CHAPTER 2

WHAT IS EVIDENCE?

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.”

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.

Clinical practice guidelines: “Systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific circumstances,” also referred to as “summaries.”

Cross-sectional study: A study that collects data about a phenomenon during a single point in time or once within a defined time interval.

Effectiveness: The extent to which an intervention or service produces a desired outcome under usual clinical conditions.
Efficacy: The extent to which an intervention or service produces a desired outcome under ideal conditions.1

Evidence: "Any empirical observation about the apparent relation between events constitutes potential evidence."6(p.6)

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).7,8

Longitudinal study: A study that looks at a phenomenon occurring over time.1

Narrative review (also referred to as a literature review): A description of prior research without a systematic search and selection strategy or critical appraisal of the studies’ merits.9

Nonexperimental design (also referred to as an observational study): A study in which controlled manipulation of the subjects is lacking; in addition, if groups are present, assignment is predetermined based on naturally occurring subject characteristics or activities.5

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.1

Physiologic study: A study that focuses on the cellular or physiologic systems levels of the subjects; often performed in a laboratory.5

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.10

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.7

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.7

Synopsis: "A succinct description of selected individual studies or systematic reviews."4

Systematic review: A method by which a collection of individual research studies is gathered and critically appraised in an effort to reach an unbiased conclusion about the cumulative weight of the evidence on a particular topic; also referred to as "syntheses."4

Systems: "Individual patient characteristics are automatically linked to the current best evidence that matches the patient’s specific circumstances and the clinician is provided with key aspects of management (e.g., computerised decision support systems)."4
Introduction

The case has been made 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” 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. 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 a research article has been 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. 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. 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. However, studies should not be rejected 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 research related to physical therapist practice. 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. 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 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. 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. 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.7

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.13 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, or “syntheses,” are composed of original evidence that has been selected and critically appraised according to preestablished criteria.5 The goal of this research design is to draw conclusions from the cumulative weight of studies that, individually, may not be sufficient 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”10 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.\textsuperscript{10} 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.\textsuperscript{7} 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.\textsuperscript{6}

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.\textsuperscript{2,9} 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. \textbf{Table 2-2} provides a list of citations from physical therapy literature that represent each type of study design described here.

\textbf{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 previous patient records from the clinic. This \textit{retrospective design} has as an opposite form—a \textit{prospective design}—in which the investigators collect information from new patients that are admitted to the clinic. As \textbf{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., \textit{cross-sectional study}), or they may wish to study a phenomenon over an extended period of time (e.g., \textit{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
TABLE 2-2 Citations from Physical Therapy Research Illustrating Different Study Designs

<table>
<thead>
<tr>
<th>Study Design</th>
<th>Citation</th>
</tr>
</thead>
</table>
  Cochrane Database Syst Rev.  
| Randomized Clinical Trial| Miyamoto GC, et al. Efficacy of the addition of modified Pilates exercises to a minimal intervention in patients with chronic low back pain: a randomized controlled trial.  
  Phys Ther.  
  Phys Ther.  
  Clin Rehabil.  
  Phys Ther.  
| Physiologic Study        | DeSimone NA, et al. 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 Staphylococcus aureus and Pseudomonas aeruginosa in vitro.  
  Phys Ther.  
  Phys Ther.  
| Summary                  | Ciesla ND. Chest physical therapy for patients in the intensive care unit.  
  Phys Ther.  

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.

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
A cuff tear (diagnosis) is likely to be addressed by a cross-sectional nonexperimental study of patients who 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. 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 more efficient for busy clinicians by developing hierarchies, or levels, of evidence. In 2011, Howick et al. at the Oxford Centre for Evidence-Based Medicine (OCEBM) in England consolidated previously developed individual hierarchies for evidence about diagnostic tests, prognostic indicators, and treatment techniques into one reference table for easier use (reprinted in Figure 2-4). 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.

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 a result, individual studies may provide stronger evidence in answer to a clinical question. At the lowest end of each OCEBM hierarchy are physiologic studies and research based on biologic plausibility. These studies are classified as such because of their focus on the mechanisms underlying a pathologic condition or treatment technique. Clinicians who locate studies that fall into this level of evidence must consider the extent to which they can reasonably apply the findings at the person level for an individual patient.

Details about each evidence level between these end points vary because of the types of questions being addressed; however, some common themes regarding use of the hierarchies 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 an intervention such as joint mobilization in patients with neck pain. Second, individual studies with strong designs should be “graded up” above 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 in a large number of subjects 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) make a stronger case 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 2-5 summarizes the commonalities among the OCEBM evidence hierarchies. Howick et al. also acknowledged some preliminary findings regarding the potential value of individual patient cases
### FIGURE 2-4  Oxford Centre for Evidence-Based Medicine 2011 Levels of Evidence.

<table>
<thead>
<tr>
<th>Question</th>
<th>Step 1 (Level 1)</th>
<th>Step 2 (Level 2)</th>
<th>Step 3 (Level 3)</th>
<th>Step 4 (Level 4)</th>
<th>Step 5 (Level 5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>How common is the problem?</td>
<td>Local and current random sample surveys (or censuses)</td>
<td>Systematic review of surveys that allow matching to local circumstances</td>
<td>Local non-random sample</td>
<td>Case-series</td>
<td>n/a</td>
</tr>
<tr>
<td>Is this diagnostic or monitoring test accurate? (Diagnosis)</td>
<td>Systematic review of cross sectional studies with consistently applied reference standard and binding</td>
<td>Individual cross sectional studies with consistently applied reference standard and binding</td>
<td>Non-consecutive studies, or studies without consistently applied reference standards</td>
<td>Case-control studies, or “poor or non-independent reference standard”</td>
<td>Mechanism-based reasoning</td>
</tr>
<tr>
<td>What will happen if we do not add a therapy? (Prognosis)</td>
<td>Systematic review of inception cohort studies</td>
<td>Inception cohort studies</td>
<td>Cohort study or control arm of randomized trial</td>
<td>Case-series or case-control studies, or poor quality prognostic cohort study</td>
<td>n/a</td>
</tr>
<tr>
<td>Does this intervention help? (Treatment Benefits)</td>
<td>Systematic review of randomized trials or n-of-1 trials</td>
<td>Randomized trial or observational study with dramatic effect</td>
<td>Non-randomized controlled cohort/follow-up study</td>
<td>Case-series, case-control studies, or historically controlled studies</td>
<td>Mechanism-based reasoning</td>
</tr>
<tr>
<td>What are the COMMON harms? (Treatment Harms)</td>
<td>Systematic review of randomized trials, systematic review of nested case-control studies, n-of-1 trial with the patient you are raising the question about, or observational study with dramatic effect</td>
<td>Individual randomized trial or (exceptionally) observational study with dramatic effect</td>
<td>Non-randomized controlled cohort/follow-up study</td>
<td>Case-series, case-control, or historically controlled studies</td>
<td>Mechanism-based reasoning</td>
</tr>
<tr>
<td>What are the RARE harms? (Treatment Harms)</td>
<td>Systematic review of randomized trials or n-of-1 trial</td>
<td>Randomized trial or (exceptionally) observational study with dramatic effect</td>
<td>Non-randomized controlled cohort/follow-up study</td>
<td>Case-series, case-control, or historically controlled studies</td>
<td>Mechanism-based reasoning</td>
</tr>
<tr>
<td>Is this (early detection) test worthwhile? (Screening)</td>
<td>Systematic review of randomized trials</td>
<td>Randomized trial</td>
<td>Randomized controlled cohort/follow-up study</td>
<td>Case-series, case-control, or historically controlled studies</td>
<td>Mechanism-based reasoning</td>
</tr>
</tbody>
</table>

*Level may be graded down on the basis of study quality, imprecision, indirectness (study PICO does not match questions PICO), because of inconsistency between studies, or because the absolute effect size is very small; Level may be graded up if there is a large or very large effect size.

**As always, a systematic review is generally better than an individual study.

Source: Adapted with permission from Oxford Center for Evidence-Based Medicine—www.cebm.net.
and anecdotes in evidence-based decision making. Although not included in the latest edition of the OCEBM hierarchies, these sources of information are reflected in the figure here to illustrate their rank relative to planned, systematic research efforts.

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 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 therapist practice include the following:

- VA/DoD Clinical Practice Guideline for Rehabilitation of Lower Limb Amputation (2007);18
- 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); and

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.”18(p.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.

---

**FIGURE 2-5** General ranking of evidence within hierarchies.
However, clinical practice guidelines should be assessed for quality in their own right before a clinician blindly adopts the practice behaviors they address.

DiCenso and colleagues have developed the “6S model” to aid in the selection of “pre-appraised” evidence such as clinical practice guidelines (Figure 2-6). In recognition of the value of a cumulative body of evidence, this hierarchy places all individual studies on the lowest level of the continuum. Computerized decision support systems that provide clinicians with the ability to integrate a specific patient’s characteristics with synthesized evidence sit at the top of the hierarchy. The levels in between comprise progressively greater degrees of abstraction from collections of studies previously evaluated for their quality. Development of these synopses, syntheses, and summaries is dependent upon groups with sufficient expertise and resources to locate, critically appraise, write, and publish cohesive analyses and practice recommendations based on the evidence gathered. As such, physical therapists may find it challenging to locate pre-appraised evidence that addresses their clinical questions.

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.” 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 or validity. Katrak et al. reported similar findings from their investigation into the utility of critical appraisal tools for evaluation of literature relevant to the allied health professions. The take-home message from these reports 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 more information regarding ratings criteria.

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, whether it is a single study or a filtered synthesis of a collection of studies, 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. Clinical judgment and expertise are essential to evidence-based physical therapist practice. 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 therapist practice 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 therapist practice.
4. Discuss the difference between a cross-sectional research design and a longitudinal research design and give an example of each that reflects a study question relevant to physical therapist practice.
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 potential difference in value between pre-appraised collections of individual studies and individual research reports. Why is the hierarchy for pre-appraised collections structured the way that it is?
8. Discuss the limitations of evidence hierarchies. Why is a hierarchy only a starting point in evidence-based physical therapist practice?

References

References
