

## Chapter 2

---

# What Is Evidence?

*The most savage controversies are those about matters as to which there is no good evidence either way.*

—Bertrand Russell

### Objectives

Upon completion of this chapter the student/practitioner will be able to:

1. Discuss the concept of “best available clinical evidence.”
2. Describe the general content and procedural characteristics of desirable evidence and their implications for the selection of studies to evaluate.
3. Describe different forms of evidence and their uses for answering clinical questions in physical therapist practice.
4. Discuss and apply the principles and purposes of evidence hierarchies for each type of clinical question.
5. Discuss the limitations of evidence hierarchies and their implications for the use of evidence in practice.

### Terms in This Chapter

**Bias:** Results or inferences that systematically deviate from the truth “or the processes leading to such deviation.”<sup>1(p.251)</sup>

**Biologic plausibility:** The reasonable expectation that the human body could behave in the manner predicted.

**Case report:** A detailed description of the management of a patient/client that may serve as a basis for future research.<sup>2</sup>

**Clinical practice guideline:** “Systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific circumstances.”<sup>3(p.2)</sup>

**Cross-sectional study:** A study that collects data about a phenomenon during a single point in time or once within a defined time interval.<sup>4</sup>

**Effectiveness:** The extent to which an intervention or service produces a desired outcome under usual clinical conditions.<sup>1</sup>

**Efficacy:** The extent to which an intervention or service produces a desired outcome under ideal conditions.<sup>1</sup>

**Evidence:** “Any empirical observation about the apparent relation between events constitutes potential evidence.”<sup>5(p.6)</sup>

**Experimental design:** A research design in which the behavior of randomly assigned groups of subjects is measured following the purposeful manipulation of an independent variable(s) in at least one of the groups; used to examine cause-and-effect relationships between an independent variable(s) and an outcome(s).<sup>6,7</sup>

**Longitudinal study:** A study that looks at a phenomenon occurring over time.<sup>1</sup>

**Narrative review (also referred to as a *summary or literature review*):** A description of prior research without a systematic search and selection strategy or critical appraisal of the studies' merits.<sup>8</sup>

**Nonexperimental design (also referred to as an *observational study*):** A study in which controlled manipulation of the subjects is lacking<sup>6</sup>; in addition, if groups are present, assignment is predetermined based on naturally occurring subject characteristics or activities.<sup>4</sup>

**Peer review:** A process by which research is appraised by one or more content experts; commonly utilized when articles are submitted to journals for publication and when grant proposals are submitted for funding.<sup>1</sup>

**Physiologic study:** A study that focuses on the cellular or physiologic systems levels of the subjects; often performed in a laboratory.<sup>4</sup>

**Prospective design:** A research design that follows subjects forward over a specified period of time.

**Quasi-experimental design:** A research design in which there is only one subject group or in which randomization to more than one subject group is lacking; controlled manipulation of the subjects is preserved.<sup>9</sup>

**Randomized clinical trial (also referred to as a *randomized controlled trial* and a *randomized controlled clinical trial*) [RCT]:** A clinical study that uses a randomization process to assign subjects to either an experimental group(s) or a control (or comparison) group. Subjects in the experimental group receive the intervention or preventive measure of interest and then are compared to the subjects in the control (or comparison) group who did not receive the experimental manipulation.<sup>6</sup>

**Retrospective design:** A research design that uses historical (past) data from sources such as medical records, insurance claims, or outcomes databases.

**Single-system design:** A quasi-experimental research design in which one subject receives in an alternating fashion both the experimental and control (or comparison) condition.<sup>6</sup>

**Systematic review:** A method by which a collection of research is gathered and critically appraised in an effort to reach an unbiased conclusion about the cumulative weight of the evidence on a particular topic.<sup>4</sup>

## Introduction

Chapter 1 made the case that physical therapists should use *evidence* to inform their decision making during the patient/client management process. This claim raises the question “What qualifies as evidence?” Guyatt and Rennie’s statement “any empirical observation about the apparent relation between events constitutes potential evidence”<sup>5(p.6)</sup> suggests that a variety of types of evidence exist that may be integrated with clinical decisions.

Options may include, but are not limited to, published research articles, clinical practice guidelines, patient/client records, and recall of prior patient/client cases. The use of the modifier “best available clinical evidence” by Sackett et al., however, indicates that a method of prioritizing the evidence according to its merits is required to guide the clinician’s selection of relevant information.<sup>10</sup> This chapter discusses the forms and general characteristics of evidence available, as well as the hierarchies that have been developed to rank them.

## General Characteristics of Desirable Evidence

In light of the variety of evidence potentially available to physical therapists, it is helpful to have some general characteristics to consider during the initial search. Desirable attributes relate both to content as well as to procedural considerations that serve as preliminary indicators of quality.

The first content criterion pertains to the type of question a physical therapist wants to answer. The patient/client management elements of examination, diagnosis, prognosis, intervention (including preventive measures), and outcomes provide potential focus areas for evidence development and application. Ideally, the evidence located will address specifically the test, measure, classification system, prognostic factor, treatment technique, clinical prediction rule, or outcome that the physical therapist is considering relative to an individual patient/client.

The second content criterion pertains to the subjects studied. Desirable evidence includes subjects whose characteristics are similar to the patient/client in order to increase the therapist’s ability to apply the research findings to this individual person. Common attributes of interest may include, but are not limited to, the subjects’ diagnosis, stage of illness, duration of the problem(s), functional status, level of disability, age, gender, race, and clinical setting in which the patient/client management is occurring.

Two basic procedural characteristics have relevance in the evidence selection process as well. Whether or not a research article is peer reviewed is an important consideration. *Peer review* is the process by which research articles are evaluated by identified content experts to determine their merit for publication. Evaluation criteria usually include the credibility of the research in terms of its design and execution, relevance of the findings for the field and/or the specific journal, contribution to the body of knowledge about the topic, and, to a lesser degree, writing style.<sup>1</sup> The scrutiny of peer review provides an initial screening process to weed out lower quality research efforts.

The time of publication may be another procedural feature of interest given that articles often appear in journals a year or more after the completion of the research project.<sup>5</sup> Direct Internet publication undoubtedly has reduced this time line in many cases. Nevertheless, the rapid evolution of medical technology and pharmaceuticals continues to alter health care dramatically. As a result, older research may not reflect current patient management. A hypothetical example might be a 15-year-old study evaluating the effectiveness of an aerobic training program in patients with multiple sclerosis that has limited relevance now that multiple disease-modifying drugs are available.<sup>11</sup> However, studies should not be rejected

**TABLE 2-1 Four Desirable Characteristics of Research Identified During a Search for Evidence**

1. The study addresses the specific clinical question the physical therapist is trying to answer.
2. The subjects in the study have characteristics that are similar to the patient/client about whom the physical therapist has a clinical question.
3. The study was published in a peer-reviewed medium (paper, electronic).
4. The context of the study and/or the technique of interest are consistent with contemporary health care.

only because of their age if the techniques in question, and the context in which they were evaluated, have remained relatively unchanged since the research was conducted.

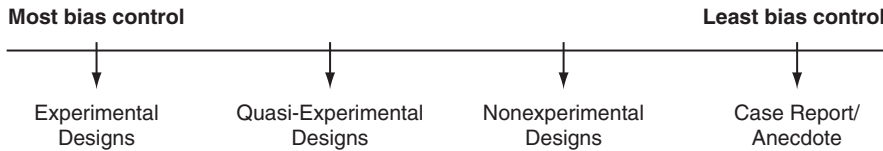
**Table 2-1** summarizes the four general characteristics of evidence that are preferable. Note that these attributes are labeled “desirable,” not “mandatory.” This word choice is purposeful because there is much work to be done to expand the depth and breadth of physical therapy research. Many of the clinical questions physical therapists have about their patients/clients have not been explored or have been addressed in a limited fashion. A search for the “best available clinical evidence” may result in the identification of studies that are not peer reviewed or do not include subjects that look like a therapist’s individual patient/client. Similarly, studies may not exist that include a test or technique of interest in the clinical setting. The evidence-based physical therapist practice challenge is to decide how best to use evidence that is limited in these ways when it is the only evidence available.

## Forms of Evidence

As noted previously, forms of evidence may include anything from published research to patient records and clinical recall. Evidence-based practice in health care emphasizes the use of research to inform clinical decisions because of its potential to provide objective, unbiased results. A variety of research design options exist, many of the details of which are discussed in Chapter 5. A key point is that different research designs are suited to answering different types of clinical questions therapists may have about their patients/ clients. The usefulness of a diagnostic test must be evaluated with methods that are different than those used to determine whether an intervention works. As a result, therapists should anticipate looking for evidence with different research designs depending on what they want to know. The remainder of this chapter provides highlights of these different designs and their relative merits.

### Research Designs: Overview

Forms of evidence fall along a continuum that is dictated by the presence and strength of a research design (**Figure 2-1**). At one end of the continuum is research that attempts



**Figure 2-1** Continuum of bias control in research designs.

to impose maximum control within the design in order to reduce the chance that bias will influence the study's results. *Bias* is a systematic deviation from the truth that occurs as a result of uncontrolled (and unwanted) influences during the study.<sup>1</sup> Various authors refer to research designs with the best features to minimize bias as *randomized clinical trials*, *randomized controlled trials*, or *randomized controlled clinical trials*.<sup>1,4,6</sup> The acronym used for all three is “RCT.” These studies also are categorized as *experimental designs*. Irrespective of the label, the researchers' intention is the same: to reduce unwanted influences in the study through randomization of study participants to two or more groups and through controlled manipulation of the experimental intervention. A variant of this approach is the *single-system design* in which only one person is studied who receives, on an alternating basis, both the experimental and control (or comparison) conditions.<sup>6</sup>

An RCT or single-system design is best suited to answer questions about whether an experimental intervention has an effect and whether that effect is beneficial or harmful to the subjects. When conducted under ideal conditions—that is, when a high degree of control is achieved—these studies are focused on treatment *efficacy*. An example might be a study in which individual subjects with traumatic brain injuries are randomized to an experimental balance-training program that is performed in a quiet research laboratory. Such an environment is free of distractions that may interfere with the subjects' ability to pay attention to directions and focus on the required activities. Alternatively, if the same subjects perform the experimental balance-training program during their regular physical therapy appointment in the outpatient rehabilitation center, then the RCT is focused on treatment *effectiveness*.<sup>12</sup> Investigators in this version of the study want to know if the balance program works in a natural clinical environment full of noise and activity.

Randomized clinical trials and single-system designs are approaches used to conduct an original research project focusing on one or more persons. These individual studies themselves may serve as the focus of another type of controlled research design referred to as a systematic review. *Systematic reviews* synthesize original evidence that has been selected and critically appraised according to preestablished criteria.<sup>4</sup> The goal of this research design is to draw conclusions from the cumulative weight of studies that, individually, may not provide enough evidence to provide a definitive answer. The preestablished criteria are used to minimize bias that may be introduced when investigators make decisions about which prior studies to include and when judgments are made about their quality. Systematic reviews may address any type of clinical question; however, most commonly they focus on well-controlled studies of interventions—in other words, on RCTs.

At the other end of the evidence continuum is the unsystematic collection of patient/client data that occurs in daily physical therapist practice. The term *unsystematic* is not meant to imply substandard care; rather, it is an indication that clinical practice is focused on the individual patient/client rather than on groups of subjects on whom controls are imposed for the purposes of ensuring research integrity. This type of evidence often is labeled “anecdotal”<sup>9</sup> and frequently is put to use when therapists recall from memory prior experiences with patients/clients similar to the person with whom they are currently dealing. In response to regulatory and reimbursement pressures, many clinical settings are creating a degree of consistency in data collection with their implementation of standardized assessment and outcomes instruments, electronic health records, and databases to capture patient/client outcomes. As a result, physical therapists working in these settings may find some evidence that is useful to inform their practice.

In between the two ends of the evidence continuum are study designs that lack one or more of the following characteristics:

1. Randomization techniques to distribute subjects into groups;
2. The use of more than one group in order to make a comparison;
3. Controlled experimental manipulation of the subjects;
4. Measures at the patient/client level (e.g., impairment in body functions and structure, activity limitations, participation restrictions); and/or
5. A systematic method for collecting and analyzing information.

These designs have fewer features with which to minimize bias and/or shift their focus away from patient/client-centered outcomes. For example, *quasi-experimental designs* maintain the purposeful manipulation of the experimental technique, but they may not randomize subjects to groups or may have only one subject group to evaluate.<sup>9</sup> *Nonexperimental (or observational) designs* have even less control than quasi-experimental studies because they have the same limitations with respect to their group(s) and they do not include experimental manipulation of subjects.<sup>6</sup> In spite of their less rigorous designs, both quasi-experimental and nonexperimental studies are used to evaluate the effectiveness of interventions, often due to ethical or pragmatic reasons related to the use of patients in research. In addition, observational designs are used to answer questions about diagnostic tests, clinical measures, prognostic indicators, clinical prediction rules, and patient/client outcomes.

Below quasi-experimental and nonexperimental designs on the continuum are research efforts that focus only on cellular, anatomic, or physiologic systems. These studies often have a high degree of control because they are grounded in the scientific method that is the hallmark of good bench research. They are lower on the continuum not because of their potential for bias, but because they do not focus on person-level function. For this reason they are referred to as *physiologic studies*.<sup>5</sup>

Even lower on the continuum are case reports and narrative reviews. These study approaches have different purposes. *Case reports* simply describe what occurred with a

patient/client, whereas *narrative reviews* summarize prior research.<sup>2,8</sup> In spite of these differences, these designs have one common element that puts them both at the bottom of the continuum: they lack a systematic approach to the issue or topic of interest. It is important to note, however, that the content of a case report or narrative review may provide a stimulus to conduct a more rigorous research project. **Table 2-2** provides a list of citations from physical therapy literature that represent each type of study design described here.

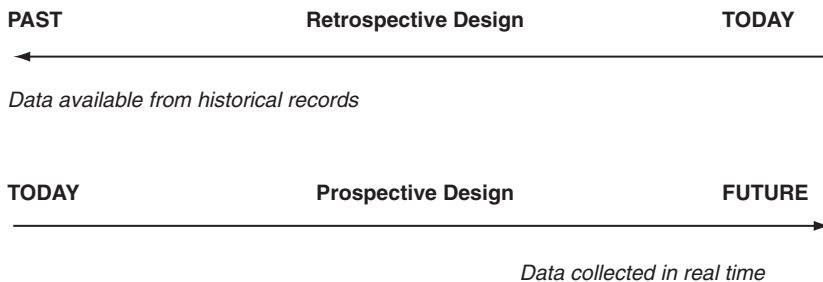
**TABLE 2-2 Citations from Physical Therapy Research Illustrating Different Study Designs**

Study Design	Citation
Systematic Review	Harvey LA, Brosseau L, Herbert RD. Continuous passive motion following total knee arthroplasty in people with arthritis. <i>Cochrane Database Syst Rev</i> . 2010;3: Art. No.: CD004260. DOI: 10.1002/14651858.CD004260.pub2.
Randomized Clinical Trial	Johansson KM, Adolfsson LE, Foldevi MOM. Effects of acupuncture versus ultrasound in patients with impingement syndrome: Randomized clinical trial. <i>Phys Ther</i> . 2005;85: 490–501.
Single-System Design	Carr S, Fairleigh A, Backman C. Use of continuous passive motion to increase hand range of motion in a woman with scleroderma: a single-subject study. <i>Physiother Can</i> . 1997;49(4):292–296.
Quasi-Experimental Study	Kileff J, Ashburn A. A pilot study of the effect of aerobic exercise on people with moderate disability multiple sclerosis. <i>Clin Rehabil</i> . 2005;19(2):165–169.
Observational Study	Kirk-Sanchez NJ. Factors related to activity limitations in a group of Cuban Americans before and after hip fracture. <i>Phys Ther</i> . 2004;84:408–418.
Physiologic Study	DeSimone NA, Christiansen C, Dore D. Bactericidal effect of 0.95-mW helium-neon and 5-mW indium-gallium-aluminum-phosphate laser irradiation at exposure times of 30, 60, and 120 seconds on photosensitized <i>Staphylococcus aureus</i> and <i>Pseudomonas aeruginosa</i> in vitro. <i>Phys Ther</i> . 1999;79:839–846.
Case Report	Shrader JA, Siegel KL. Nonoperative management of functional hallux limitus in a patient with rheumatoid arthritis. <i>Phys Ther</i> . 2003;83:831–843.
Summary	Ciesla ND. Chest physical therapy for patients in the intensive care unit. <i>Phys Ther</i> . 1996;76:609–625.

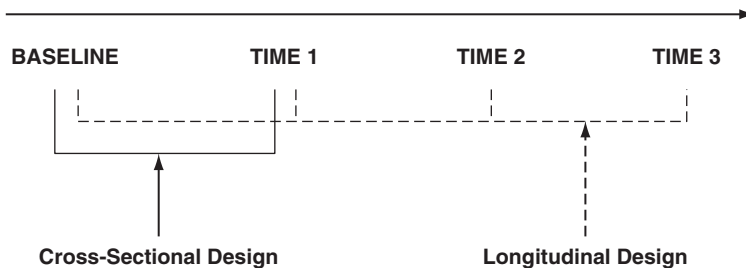
## Research Designs: Timing

Research designs also may be categorized according to the time line used in the study. For example, physical therapy researchers may want to know the relationship between the number of visits to an outpatient orthopedic clinic and the workers' compensation insurance status of patients treated over a 3-year period. Such a question may be answered through an analysis of 3 years of patient records from the clinic. This *retrospective design* has as an opposite form—a *prospective design*—in which the investigators collect information from new patients that are admitted to the clinic. As **Figure 2-2** illustrates, a retrospective approach takes advantage of data that already exist, whereas a prospective approach requires that new data be collected in real time.

In a similar fashion, researchers may be interested in a single point in time or a limited time interval (e.g., *cross-sectional study*), or they may wish to study a phenomenon over an extended period of time (e.g., *longitudinal study*). In the cross-sectional approach, investigators may have an interest in the functional outcome at discharge (a single point in time) from the hospital of patients receiving physical therapy following total hip replacement. In contrast, a longitudinal approach would include follow-up of these patients to assess outcomes at discharge and at a specified point or points in time in the future (e.g., 3 months, 6 months, 1 year). **Figure 2-3** illustrates these design options.



**Figure 2-2** Graphic depiction of retrospective and prospective research designs.



**Figure 2-3** Graphic depiction of cross-sectional and longitudinal research designs.

The sequence of events across time in a study is important, particularly when an investigator is trying to determine whether a change in the patient/client's condition was the direct result of the intervention or preventive measure applied. Specifically, the intervention must have occurred before the outcome was measured to increase one's confidence that it was the technique of interest that made a difference in the subjects.

### Research Designs: What Is the Question?

Remember that the clinical question the physical therapist wants to answer will determine which of these forms of evidence to seek. For example, a question about the best test to identify a rotator cuff tear (diagnosis) is likely to be addressed by a cross-sectional nonexperimental study of patients that are suspected to have the problem based on clinical exam. However, a question about risk factors for falls in the elderly may be answered in one of two ways: (1) a longitudinal study in which two groups of elderly subjects are followed in real time (e.g., prospectively) to determine who falls and who does not, or (2) a retrospective study that starts with subjects with documented falls and evaluates possible precipitating characteristics (e.g., visual deficits) in comparison to non-fallers. Finally, a question about the effectiveness of joint mobilization in the management of neck pain is best answered by a prospective RCT of patients classified with neck pain. Physical therapists should anticipate these differences when planning their search strategies to increase the efficiency of the process.

One must also recall that a search for the “best available clinical evidence” may result in the discovery of research that is limited in content and/or quality. In other words, the current state of knowledge in an area may be such that the best (and only) evidence available is from studies in which the chance of bias is higher because of weaknesses in the research designs. Physical therapists will find this scenario to be true for many of the clinical questions they pose in practice. This reality is not a reason to reject evidence-based physical therapist practice; rather, it is a reaffirmation that clinical judgment and expertise are required to decide how to use evidence that is limited in form.

## Hierarchies of Evidence

Previous research has identified a number of barriers to using evidence in physical therapist practice, one of which is the lack of time available to search for, select, and read professional literature.<sup>13</sup> The selection process may be eased somewhat by ranking research designs based on their ability to minimize bias. Proponents of evidence-based medicine have attempted to make the study selection process easier by developing hierarchies, or levels, of evidence.<sup>14</sup> Guyatt and Rennie have focused on a hierarchy for studies about treatment<sup>5</sup>; Sackett et al.<sup>15(p.6)</sup> and the Centre for Evidence-Based Medicine in Oxford, England,<sup>16</sup> have developed separate hierarchies for evidence about diagnostic tests, prognostic indicators, treatment, and economic and decision analyses. The variety of hierarchies is necessary because of the point made previously: different research designs are required to answer different types of clinical questions. Understanding the nuances of each hierarchy is an important skill to develop in order to use them appropriately.

The remainder of this chapter focuses on the hierarchies adapted from the Centre for Evidence-Based Medicine Oxford Web site.<sup>16</sup> **Table 2-3** depicts the hierarchy for intervention papers. All of the hierarchies are included in Appendix A.

These ranking schemes are similar to one another in that they place systematic reviews at the top of each list. Systematic reviews are valued because they may produce conclusions based on a critical appraisal of a number of individual studies that have been selected according to preestablished criteria. Ideally, the studies reviewed have research designs that minimize the chance of bias (e.g., “high-quality evidence”), are pertinent to the therapist’s question, and provide a more definitive answer to the question. This ideal is akin to the “holy grail” in evidence-based practice; however, systematic reviews also have their limitations, as discussed in Chapter 16.

At the other end of each hierarchy are several forms of evidence that are least desirable because of their potential for bias or because of their lack of focus on patient/client level information, including:

- Expert opinion without critical appraisal;
- Anecdotal evidence;
- Physiologic studies; and
- Studies based only on *biologic plausibility*.

Clinicians who locate studies that fall into this level of evidence must identify and consider the implications and limitations for each study type when deciding whether to use the information with a patient/client. Despite their limitations, however, it is worth noting that these study types are included in the hierarchies because they constitute evidence.

Details for each level between these end points on the hierarchies vary because of the types of questions being addressed; however, some additional common themes can be identified. First, level of rank depends on the strength of the study design. For example, an RCT is highly ranked because it is a more rigorous research design than an observational study for investigation of the therapeutic effects of joint mobilization in patients with neck pain. Second, individual studies with strong designs are ranked more highly than systematic reviews of studies with weaker designs. For example, a single prospective study of fall risk in the elderly that includes a comprehensive list of predisposing factors for falls is more valuable than a systematic review of retrospective studies that failed to include medications, living environment, and mental status as potential contributors to fall risk. Third, systematic reviews of studies with similar directions and degrees of results (e.g., subjects improved in most studies) are ranked higher as a result of this homogeneity than systematic reviews of studies with significant variation in their individual findings (e.g., subjects improved in some studies and not others). **Figure 2-4** summarizes the commonalities among evidence hierarchies.

Selection of studies through the use of hierarchies may improve the efficiency of the search process for busy clinicians. These schemas also are used regularly to grade evidence to facilitate the decision-making process about which information to use. This strategy is

TABLE 2-3 Hierarchy of Evidence for Articles About Therapy

Level	Therapy/Prevention, Aetiology/Harm
1a	<ul style="list-style-type: none"> <li>• <b>Systematic Review</b> of Randomized Clinical Trials that <i>do not</i> have a statistically significant variation in the direction or degrees of results.</li> </ul>
1b	<ul style="list-style-type: none"> <li>• <b>Individual</b> Randomized Clinical Trial with a narrow Confidence Interval.</li> </ul>
1c	<ul style="list-style-type: none"> <li>• All or none study.<sup>a</sup></li> </ul>
2a	<ul style="list-style-type: none"> <li>• <b>Systematic Review</b> of cohort studies<sup>b</sup> that <i>do not</i> have a statistically significant variation in the direction or degrees of results.</li> </ul>
2b	<ul style="list-style-type: none"> <li>• <b>Individual</b> cohort study (including low-quality Randomized Clinical Trial; e.g., &lt;80% subject follow-up).</li> </ul>
2c	<ul style="list-style-type: none"> <li>• “Outcomes” Research.<sup>c</sup></li> </ul>
3a	<ul style="list-style-type: none"> <li>• <b>Systematic Review</b> of case-control studies<sup>d</sup> that <i>do not</i> have a statistically significant variation in the direction or degrees of results.</li> </ul>
3b	<ul style="list-style-type: none"> <li>• <b>Individual</b> Case-Control Study.</li> </ul>
4	<ul style="list-style-type: none"> <li>• Case-series study.<sup>e</sup></li> <li>• Cohort or case-control study that did not: <ul style="list-style-type: none"> <li>◦ define comparison groups adequately; or,</li> <li>◦ did not measure exposures and outcomes objectively or in a blinded fashion; or,</li> <li>◦ control for confounders; or,</li> <li>◦ have sufficient follow-up (cohort studies only).</li> </ul> </li> </ul>
5	<ul style="list-style-type: none"> <li>• Expert opinion without explicit critical appraisal, or based on physiology, bench research or “first principles.”<sup>f</sup></li> </ul>

<sup>a</sup>All or none: A study in which some or all patients died before treatment became available, now none die.

<sup>b</sup>Cohort Study: In intervention papers, a prospective research design used to evaluate the relationship between a treatment and an outcome; two groups of subjects—one of which receives the intervention and one of which does not—are monitored over time to determine who develops the outcome and who does not. This label could be applied to quasi-experimental and nonexperimental research designs in which two non-randomized groups of subjects are evaluated.

<sup>c</sup>Outcomes Research: Nonexperimental research that evaluates outcomes of care in “real-world” clinical conditions.

<sup>d</sup>Case-Control Study: A retrospective epidemiologic research design used to evaluate the relationship between a potential risk factor and a disease or disorder; two groups of subjects—one of which has the disease/disorder (the *case*) and one which does not (the *control*)—are compared to determine which group has a greater proportion of individuals with the risk factor.

<sup>e</sup>Case Series: a description of the management of several patients/clients for the same purposes as a case report; the use of multiple individuals increases the potential importance of the observations as the basis for future research.

<sup>f</sup>First Principles: Biologically plausible rationales for management of pathophysiology.

Source: Adapted with permission from Oxford Centre for Evidence-Based Medicine; www.cebm.net.



**Figure 2-4** General ranking of evidence within hierarchies.

most apparent in published *clinical practice guidelines*. National and international government agencies and professional associations produce guidelines in an effort to promote effective and efficient health care. A few examples relevant to physical therapy include the following:

- VA/DoD Clinical Practice Guideline for Rehabilitation of Lower Limb Amputation (2007)<sup>17</sup>;
- American College of Chest Physicians and American Association of Cardiovascular and Pulmonary Rehabilitation’s “Pulmonary Rehabilitation: Joint ACCP/AACVPR Evidence-Based Clinical Practice Guidelines” (2007)<sup>18</sup>; and
- The Ottawa Panel Evidence-Based Clinical Practice Guidelines for Aerobic Fitness Exercises in the Management of Fibromyalgia: Part I (2008).<sup>19</sup>

Each of these documents, as well as numerous other similar publications, contain recommendations based on a review and ranking of available evidence. Grading schemes are described in the guidelines and used to qualify the recommendations made.

For example, the Department of Veterans Affairs/Department of Defense (VA/DoD) assigned the letter grades A, B, C, D, and I based on the quality of the evidence and the net benefit of the intervention recommended. The letter “A” is the highest designation and signifies that a recommendation in the guideline is supported by the following statement: “Good evidence was found [at least one properly done RCT] that the intervention improves important health outcomes and concludes that benefits substantially outweigh harm.”<sup>17</sup>(pp.116–117) The letter “B” indicates that recommendations are supported by “fair evidence” that benefit outweighs harm, and the letter “C” indicates a neutral position based on “fair evidence” that does not clearly indicate benefits outweigh harm. Finally, the letter “I” indicates insufficient or “poor evidence” (e.g., expert opinion, descriptive studies, case reports) on which to base a recommendation.

In theory, physical therapists using any of these guidelines could go straight to the recommendations and make decisions about how to change their practice based on these evidence grades. However, there are several limitations to these levels that should be recognized before a clinician blindly adopts the practice behaviors addressed in the guidelines.

## Limitations of Evidence Hierarchies

In 2002, the Agency for Healthcare Research and Quality (formerly known as the Agency for Healthcare Policy and Research) published an evidence report entitled “Systems to Rate the Strength of Scientific Evidence.”<sup>14</sup> The authors of this report performed an extensive literature review to identify quality assessment methods used to assess the strength of evidence for systematic reviews and meta-analyses, RCTs, observational studies, and diagnostic studies, as well as methods for evaluating the strength of an entire body of evidence on a particular topic. In addition, they examined evidence evaluation methods used by agency-sponsored Evidence-Based Practice Centers and other organizations focused on evidence-based medicine, such as the Cochrane Collaboration.

Of the 121 systems reviewed, only 26 fully addressed quality criteria established by the authors for each type of study. Many of these lengthy systems required an inconvenient amount of time to complete. Also noted was the greater number of quality assessment methods for RCTs as compared with other types of research. The other 95 assessment methods that the authors reviewed were limited in the quality domains addressed, by a “one-size-fits-all” approach that did not distinguish among critical features of different study designs, or by lack of validation. Few of the methods had been tested for reliability. The take-home message from this report is that the strength of evidence depends, in part, on the scale against which it is being rated. In response to the potential misuse of evidence grading systems, Glasziou et al. suggested that quality ratings or scales should address different types of research and would be improved by the addition of qualitative statements, as well as details regarding ratings criteria.<sup>20</sup>

Understanding the details and bases for evidence hierarchies will help physical therapists select evidence to answer clinical questions about patients/clients. However, a hierarchy is only a tool to facilitate the process; it should not be used to make a final judgment about a study’s value and relevance. Physical therapists must still read and critically appraise the evidence they find before incorporating any results into their clinical decisions. This point is emphasized by an ongoing debate about the relative merits of RCTs versus quasi-experimental and observational studies. Some evidence indicates that the bias in the latter study designs results in overestimations of treatment effects, whereas other authors have reported that none of the study designs consistently estimate an intervention’s impact.<sup>21–23</sup> As noted in Chapter 1, clinical judgment and expertise are essential to evidence-based physical therapy. The variability in research quality requires that physical therapists use their knowledge and skills to determine whether the evidence they find, no matter how high or low on a hierarchy, is useful for an individual patient/client.

## Summary

Evidence-based physical therapist practice requires clinicians to select the “best available evidence” from studies whose quality depends on their relevance to the question asked, their timeliness, and the level of prior scrutiny of their merits, as well as on their research

design and execution. Evidence hierarchies may facilitate study selection because of the ranking structure they create based on important research attributes. Different hierarchies have been designed to address evidence about diagnostic tests, prognostic indicators, and interventions. Producers of clinical practice guidelines also have defined various levels of evidence to demonstrate the degree to which their recommendations are supported by research. No matter what form a hierarchy takes, it is only a tool to facilitate the process; it should not be used to make a final judgment about a study's value and relevance. Physical therapists must still read and critically appraise the evidence they find before incorporating any results into their clinical decisions.

## Exercises

1. What does the phrase “best available clinical evidence” mean with respect to a physical therapist's selection and use of studies?
2. Discuss the differences between a randomized controlled trial and an observational study. Under which circumstances might each study design be appropriate? Provide examples of each relevant to physical therapy to illustrate your points.
3. Discuss the difference between a retrospective and prospective research design and give an example of each that reflects a study question relevant to physical therapy.
4. Discuss the difference between a cross-sectional and longitudinal research design and give an example of each that reflects a study question relevant to physical therapy.
5. Describe the common organizational characteristics of evidence hierarchies.
6. Discuss the rationale behind the creation of different hierarchies for evidence about diagnostic tests, prognostic factors, and interventions.
7. Discuss the limitations of evidence hierarchies. Why is a hierarchy only a starting point in evidence-based physical therapist practice?

## References

1. Helewa A, Walker JM. *Critical Evaluation of Research in Physical Rehabilitation: Towards Evidence-Based Practice*. Philadelphia, PA: W.B. Saunders; 2000.
2. McEwen I. *Writing Case Reports: A How-to Manual for Clinicians*. 3d ed. Alexandria, VA: American Physical Therapy Association; 2009.
3. Lohr KN, Field MJ. A provisional instrument for assessing clinical practice guidelines. In: Field MJ, Lohr KN, eds. *Guidelines for Clinical Practice: From Development to Use*. Washington, DC: National Academy Press; 1992.
4. Straus SE, Richardson WS, Glaziou P, Haynes RB. *Evidence-Based Medicine: How to Practice and Teach EBM*. 3d ed. Edinburgh, Scotland: Elsevier Churchill Livingstone; 2005.

5. Guyatt G, Rennie D. *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice*. Chicago, IL: AMA Press; 2002.
6. Domholdt E. *Rehabilitation Research: Principles and Applications*. 3d ed. St Louis, MO: Elsevier Saunders; 2005.
7. Campbell DT, Stanley JC. *Experimental and Quasi-Experimental Designs for Research*. Boston, MA: Houghton Mifflin; 1963.
8. Herbert R, Jamtvedt G, Mead J, Hagen KB. *Practical Evidence-Based Physical Therapy*. Edinburgh, Scotland: Elsevier Butterworth Heinemann; 2005.
9. Cook TD, Campbell DT. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Boston, MA: Houghton Mifflin; 1979.
10. Sackett DL, Rosenberg WMC, Gray JAM, et al. Evidence-based medicine: what it is and what it isn't. *BMJ*. 1996;312(7023):71-72.
11. The MS Disease-Modifying Drugs. National Multiple Sclerosis Society Web site. Available at: [www.nationalmssociety.org/download.aspx?id=45](http://www.nationalmssociety.org/download.aspx?id=45). Accessed May 20, 2010.
12. Winstein CJ, Lewthwaite R. *Efficacy and Effectiveness: Issues for Physical Therapy Practice and Research. Examples from PTClinResNet*. Eugene Michels Forum. Combined Sections Meeting. American Physical Therapy Association. 2004. Available at: <http://pt.usc.edu/clinresnet/CSM04/Eugene%20Michels%20pdf%20files/Winstein%20EM%202004.pdf>. Accessed April 20, 2010.
13. Jette DU, Bacon K, Batty C, et al. Evidence-based practice: beliefs, attitudes, knowledge, and behaviors of physical therapists. *Phys Ther*. 2003;83(9):786-805.
14. West S, King V, Carey TS, et al. *Systems to Rate the Strength of Scientific Evidence. Evidence Report/Technology Assessment Number 47* (Prepared by the Research Triangle Institute–University of North Carolina Evidence-Based Practice Center under Contract No. 290-97-0011.). AHRQ Publication No. 02-E016. Rockville, MD: Agency for Health Care Research and Quality; April 2002.
15. Sackett DL, Straus SE, Richardson WS, et al. *Evidence-Based Medicine: How to Practice and Teach EBM*. 2d ed. Edinburgh, Scotland: Churchill Livingstone; 2000.
16. Levels of Evidence. Centre for Evidence-Based Medicine: Oxford Web site. Available at: <http://www.cebm.net>. Accessed April 20, 2010.
17. Department of Veterans Affairs, Department of Defense. VA/DoD clinical practice guideline for rehabilitation of lower amputation. Washington, DC: Department of Veterans Affairs, Department of Defense; 2007.
18. Ries AL, Bauldoff GS, Carlin BW, et al. Pulmonary rehabilitation: Joint ACCP/AACVPR Evidence-Based Clinical Guidelines. *Chest*. 2007;131(5 suppl):4S-42S.
19. Brousseau L, Wells GA, Tugwell P, et al. Ottawa Panel Evidence-Based Clinical Practice Guidelines for Aerobic Fitness Exercises in the Management of Fibromyalgia—Part I. *Phys Ther*. 2008;88(7):857-871.
20. Glasziou P, Vandenbroucke J, Chalmer I. Assessing the quality of research. *BMJ*. 2004;328(7430):39-41.
21. Britton A, McKee M, Black N, et al. Choosing between randomized and non-randomised studies: a systematic review. *Health Technol Assess*. 1998;2(13):i-iv,1-124.
22. MacLehose RR, Reeves BC, Harvey IM, et al. A systematic review of comparisons of effect sizes derived from randomised and non-randomised studies. *Health Technol Assess*. 2000;4(34):1-154.
23. Glasziou P, Chalmers I, Rawlins M, McCulloch P. When are randomised trials unnecessary? Picking signal from noise. *BMJ*. 2007;334(7589):349-351.

