

PT Bulletin Online Article:
All Evidence is Not Created Equal

All Evidence is Not Created Equal: A Discussion of Levels of Evidence

By Steven Glaros

Part of evidence-based practice is knowing what evidence to use. Levels-of-evidence models provide guidance for critically appraising literature.

You have a patient and want to make sure that your intervention is effective and appropriate. You devise a clinical question, log onto PubMed (the Internet version of MEDLINE), and search for evidence on the effectiveness of the intervention. PubMed displays an intimidating 30 "hits" that might be appropriate for your patient. Which articles should you use? Is all this evidence of equal value?

Jules Rothstein, PT, PhD, FAPTA, editor in chief of *Physical Therapy*, has written, "Evidence-based practice means using the best possible evidence and recognizing that not all evidence is created equal."^{1(p7)} Rothstein, a vocal proponent of evidence-based practice, explains, "Think of a dramatized court trial. Some evidence clearly speaks to the issue more than other evidence."

One commonly cited definition of evidence-based practice--"the integration of best research evidence with clinical expertise and patient values"^{2(p1)}--indicates that some evidence is better than other evidence. David Scalzitti, PT, MS, OCS, APTA's associate director of research, manages the Hooked on Evidence initiative, a grassroots effort to collect the evidence on physical therapy interventions. He says, "The evidence you choose will depend on the patient's problem and values. This evidence will then be combined with your clinical experience to determine how it will fit into your management of the patient."

Marianne Orest, PT, MEd, is one of three clinical research educators in the Rehabilitation Therapies Department at Fletcher Allen Health Care in Burlington, Vermont. She and her colleagues have begun a formal process of appraising the literature. "The first part of carrying out evidence-based practice is knowing what evidence to use and what evidence not to use," she says. "People often look at the abstract or look at the conclusions of an article and take them at face value; however, we want people to be able to critically appraise articles, to really look at some of their strengths and weaknesses."

To help clinicians critically appraise the literature, some of the foremost proponents of evidence-based practice, led by David Sackett, MD, have developed a model for levels of evidence, which stratifies study designs based on the level of confidence that a clinician can have in applying a study's results to an individual patient (Table, below).² Certain study designs are given a greater level of confidence because they rigorously address bias and confounding factors. Because not all research designs are appropriate for all aspects of patient/client management, different aspects of patient care have different levels-of-evidence hierarchies.³

Levels of Evidence

The study designs described in the Sackett levels of evidence for intervention are:

- Systematic reviews
- Randomized controlled trials
- Cohort studies
- Case-control studies
- Case series studies

Systematic review-"A summary of the medical literature that uses explicit methods to perform a thorough literature search and critical appraisal of individual studies and that uses appropriate statistical techniques to combine these valid studies."²

The systematic review, according to Sackett et al, is the highest level of evidence. The goal of a systematic review is "to minimize both bias (usually by not only restricting itself to randomized trials, but also seeking published and unpublished reports in every language) and random error (by amassing very large numbers of individuals)."^{2(p133)} The results of these individual trials sometimes are combined using statistical methods in a process called meta-analysis.

Because systematic reviews summarize and combine the results of several studies, they may be ideal sources of evidence for busy clinicians. As Guyatt et al state, "Efficient evidence-based practice dictates bypassing the critical assessment of primary studies and moving straight to the evaluation of rigorous systematic literature reviews."^{3(p2)}

Like evidence in general, however, all systematic reviews are not created equal; they are only as strong as the design of the studies they summarize. For instance, systematic reviews of randomized controlled trials (RCTs) are ranked higher than systematic reviews of cohort studies or case-control studies (see below). A systematic review of RCTs further reduces possible bias and random error that may be found in RCTs, whereas "systematic reviews of nonrandomized trials can compound the problems of individually misleading trials and produce a lower quality of evidence."^{2(p134)}

In addition, many steps in conducting a systematic review involve making judgments or choices, which may introduce bias. However, the use of rigorous procedures can reduce the likelihood that bias will occur.⁴ Nevertheless, Rothstein says, "Systematic reviews are not as bias-free as their proponents would have you believe."

Researchers may reach different conclusions if they use different criteria to select articles for the review. Physical therapists, Rothstein argues, should look closely at the systematic review to see if they agree with the authors' methods and conclusions. "The systematic review takes an active, thinking reader to apply it," he says.

For physical therapists, searches for systematic reviews that are applicable to their patients often may be fruitless. According to Charles Ciccone, PT, PhD, editor for Evidence in Practice and Reviews for Physical Therapy, systematic reviews "are not available in the physical therapy literature to any great extent" because the studies needed for systematic review are not available. Rothstein explains, "Systematic reviews are an excellent way of aggregating lots of data, but you have to have lots of data to aggregate; you have to have reasonably large numbers of strong studies."

A notable exception to the lack of systematic reviews in the physical therapy literature are the clinical practice guidelines for management of musculoskeletal

conditions published by the Philadelphia Panel in the October 2001 special issue of Physical Therapy. In view of the relative dearth of systematic reviews in the physical therapy literature, one of the long-range goals of the Hooked on Evidence initiative is to develop systematic reviews based on the extracted articles in the database.⁵

Randomized controlled trial-"A group of patients is randomized into an experimental group and a control group. These groups are followed up for the variables/outcomes of interest."²

Randomized controlled trials constitute the next highest level of evidence. The term is roughly equivalent to "randomized clinical trial" used in the Hooked on Evidence database.⁶ Scalzitti estimates that more than half of the 837 articles that have been entered into Hooked on Evidence so far are some form of clinical trial. Well-conducted RCTs offer some advantages over other types of evidence, the most important of which is that bias has been systematically addressed through randomization, blinding, and other methods.² Rothstein says, "RCTs are considered the highest level of evidence because they are the one place where you isolate and manipulate the treatment effect. By using a control group and all the other components of an RCT, we can say 'this caused that.' There's really no other study that allows us to say this as strongly."

In theory, randomization, by creating experimental and control groups that are similar, allows the outcome of interest—typically the outcome of an intervention—to be isolated and helps to minimize both known and unknown factors from affecting the results of the study.⁷ In well-designed RCTs, both the subjects and the researchers are unaware of ("blinded" to) the group assignment. When the subjects or the investigators are aware of group assignments, the possibility increases that they will unintentionally distort a measurement or overestimate or underestimate the effectiveness of an intervention.⁷ Ciccone explains, "If a clinician measuring range of motion isn't blinded or senses that a subject received the intervention, he or she unintentionally might put a little more force on the limb to get a little more out of it or might encourage the subject a bit more." Similarly, he adds, "Patients often want to please the therapist, and they'll try a little harder if they think that they're supposed to show a reaction."

RCTs, however, are not infallible. Bias still may exist, especially if the subject groups contain volunteers, which would be a form of selection bias. RCTs also are expensive and difficult to conduct, often costing millions of dollars.⁸ In addition, Rothstein says, "RCTs are difficult because it's often hard to get the proper patients in the proper setting under the proper advisement." These factors may limit the number of RCTs your literature search produces.

Table

Levels of Evidence for Interventions

Level of Evidence	Type of Study
1a	Systematic reviews of randomized controlled trials (RCTs)
1b	Individual RCTs with narrow confidence interval
2a	Systematic reviews of cohort studies
2b	Individual cohort studies and low-quality RCTs
3a	Systematic reviews of case-control studies
3b	Case-controlled studies
4	Case series and poor-quality cohort and case-control studies
5	Expert opinion
Adapted from Sackett DL, Straus SE, Richardson WS, et al. <i>Evidence-Based Medicine: How to Practice and Teach EBM</i> . 2nd ed. Edinburgh, Scotland: Churchill Livingstone Inc; 2000:173-177.	

An RCT also may not be applicable because it is an experimental study. As Scalzitti explains, "Sackett's hierarchy puts greater weight on studies of higher quality and with more control to show whether, in experimental situations, a treatment is more effective. Sometimes, however, these same controls can't be reproduced in a clinical setting. Other forms of evidence such as cohort studies, therefore, might take precedence in decision making over either a systematic review or an RCT." Orest says, "What we've tried to do is avoid saying that the randomized controlled trial is the only evidence that you can use because many of those articles may not be directly relevant to your patient."

Some physical therapists claim that RCTs cannot be conducted in many areas of practice because it is impossible to create a "pure" control group that receives either no intervention or a placebo intervention. Rothstein disagrees. "The total absence of a treatment is not the only way to run a randomized controlled trial," he says. "You can compare one intervention with an alternative intervention. Obviously, when you compare two interventions, you cannot discuss the total absence of intervention." In addition, your search for an RCT on the potential harm of an intervention is likely to be fruitless. Medical and research ethics prevent randomly assigning subjects to a potentially harmful intervention or exposure (eg, smoking); furthermore, the length of time or the number of subjects needed to reveal a harmful effect may be prohibitive. When addressing issues of harm, you must look for other types of studies such as a nonrandomized clinical trial or cohort study.⁹

Cohort study-"Involves identification of two groups (cohorts) of patients, one [that] received the exposure of interest, and one [that] did not, and following these cohorts forward for the outcome of interest."²

The cohort study*--sometimes referred to as a prospective study, longitudinal study, inception cohort, or historical cohort--is the next level of evidence on Sackett's hierarchy. Although this study design is ranked lower than an RCT for interventions, it arguably is the highest level of evidence (other than a systematic review) for other aspects of patient/client management-examination, evaluation, diagnosis, and prognosis--or when ethical issues preclude the use of RCTs, such as possibly harmful interventions or exposures.^{9,10}

A cohort study is an observational study design--that is, outcomes are observed as they occur rather than being manipulated experimentally with an intervention. Cohort studies are used in a number of ways. When studying potential harm, groups of patients who have been have not been exposed to a particular agent (eg, smoking) are followed to see if they develop a particular outcome (eg, cancer).¹¹ To determine the criterion-based validity of a test or measure, for instance, researchers compare the new test to a "gold standard" (a test with previously accepted validity that is considered to be the standard for a particular situation) to see if the new measure provides valid measurements of the target condition or clinical situation.^{11,12} In the case of prognosis, researchers identify

** The definitions of cohort studies and case-control studies used by other epidemiologists and in Hooked on Evidence both have prospective and retrospective designs, whereas Sackett et al define cohort studies as prospective studies and case-control studies as retrospective studies. According to Scalzitti, both cohort studies and case-control studies can be retrospective if the investigation is initiated after both the exposure and outcome of interest have occurred. The difference is that a retrospective cohort study classifies participants on the basis of exposure of interest to determine the outcome of interest, whereas a retrospective case-control study classifies participants on the basis of the outcome of interest to determine the exposure of interest. For the sake of simplicity, this article will use the Sackett definitions.*

groups of patients who have or do not have a factor that might modify the prognosis and follow the groups to see if the patients have a particular outcome.¹¹

Cohort studies do have disadvantages. They do not use randomization and blinding procedures. Unknown confounding variables also may be responsible for differences in outcomes. The patients who have been exposed to a possible harmful agent may have a different level of risk for developing a specific outcome than patients who have not been exposed. Finally, rare outcomes or those that develop over an extended period of time may make a prospective cohort study impossible to conduct. In this case, a retrospective case-control study may offer the best evidence.⁹

Case-control study--"A study [that] involves identifying patients who have the outcome of interest (cases) and control patients without the same outcome, and looking back to see if they had the exposure of interest."²

The case-control study*--also called a prevalence study, cross-sectional study, and retrospective study--may be the only viable option when looking at rare outcomes or outcomes that occur long after the initial exposure. Generally, in a case-control study, patients who already have developed a particular outcome (ie, cases) are compared with a group of patients who do not have that outcome and who are similar to the patients in the case group in age, sex, and other important variables. The researchers then examine patient histories and adjust for unknown variables to see if the case group received the exposure of interest.⁹

Case-control studies, however, suffer from the same problems as cohort studies. The retrospective nature of most case-control studies makes this study design even more susceptible to the problems of confounding factors "because confounders that are transient or lead to early death won't even be able to be measured."^{2(p158)} In addition, this design relies on patient recollection or records to establish a patient's exposure, introducing potential recall and selection bias.⁸

Case series-"A report on a series of patients with an outcome of interest. No control group is involved."²

Case series reports include *Physical Therapy's* "case report" category. Case reports and case series studies describe the process of patient/client management--examination, evaluation, diagnosis, prognosis, intervention--and the outcomes for a single patient or a group of patients. Case reports also can focus on other issues beyond patient/client management, including "ethical dilemmas, use of equipment or devices, or administrative or educational concerns."^{13(p4)} Because case series studies and case reports are strictly descriptions of practice and do not have comparison or control groups, cause-and-effect relationships cannot be drawn from these reports and the outcomes they describe cannot be generalized to other patients.¹³ Sackett and colleagues, therefore, assign these studies to a low level of evidence.

Case reports and case series studies, however, should not be ignored in a literature search. Ciccone says, "It's the best way for a clinician to communicate with other clinicians. It's communication and documentation of what we do." Moreover, case reports can stimulate new research and "help develop practice guidelines and critical pathways."^{13(p3)} Research rarely addresses patient preferences, but clinicians must consider these preferences every day. Case series studies and case reports arguably can perform a critical function in evidence-based practice-by illustrating "how clinicians integrate the best available research evidence, clinical experience, and patient choice."^{13(p8)}

Expert opinion

Many advocates of evidence-based practice consider expert opinion to be the lowest level of evidence. Expert opinion should not be confused with clinical expertise--"the ability to use our clinical skills and past experience to rapidly identify each patient's unique health state and diagnosis, their individual risks and benefits of potential interventions, and their personal values and expectations"^{2(p1)}--which is an essential part of evidence-based practice. Sackett points out that "the non-experimental evidence that forms the recalled experiences of seasoned clinicians will tend to overestimate efficacy" of an intervention because the clinician is more likely to remember favorable responses and because of "the desire of patients and clinicians for success."^{14(p8)}

The unsystematic nature and unknown quality of expert opinion also is troubling. Expert opinion can be limited by small sample sizes and, "more importantly, by deficiencies in human processes of making inferences."^{3(p11)} Experts often make an inferential leap when attributing a result to an intervention rather than to a natural resolution of the problem or to other factors.

"Experts often say, 'That's what I've seen over the years,'" says Rothstein, "and you realize that they haven't even systematically looked at their own clinical information. As someone who's been in clinical practice, I believe you often get a bias about what works and doesn't work in

Examples of Study Designs Found in Physical Therapy^a

Systematic Review

Mortenson PA, Eng JJ. The use of casts in the management of joint mobility and hypertonia following brain injury in adults: a systematic review. *Phys Ther.* 2003;83:648-658.

Randomized Controlled Trial

Beaupre LA, Davies DM, Jones CA, Cinats JG. Exercise combined with continuous passive motion or slider board therapy compared with exercise only: a randomized controlled trial of patients following total knee arthroplasty. *Phys Ther.* 2001;81:1029-1037.

Cohort Study (Prospective)

Guccione AA, Fagerson TL, Anderson JJ. Regaining functional independence in the acute care setting following hip fracture. *Phys Ther.* 1996;76:818-826.

Cohort Study (Retrospective)

Kirk-Sanchez NJ, Roach KE. Relationship between duration of therapy services in a comprehensive rehabilitation program and mobility at discharge in patients with orthopedic problems. *Phys Ther.* 2001;81:888-895.

Case-Control Study (Prospective)

None available

Case-Control Study (Retrospective)

Mizner RL, Stevens JE, Snyder-Mackler L. Voluntary activation and decreased force production of the quadriceps femoris muscle after total knee arthroplasty. *Phys Ther.* 2003;83:359-365.

Case Study/Case Report

Pesanelli KE, Cigna JA, Basu SG, Morin AR. Occupational rehabilitation following open mesh surgical repair of an inguinal hernia. *Phys Ther.* 2003;83:58-66.

^a The studies listed here are examples that were selected randomly from *Physical Therapy's archives* and no attempt was made to assess their quality for the purposes of this article.

your own practice because you selectively remember things. In addition, experts often only look at certain evidence that supports and justifies their opinions and what they already do. That is not what evidence-based practice is all about. In evidence-based practice, you look at the entire body of evidence on the topic to make clinical decisions about the care of patients. Sometimes the evidence will support current practice, sometimes it will indicate that current practice should be changed." Part of the problem with expert opinion, both Ciccone and Rothstein say, is that experts do not contribute their work to the primary literature, even in the form of case reports. "Even articles on the lowest levels on the Sackett hierarchy undergo peer review. It casts a lot of doubt when you don't take even that step and document how you examined the patient, how you intervened, and how you measured the outcome," Ciccone says.

Not Etched in Stone

This levels-of-evidence model assigns a higher level of confidence to study designs that use more rigorous methods to ensure that bias has not affected the outcome of the study. The levels of evidence, however, are not a rigid set of rules; they are guidelines for the critical appraisal of literature. "You don't have to regard Sackett's levels of evidence as etched in stone," says Ciccone, "It is a place to begin, and the development of the hierarchy is based on fundamentally sound principles."

Ciccone also says that clinicians should not automatically reject studies with a lower level of evidence in favor of an RCT. "Some people treat levels of evidence like a poker game where a randomized controlled trial beats two cohort studies. That's not always true in evidence-based practice. Studies from lower levels may be better for your purposes or they may be better in terms of the quality of the study." As the Table illustrates, low-quality RCTs are given roughly the same weight in the Sackett hierarchy as well-conducted cohort studies.

Although you should look for the highest levels of evidence possible when attempting to answer your clinical problem, you ultimately are the judge of the applicability or generalizability of any study to your patient.

Using Evidence

In the mid-1980s, Tiziano Marovino, PT, a practice manager at Preferred Medicine Spine, Sports, and Occupational Medicine Inc in Allen Park, Michigan, was a student at McMaster University, a hotbed of the evidence-based practice movement, where Sackett and other proponents were faculty members. Not surprisingly, he is familiar with Sackett's levels-of-evidence model and has been using it since his student days. "By using the hierarchy outlined by Sackett," Marovino says, "clinicians can begin to categorize the various studies they've retrieved and assign a level of confidence based on the rigor of the study design employed."

Both Orest and Marovino use literature searches to support or change current practice. Orest is involved in a quality improvement project to critically appraise the literature to determine when patients should begin mobility after total knee or total hip replacement surgery. "We are looking at the literature to see if our current pathway is meeting the needs of our patients or whether we can improve functional outcomes or length of stay in the acute care environment."

Marovino practices in a small clinic where "we're always changing our practice based on new evidence." He describes literature searches as a routine part of his day. In addition to seeking information on the effectiveness of interventions or on the validity and reliability of tests and measures, he also searches for information to answer his patients' questions. He says, "Ultimately, we as clinicians are expected by payers, patients, and the clinical community to use tests and interventions that have an acceptable level of research support, so the levels-of-evidence framework should be applied and integrated into daily practice. The results can only lead to an improved quality of care."

Rothstein agrees. "I think you are hard pressed to tell payers that you should be reimbursed for an intervention if you have very little evidence to support it. On the other hand, if you have high-quality evidence to support it, you really should fight for the patient's right to get a treatment."

Rothstein also sees the levels of evidence as a way of improving the dialogue among physical therapists. "When people write letters to the Journal to discuss an article," he says, "they often cite opinions of 'experts' who have never published in the primary literature instead of citing a body of literature that supports their point! We need to make understanding the levels of evidence and the use of the best possible evidence part of our culture, in our discussions as well as in our clinical decision making."

Can every physical therapist incorporate the levels of evidence into their practice? Marovino thinks so. "As a profession, we are systematically curious; all evidence-based practice does is provide an organized method for that exploration," he says. "The whole point is to look for the best and most effective intervention, especially as new knowledge is developed," says Ciccone. "If clinicians can embrace that idea, they'll feel more comfortable with evidence-based practice and the levels of evidence."

Steven Glaros is Assistant Editor-Special Content for *Physical Therapy*. He can be reached at 800/999-2782 ext 3158, or steveglaros@apta.org.

References

- 1 Rothstein JM. Thirty years later...[editor's note]. *Phys Ther*. 2000;81:6-7.
- 2 Sackett DL, Straus SE, Richardson WS, et al. *Evidence-Based Medicine: How to Practice and Teach EBM*. 2nd ed. Edinburgh, United Kingdom: Churchill Livingstone Inc; 2000.
- 3 Guyatt G, Haynes B, Jaeschke R, et al. Introduction: the philosophy of evidence-based medicine. In: Guyatt G, Rennie D, eds. *Users' Guides to the Medical Literature: Essentials of Evidence-Based Clinical Practice*. Chicago, Ill: American Medical Association; 2002:5-20.
- 4 Guyatt G, Hayward R, Richardson WS, et al. Moving from evidence to action. In: Guyatt G, Rennie D, eds. *Users' Guides to the Medical Literature: Essentials of Evidence-Based Clinical Practice*. Chicago, Ill: American Medical Association; 2002:271-310.
- 5 Goldstein M, Scalzitti DA. Hooked on Evidence: adding an article to the database. Presentation at: PT2003: Annual Conference and Exposition of the American Physical Therapy Association; June 20, 2003; Washington, DC.
- 6 Design Type Definitions page. Hooked on Evidence Web site. Available at: www.apta.org/hookedonevidence/formsubmit/popups/designtypedef.cfm. Accessed July 9, 2003.
- 7 Guyatt G, Cook D, Devereaux PJ, et al. Therapy. In: Guyatt G, Rennie D, eds. *Users' Guides to the Medical Literature: Essentials of Evidence-Based Clinical Practice*. Chicago, Ill: American Medical Association; 2002:81-120.
- 8 Study Design page. Centre for Evidence-Based Medicine Web site. Available at: www.cebm.net/study_designs.asp. Accessed: June 24, 2003.
- 9 Levine M, Haslam D, Water S, et al. Harm. In: Guyatt G, Rennie D, eds. *Users' Guides to the Medical Literature: Essentials of Evidence-Based Clinical Practice*. Chicago, Ill: American Medical Association; 2002:121-153.
- 10 Levels of Evidence and Grades of Recommendation page. Centre for Evidence-Based Medicine Web site. Available at: www.cebm.net/levels_of_evidence.asp. Accessed: June 24, 2003.
- 11 McKibbon A, Hunt D, Richardson WS, et al. Finding the evidence. In: Guyatt G, Rennie D, eds. *Users' Guides to the Medical Literature: Essentials of Evidence-Based*

- Clinical Practice*. Chicago, Ill: American Medical Association; 2002:21-71.
- 12 Rothstein JM, Echternach JL. *Primer on Measurement: An Introductory Guide to Measurement Issues*. Alexandria, Va: American Physical Therapy Association; 1993.
- 13 McEwen I, ed. *Writing Case Reports: A How-to Manual for Clinicians*. 2nd ed. Alexandria, Va: American Physical Therapy Association; 2001.
- 14 Sackett DL. Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Chest*. 1986;89(suppl 1):2S-3S.

Visit [PT Bulletin Online](#)

[All contents © 2000 American Physical Therapy Association. All Rights Reserved.](#)