is very possible that the outcomes measured for the remaining subjects are not like the outcomes of the dropout subjects. If so, the study results do not accurately represent the entire cohort that began the study. Studies that are able to follow and measure outcomes for at least 80% of their original subjects have the highest quality for prognosis evidence.

Figure 3 *Strength of Evidence Pyramid for Human Response / Meaning*



(Human response / meaning pyramid reprinted from Grace JT, Powers BA. Claiming our core: appraising qualitative evidence for nursing questions about human response and meaning. Nurs Outlook. 2009;57(1):27-34, with permission from Elsevier.)

Like the prognosis and diagnosis domains, strongest evidence for the clinical question domain of human response / meaning comes from observational studies. The nature of the clinically useful evidence for human response / meaning, however, is different from that for diagnosis and prognosis and is of particular interest to nurses. Best evidence for diagnosis and prognosis provides information about what is generally true in a specific clinical situation. We are interested in what is generally true about our patients' responses to health situations and the meanings they attach to their experiences, but we are also interested in the potential variety of unique responses and meanings our patients may be experiencing. As a result, we draw on evidence from both the quantitative research traditions and from the qualitative (interpretive) research traditions.

Figure 3 presents the strength of evidence pyramid Bethel Powers and I have developed for human response and meaning questions. The hallmarks of observational study design quality in the quantitative tradition are systematic

observation of a representative sample of subjects and accurate measurement of their responses to health situations. This is depicted on the right side of the evidence pyramid in a progression of study designs providing increasingly strong evidence for what is generally true.

The left side of the pyramid represents evidence from qualitative research traditions, where no specific tradition or study design necessarily provides stronger evidence for practice than any other. Evidence from a metasynthesis (the qualitative counterpart for a systematic review) is not necessarily stronger than the evidence from an individual study. Qualitative research provides evidence about possibilities of human response and meaning, not probabilities. The nurse cannot assume the qualitative evidence applies to his/her patient, but awareness of the possibilities heightens the nurse's potential for empathy and enhances the nurse's ability to make effective nurse – patient partnerships for health.<sup>3</sup>

For a growing number of clinical problems, the available evidence that addresses multiple

question domains around a specific health problem has been synthesized into a comprehensive set of recommendations, known as a clinical practice guideline. The intent of a practice guideline is to suggest the most effective and efficient ways to assess and manage some health care issue, based on the best available evidence. Because a practice guideline addresses all the care needs of the patient with that problem, some of the recommendations are based on stronger evidence than others. Guideline developers may include ratings to reflect this in the guideline itself.

There are two types of strength ratings that may be included in clinical practice guidelines. One is the rating of the strength of evidence and the other is the rating of the strength of the recommendation. There are many different rating scales in use, and you must read the guideline carefully to determine which ones are being used.

Rankings for strength of evidence indicate how strong the evidence supporting the specific recommendation is, and they generally mirror the strength of evidence pyramids for therapy and harm. The highest rankings go to recommendations supported by a well-done systematic review or more than one well-designed study with consistent findings. Lower rankings go to recommendations for which the available evidence is not consistent or is based on less desirable study designs, and the lowest rankings represent recommendations based on expert clinical opinion, unsupported by any systematic studies.

Strength of recommendation ratings indicates how important the specific recommendation is, clinically. In some situations, there is very good evidence that two or more therapies result in the same outcome. The strength of evidence ranking for the suggested course of action would be high, while

the strength of recommendation ranking would be low. In other words, there is very good evidence that either choice of treatment would accomplish the desired outcome, so the choice can be made based on patient preferences, not relative effectiveness. There are also situations where one treatment is strongly recommended over the other because of effectiveness or safety (high strength of recommendation rating), but the support for that recommendation is clinical opinion only (low strength of evidence rating). Although the nursing knowledge base is growing rapidly, we encounter many clinical problems where we do not yet have a strong evidence base for practice. When we are faced with a lack of evidence from the strongest research designs, the evidence pyramids help us identify the relative merits of other study designs as sources of clinical evidence. It is also important to remember that study design is only one aspect of evidence quality: the care with which the study is conducted, the relevance of the outcomes studied and the similarity of the study subjects to the patients of interest are also important considerations.

## References

1. Guyette, G, Rennie, D. Users' guides to the medical literature. Chicago (IL): AMA Press; 2002.
2. Ebell MH, Siwek J, Weiss BD, Woolf SH, Susman J, Ewigman B, et al. Strength of recommendation taxonomy (SORT): A patient-centered approach to grading evidence in the medical literature. *Am Fam Physician*. 2004;69(3):548-56.
3. Grace JT, Powers BA. Claiming our core: appraising qualitative evidence for nursing questions about human response and meaning. *Nurs Outlook*. 2009;57(1):27-34.