CHAPTER 2

WHAT IS EVIDENCE?

OBJECTIVES

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

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."[1(p.251)]

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 or client that may serve as a basis for future research,[2] and describes the overall management of an unusual case or a condition that is infrequently encountered in practice or poorly described in the literature.[3]

Clinical practice guidelines: ". . . statements that include recommendations intended to optimize patient care. They are informed by a systematic review of evidence and an assessment of the benefits and harm of alternative care options."[4] also referred to as "summaries."[5]

Cross-sectional study: A study that collects data about a phenomenon during a single point in time or once within a single defined time interval.[6]

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

Efficacy: The extent to which an intervention or service produces a desired outcome under ideal conditions.[1]

Evidence: "A broad definition of evidence is any empirical observation, whether systematically collected or not. Clinical research evidence refers to the systematic observation of clinical events. . . ."[7]

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).[8,9]

Longitudinal study: A study that looks at a phenomenon occurring over an extended period of 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.[10]

Nonexperimental design (also referred to as an *observational study*): A study in which controlled manipulation of the subjects is lacking[8]; in addition, if groups are present, assignment is predetermined based on naturally occurring subject characteristics or activities.[6]

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.[6]

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.[11]

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.[8]

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.[8]

Synopsis: "A succinct description of selected individual studies or systematic reviews."[5]

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[6]; also referred to as "syntheses."[5]

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., computerized decision support systems)."[5]

CLINICAL SCENARIO

Relevant Patient Characteristics

Anne's case history contains information about her that may influence your search for evidence and your judgment about its relevance. Examples include age and fitness level.

What other characteristics can you identify that may influence your search for and application of evidence in Anne's situation?

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?" In their original work, Guyatt and Rennie stated, "any empirical observation about the apparent relation between events constitutes potential evidence."[7] This observation acknowledged that a variety of information types exist that may be integrated with clinical decisions. Options include, but are not limited to, published research articles, clinical practice guidelines, patient or client records, and recall of prior patient or client cases. As these authors acknowledged, however, clinical research is the preferred source of information. Sackett et al. put a finer point on it with their use of the modifier "best available" clinical evidence. They proposed that a method of prioritizing the evidence according to its merits is required to guide the clinician's selection of relevant information.[12] 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 or client.

The second content criterion pertains to the individuals studied. Desirable evidence includes subjects whose personal and/or clinical characteristics are similar to the patient or client in order to increase the therapist's ability to apply the research findings to this individual. Common attributes

of interest include, but are not limited to, the subjects' age, gender, race/ethnicity, education level, occupation, diagnosis(es), stage of illness, duration of the problem(s), functional status, level of disability, and clinical setting in which patient/client management occurs. Subjects in a research study whose personal and/or clinical characteristics differ markedly from those of a patient or client may have different therapeutic experiences than can be achieved by the individual with whom the physical therapist wishes to use the evidence located.

Two basic procedural characteristics have relevance in the evidence selection process as well. The period in time during which the evidence was developed is often of interest given the rapid evolution of medical technology and pharmaceutical agents. This is particularly relevant for research publications given that articles often appear in journals a year or more after the completion of the project.[7] Early release on the Internet ahead of the printed version undoubtedly has reduced this time line in many cases. Nevertheless, older evidence may not reflect current patient management. A hypothetical example might be a 15-year-old study evaluating the effectiveness of an aerobic training program for individuals with multiple sclerosis that has limited relevance now that a variety of disease-modifying drugs are available.[13] However, evidence should not be rejected only because of its age if the techniques in question, and the context in which they were evaluated, have remained relatively unchanged since the data were collected.

A procedural characteristic specific to scientific journals is the application of peer review. *Peer review* is the process by which manuscripts are evaluated by identified content experts to determine their merit for publication. Evaluation criteria usually include the credibility of a research study 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.[1] Peer review acts as an initial screening process to weed out lower quality efforts.

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 evidence related to physical therapist practice. Many of the clinical questions physical therapists have about their patients or 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 or client. Similarly, studies may not exist that include a test or technique of interest in the clinical setting. The evidence-based physical therapist (EBPT) practice challenge is to decide whether and how to use evidence that is limited in these ways when it is the only evidence available.

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

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

Forms of Evidence

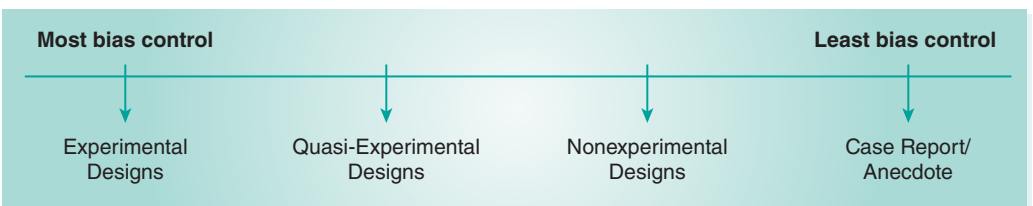
As noted previously, forms of evidence may include anything from patient records and clinical recall to published research. Evidence-based practice in health care emphasizes the use of research to inform clinical decisions because of the systematic way in which data are gathered and because of its potential to provide objective results that minimize bias. 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 or clients. The usefulness of a diagnostic test must be evaluated with methods that are different from 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 that was established prior to data collection (**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.[1] 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*.[1,6,8] 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 random assignment 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.[8]

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 some individuals 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 their 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

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



in the outpatient rehabilitation center, then the RCT is focused on treatment *effectiveness*.[14] 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," comprise original evidence that has been selected and critically appraised according to pre-established criteria.[6] 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 pre-established 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 or 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 or client rather than on groups of subjects on whom behavioral and data collection controls are imposed to ensure research integrity. This type of evidence often is labeled "anecdotal"[11] and frequently put to use when therapists recall from memory prior experiences with patients or clients similar to the person with whom they are currently dealing. In response to regulatory and payment 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 or 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 or 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- or 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.[11] *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.[8] 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 or 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*.[7]

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 or client, whereas *narrative reviews* summarize prior research.[2,3,10] In spite of these differences, these designs have one common element that puts them both at the bottom of the continuum: they lack the kind of systematic approach necessary to reduce bias. 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.

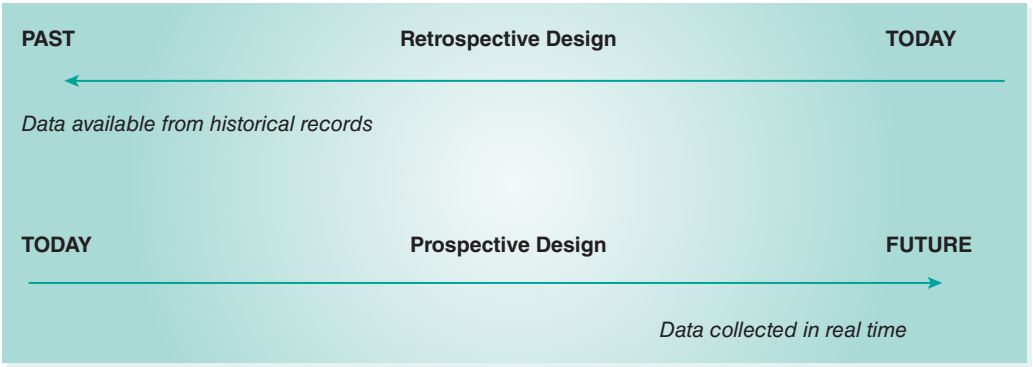
Research Designs: Timing

Research designs also may be categorized according to the time line used in the study. For example, physical therapist 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 historical 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**

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

Study Design	Citation
Systemic Review	Vanti C, et al. Effect of taping on spinal pain and disability: systematic review and meta-analysis of randomized trials. *Phys Ther*. 2015;95(4):493–506.
Randomized Clinical Trial	Fox EE, et al. Effect of Pilates-based core stability training on ambulant people with multiple sclerosis: multi-center, assessor blinded randomized controlled trial. *Phys Ther*. 2016;96(8):1170–1178.
Single-System Design	Chen YP, et al. Use of virtual reality to improve upper-extremity control in children with cerebral palsy: a single subject design. *Phys Ther.* 2007;87(11):1441–1457.
Quasi-Experimental Study	Drolett A, et al. Move to improve: the feasibility of using an early mobility protocol to increase ambulation in the intensive and intermediate care settings. *Phys Ther*. 2013;93(2):197–207.
Observational Study	Farley MK, et al. Clinical markers of the intensity of balance challenge: observational study of older adult responses to balance tasks. *Phys Ther*. 2016;96(3):313–323.
Physiologic Study	Chung JI, et al. Effect of continuous-wave low-intensity ultrasound in inflammatory resolution of arthritis-associated synovitis. *Phys Ther*. 2016;96(6):808–817.
Case Report	Parkitny L, et al. Interdisciplinary management of complex regional pain syndrome of the face. *Phys Ther*. 2016;96(7):1067–1073.
Summary	Barr AE, et al. Pathophysiologic tissue changes associated with repetitive movement: a review of the evidence. *Phys Ther.* 2002;82(2):173–187.

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



illustrates, a retrospective approach takes advantage of data that already exist, whereas a prospective approach requires that new data be collected in real time.

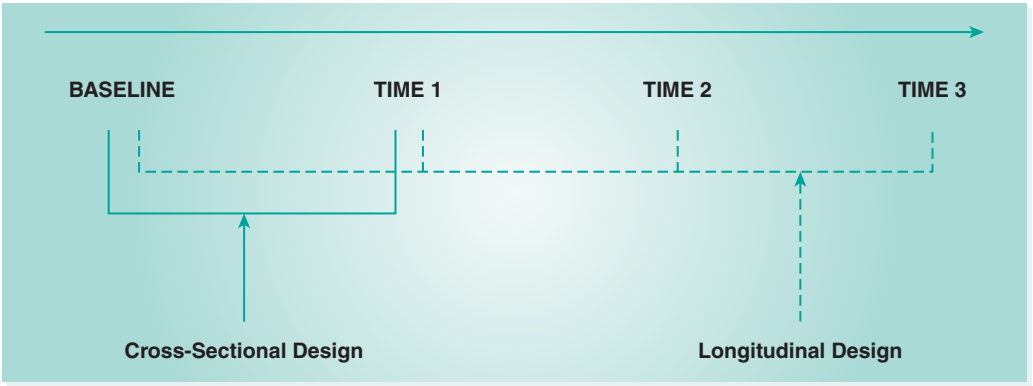
In a similar fashion, researchers may be interested in a single point in time or a limited time interval (e.g., *cross-sectional study*), or they may wish to study a phenomenon over an extended period of time (e.g., *longitudinal study*). In the cross-sectional approach, investigators may have an interest in the functional outcome at discharge (a single point in time) from the hospital of patients receiving physical therapist services 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.

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 or 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 subject's status or performance.

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 clinical test to identify a rotator 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 results from the physical examination. 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 (i.e., 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 cervical spine disorders 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 evidence identification process.

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



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 (or unpublished clinical data) 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 EBPT practice; rather, it is a reaffirmation that clinical judgment and expertise are required to decide whether and 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.[15] 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 **Table 2-3**).[16-18] 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 (i.e., "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

TABLE 2-3 Oxford Centre for Evidence-Based Medicine 2011 Levels of Evidence

Question	Step 1 (Level 1*)	Step 2 (Level 2*)	Step 3 (Level 3*)	Step 4 (Level 4*)	Step 5 (Level 5)
How common is the problem?	Local and current random sample surveys (or censuses)	Systematic review of surveys that allow matching to local circumstances**	Local nonrandom sample**	Case-series**	n/a
Is this diagnostic or monitoring test accurate? (Diagnosis)	Systematic review of cross-sectional studies with consistently applied reference standard and blinding	Individual cross-sectional studies with consistently applied reference standard and blinding	Nonconsecutive studies, or studies without consistently applied reference standards**	Case-control studies, or "poor or nonindependent reference standard**	Mechanism-based reasoning
What will happen if we do not add a therapy? (Prognosis)	Systematic review of inception cohort studies	Inception cohort studies	Cohort study or control arm of randomized trial*	Case-series or case-control studies, or poor quality prognostic cohort study**	n/a
Does this intervention help? (Treatment Benefits)	Systematic review of randomized trials or *n*-of-1 trials	Randomized trial or observational study with dramatic effect	Nonrandomized controlled cohort/follow-up study**	Case-series, case-control studies, or historically controlled studies**	Mechanism-based reasoning

Question	Step 1 (Level 1*)	Step 2 (Level 2*)	Step 3 (Level 3*)	Step 4 (Level 4*)	Step 5 (Level 5)
What are the COMMON harms? (Treatment Harms)	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	Individual randomized trial or (exceptionally) observational study with dramatic effect	Nonrandomized controlled cohort/follow-up study (post-marketing surveillance) provided there are sufficient numbers to rule out a common harm. (For long-term harms the duration of follow-up must be sufficient.)**	Case-series, case-control, or historically controlled studies**	Mechanism-based reasoning
What are the RARE harms? (Treatment Harms)	Systematic review of randomized trials or *n*-of-1 trial	Randomized trial or (exceptionally) observational study with dramatic effect			
Is this (early detection) test worthwhile? (Screening)	Systematic review of randomized trials	Randomized trial	Nonrandomized controlled cohort/follow-up study**	Case-series, case-control, or historically controlled studies**	Mechanism-based reasoning

*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.

Reproduced with permission from Oxford Center for Evidence-Based Medicine–www.cebm.net.

FIGURE 2-4 General ranking of evidence within hierarchies.

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 anatomic and/or physiologic 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 people 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 2-4** summarizes the commonalities among the OCEBM evidence hierarchies. Howick et al. also acknowledged some preliminary findings regarding the potential value of individual patient cases and anecdotes in evidence-based decision making.[16] Although not included in the current 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: Management of Concussion/Mild Traumatic Brain Injury (2016)[19];
- British Thoracic Society's "The BTS Guideline on Pulmonary Rehabilitation in Adults" (2013)[20]; and
- Orthopedic Section – APTA's "Nonarthritic Hip Joint Pain" (2014).[21]

TABLE 2-4 Evidence Grades in a Clinical Practice Guideline about Nonarthritic Hip Joint Pain

Grades Of Recommendation Based On	Strength of Evidence
A. Strong evidence	A preponderance of level I and/or level II studies support the recommendation. This must include at least 1 level I study
B. Moderate evidence	A single high-quality randomized controlled trial or a preponderance of level II studies support the recommendation
C. Weak evidence	A single level II study or a preponderance of level III and IV studies, including statements of consensus by content experts, support the recommendation
D. Conflicting evidence	Higher-quality studies conducted on this topic disagree with respect to their conclusions. The recommendation is based on these conflicting studies
E. Theoretical/foundational evidence	A preponderance of evidence from animal or cadaver studies, from conceptual models/principles, or from basic science/bench research supports this conclusion
F. Expert opinion	Best practice based on the clinical experience of the guidelines development team

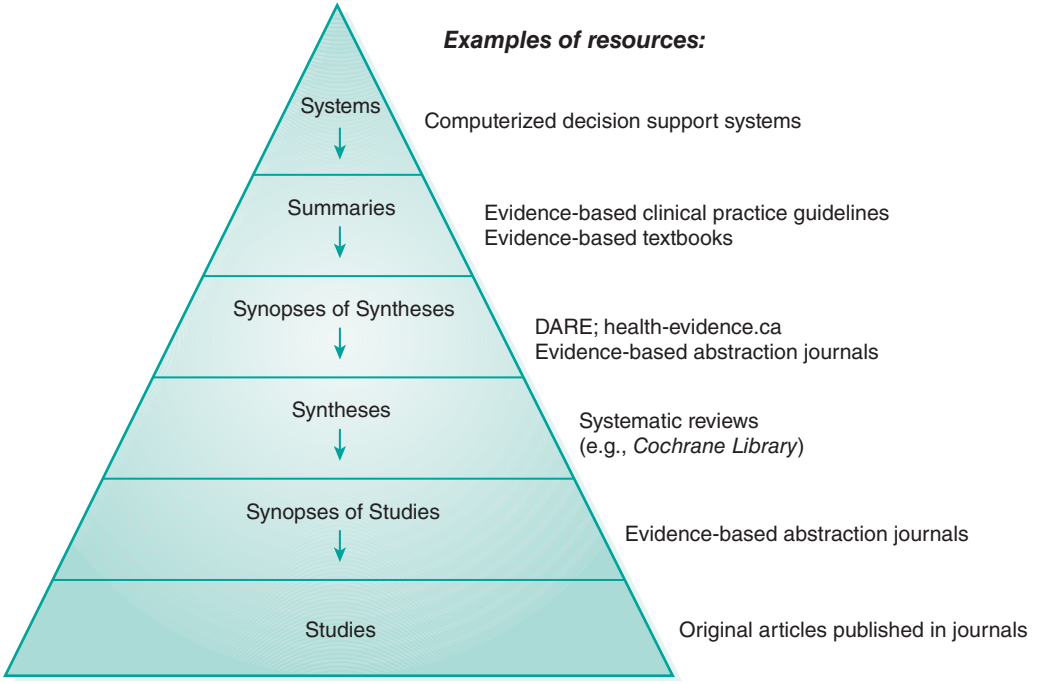
Reprinted with permission from Enseki K, Harris-Hayes M, White DM, et al. Nonarthritic hip joint pain. *J Orthop Sports Phys Ther*. 2014;44:A1-A32. https://doi.org/10.2519/jospt.2014.0302. ©Journal of Orthopaedic & Sports Physical Therapy®

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 Orthopedic Section – APTA assigned the letter grades A, B, C, D, E, and F to rate the strength of the evidence supporting the guideline's recommendations (**Table 2-4**). On the other hand, the Department of Veterans Affairs/Department of Defense (VA/DoD) used the categories "strong for," "weak for," "weak against," "strong against" based on the quality of the evidence and the balance between desirable and undesirable outcomes of the intervention(s) recommended.

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, 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 developed the "6S model" to aid in the selection of "pre-appraised" evidence such as clinical practice guidelines (**Figure 2-5**).[5] 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. Robeson and colleagues explored the availability of these

FIGURE 2-5 The 6S hierarchy of pre-appraised evidence.



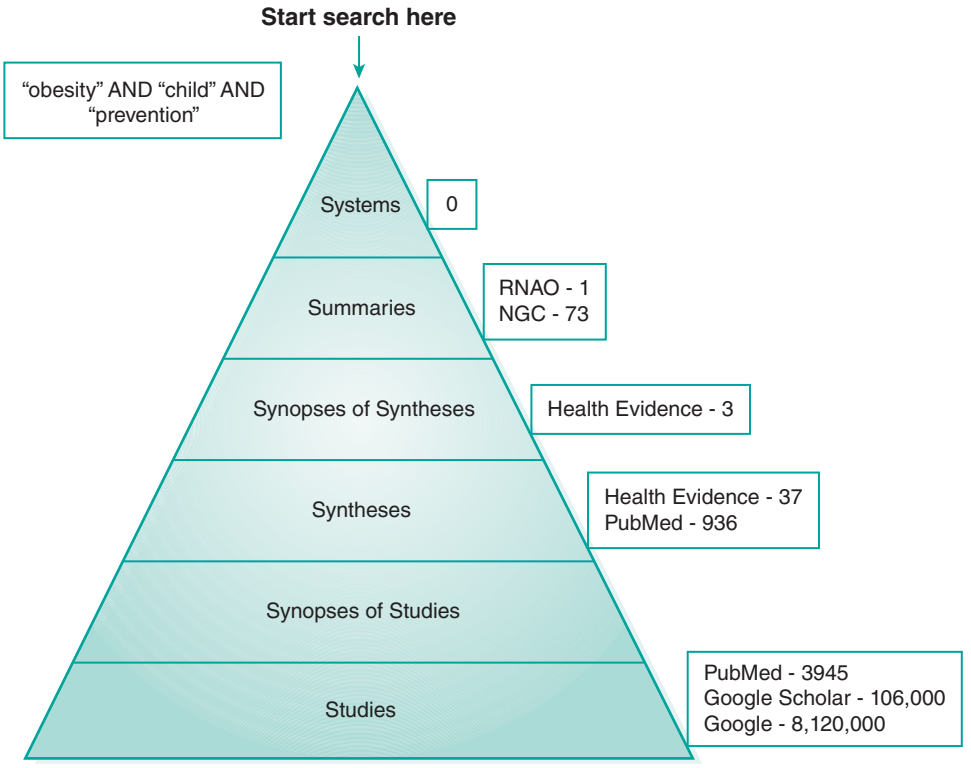
Reprinted from Evidence Based Nursing, DiCenso, A., Bayley, L., Haynes, RB., 12, 99-101, 2009 with permission from BMJ Publishing Group Ltd.

different forms of preprocessed evidence for questions related to effectiveness of public health.[22] As **Figure 2-6** suggests, the number of research products in each category varies considerably with a predictably higher volume of traditional evidence formats (i.e., individual studies, systematic reviews, and meta-analyses). Development of *synopses*, *syntheses*, *summaries,* and *systems* 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.

The Quality of Evidence Generated from Words

The research designs described above and the hierarchies used to rank them are consistent with an investigative paradigm that emphasizes objectivity, faithfulness to rules of engagement with subjects, and use of quantitative data to describe clinical phenomena. Even the subjective perspectives of patients or clients are standardized into controlled survey responses through questions with

FIGURE 2-6 **Search results mapped to 6S pyramid.**



Modified from Accessing pre-appraised evidence: fine-tuning the 5S model into a 6S model, DiCenso, Bayley, & Haynes, 2009, 12, 99-101, 2010 with permission from BMJ Publishing Group Ltd.

a limited set of options of known value. The evidence-based practice movement is solidly built on these scientific traditions.

More recently, research generated by thematic analysis of the words people use to describe their experiences, beliefs, opinions, and attitudes has gained legitimacy as a contributor to evidence-based practice in the health care professions. As these qualitative studies have proliferated and their potential value been embraced, a limited effort has been made to help clinicians understand what makes a strong qualitative research design. Daly et al. proposed the four-level hierarchy depicted in **Table 2-5**.[23] Consistent with evidence hierarchies constructed for use with quantitative studies, these authors suggest a progression of increasingly rigorous designs that follow established methods for subject recruitment, data gathering, and analysis. This hierarchy has not been universally adopted by academics, research groups, journal publishers, or professional societies but may be useful to readers unfamiliar with the research methods implemented in this body of work.

TABLE 2-5 A Hierarchy of Evidence for Practice in Qualitative Research—Summary Features

Study Type	Features	Limitations	Evidence for Practice
Generalizable studies (level I)	Sampling focused by theory and literature, extended as a result of analysis to capture diversity of experience. Analytic procedures comprehensive and clear. Located in literature to assess relevance to other settings.	Main limitations are in reporting when the word length of articles does not allow a comprehensive account of complex procedures.	Clear indications for practice or policy may offer support for current practice, or critique with indicated directions for change.
Conceptual studies (level II)	Theoretical concepts guide sample selection, based on analysis of literature. May be limited to one group about which little is known or a number of important subgroups. Conceptual analysis recognizes diversity in participants' views.	Theoretical concepts and minority or divergent views that emerge during analysis do not lead to further sampling. Categories for analysis may not be saturated.	Weaker designs identify the need for further research on other groups, or urge caution in practice. Well-developed studies can provide good evidence if residual uncertainties are clearly identified.
Descriptive studies (level III)	Sample selected to illustrate practical rather than theoretical issues. Record a range of illustrative quotes including themes from the accounts of "many," "most," or "some" study participants.	Do not report full range of responses. Sample not diversified to analyze how or why differences occur.	Demonstrate that a phenomenon exists in a defined group. Identify practice issues for further consideration.
Single case study (level IV)	Provides rich data on the views or experiences of one person. Can provide insights in unexplored contexts.	Does not analyze applicability to other contexts.	Alerts practitioners to the existence of an unusual phenomenon.

Reprinted from *Journal of Clinical Epidemiology*, 60(1), Daley J, Willis K, Small R et al., A hierarchy of evidence assessing qualitative health research, 43-49, Copyright 2007, with permission from Elsevier.

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."[24] 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 and colleagues reported similar findings from their investigation into the utility of critical appraisal tools for evaluation of literature relevant to the allied health professions.[25] In an effort to address some of these limitations, Atkins and colleagues conducted a pilot study to refine one of the established grading systems. These authors discovered that statistical agreement among raters on evidence quality ranged from zero to a nearly perfect score (kappa = 0 to 0.82).[26] 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.[27] More recently, Gugiu and Gugiu proposed a new grading system intended to be more inclusive of treatment studies that are not RCTs.[28, 29]

Understanding the details and bases for evidence hierarchies will help physical therapists select evidence to answer clinical questions about patients or 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.[30-32] Clinical judgment and expertise are essential to EBPT 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 or client.

Summary

EBPT practice requires clinicians to select the "best available clinical 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. Explain the importance of similarities between an individual patient or client and subjects in a research study. Provide three examples of personal or clinical characteristics from a patient with whom you have worked recently to illustrate your points.
6. Describe the common organizational characteristics of evidence hierarchies.
7. Discuss the rationale behind the creation of different hierarchies for evidence about diagnostic tests, prognostic factors, and interventions.
8. Discuss the potential difference in value between preappraised collections of individual studies and individual research reports. Why is the hierarchy for preappraised collections structured the way that it is?
9. Discuss the limitations of evidence hierarchies. Why is a hierarchy only a starting point in EBPT 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.* 3rd ed. Alexandria, VA: American Physical Therapy Association; 2009.
3. Information for Authors: "Full" Traditional Case Reports. Available at: http://ptjournal.apta.org/content/information-authors-full-traditional-case-reports. Accessed July 31, 2016.
4. Graham R, Mancher M, Wolman DM, Greenfield S, Steinberg E, eds. *Clinical Practice Guidelines We Can Trust.* Available at: http://www.nap.edu/catalog/13058/clinical-practice-guidelines-we-can-trust. Accessed July 31, 2016.
5. DiCenso A, Bayley L, Haynes RB. Accessing pre-appraised evidence: fine-tuning the 5S model into a 6S model. *Evid Based Nurs.* 2009;12(4):99–101.
6. Straus SE, Richardson WS, Glaziou P, Haynes RB. *Evidence-Based Medicine: How to Practice and Teach EBM.* 3rd ed. Edinburgh, Scotland: Elsevier Churchill Livingstone; 2005.
7. Guyatt G, Rennie D. *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice.* Chicago, IL: AMA Press; 2002.
8. Carter RE, Lubinsky J, Domholdt E. *Rehabilitation Research: Principles and Applications.* 4th ed. St Louis, MO: Elsevier Saunders; 2011.
9. Campbell DT, Stanley JC. *Experimental and Quasi-Experimental Designs for Research.* Boston, MA: Houghton Mifflin; 1963.

10. Herbert R, Jamtvedt G, Hagen KB, Mead J. *Practical Evidence-Based Physiotherapy*. 2nd ed. Edinburgh, Scotland: Elsevier Butterworth-Heinemann; 2011.
11. Cook TD, Campbell DT. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Boston, MA: Houghton Mifflin; 1979.
12. Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence-based medicine: what it is and what it isn't. *BMJ.* 1996;312(7023):71–72.
13. Disease Modification. National Multiple Sclerosis Society website. Available at: http://www.nationalmssociety.org/For-Professionals/Clinical-Care/Managing-MS/Disease-Modification. Accessed August 1, 2016.
14. 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://www.ptresearch.org/members/Weinstein.pdf. Accessed August 1, 2016.
15. 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.
16. Howick J, Chalmers I, Glasziou P, et al. Explanation of the 2011 Oxford Centre for Evidence-Based Medicine (OCEBM) Levels of Evidence (Background Document). Oxford Centre for Evidence-Based Medicine. Available at: www.cebm.net/index.aspx?o=5653. Accessed August 1, 2016.
17. Howick J, Chalmers I, Glasziou P, et al. The 2011 Oxford CEBM Levels of Evidence (Introductory Document). Oxford Centre for Evidence-Based Medicine. Available at: www.cebm.net/index.aspx?o=5653. Accessed August 1, 2016.
18. OCEBM Levels of Evidence Working Group. The Oxford Levels of Evidence 2. Oxford Centre for Evidence-Based Medicine. Available at: http://www.cebm.net/wp-content/uploads/2014/06/CEBM-Levels-of-Evidence-2.1.pdf. Accessed August 1, 2016.
19. Department of Veterans Affairs, Department of Defense. VA/DoD Clinical Practice Guideline: Management of Concussion/Mild Traumatic Brain Injury. Washington, DC: Department of Veterans Affairs, Department of Defense; 2016. Available at: http://www.healthquality.va.gov/guidelines/Rehab/mtbi/. Accessed August 8, 2016.
20. Bolton CE, Brevin-Smith EF, Blakey JD, et al. The BTS guideline on pulmonary rehabilitation in adults. *Thorax.* 2013;68(S2):ii1–ii30.
21. Enseki K, Harris-Hayes M, White DM, et al. Nonarthritic hip joint pain. *JOSPT*. 2014;44(6):A1–A32.
22. Robeson P, Dobbins M, DeCorby K, et al. Facilitating access to pre-processed research evidence in public health. *BMC Public Health*. 2010;95.
23. Daley J, Willis K, Small R, et al. A hierarchy of evidence assessing qualitative health research. *J Clin Epidemiol*. 2007;60(1):43–49.
24. 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.
25. Katrak P, Bialocerkowski AE, Massy-Westropp N, et al. A systematic review of the content of critical appraisal tools. *BMC Med Res Methodol.* 2004;4(22).
26. Atkins D, Briss PA, Eccles M, et al. Systems for grading the quality of evidence and strength of recommendations II: pilot study of a new system. *BMC Health Serv Res*. 2005;5(1):25.
27. Glasziou P, Vandenbroucke J, Chalmer I. Assessing the quality of research. *BMJ.* 2004;328(7430):39–41.
28. Gugiu PC, Gugiu MR. A critical appraisal of standard guidelines for grading levels of evidence. *Eval Health Prof*. 2010;33(3):233–255.
29. Gugiu PC. Hierarchy of evidence and appraisal limitations (HEAL) grading system. *Eval Program Plann*. 2015;48:149–159.
30. Britton A, McKee M, Black N, McPherson K, Sanderson C, Bain C. Choosing between randomised and non-randomised studies: a systematic review. *Health Technol Assess.* 1998;2(13):i–iv,1–124.
31. MacLehose RR, Reeves BC, Harvey IM, Sheldon TA, Russell IT, Black AM. A systematic review of comparisons of effect sizes derived from randomised and non-randomised studies. *Health Technol Assess.* 2000;4(34):1–154.
32. Glasziou P, Chalmers I, Rawlins M, McCulloch P. When are randomised trials unnecessary? Picking signal from noise. *BMJ*. 2007;334(7589):349–351.