Chapter 6
Evaluating Research: Research Designs in Evidence-Based Medicine/Evidence-Based Practice

Once you have located some research reports that can help answer your practice question, Step 3 in the evidence-based medicine (EBM) and evidence-based practice (EBP) decision making model is to appraise the quality and relevance of this research. An initial inspection of materials should help differentiate those that are relevant for your purposes from those that are not. Relevance may often be determined by examining the research question that each study addresses. Studies should have clear and relevant research questions, fitting your practice needs. That is, the topics should fit your clinical question, and the sample should be similar in age and other background criteria. Once these 'apparently relevant' studies are identified, the appraisal shifts to issues of research methodology. Even studies that appear be quite relevant initially may later on prove to have important limitations as the details of their methods are explored.

Evaluating the quality of individual research reports can be a complex process. It involves several components. This chapter will review the research designs that orient studies used in EBP. While many of these designs should be familiar to social workers, they may be described using different terminology in EBM and EBP research reports. Yet research designs are one key aspect of establishing research quality. When clinical social workers have to review individual research articles, determining their research design will be a key aspect of establishing their quality and rigor. Chapter 7 will review several other methodological steps in appraising individual research reports (sampling, defining the treatment or other intervention, tests and measures, and statistics). Together Chapters 6 and 7 provide a basis for assessing the quality of *individual* research studies. These same methodological issues are also vital to the aggregation of multiple research studies on a particular topic like depression. In Chapter 8 will we examine how multiple studies are aggregated for review in EBM/EBP through meta-analysis and systematic reviews. Both of these aggregation methods build upon specific quality standards for individual research reports. This is why we will begin our review of research methods at the level of appraising the individual research report.

J. W. Drisko and M. D. Grady, *Evidence-Based Practice in Clinical Social Work*, Essential Clinical Social Work Series, DOI: 10.1007/978-1-4614-3470-2_6,

Research design is the first methodological issue a clinical social worker must identify in appraising the quality of a research study. A research design is the orienting plan that shapes and organizes a research project. Researchers use different research designs for projects with distinct goals and purposes. Sometimes this is a researcher-determined choice, and other times practical and ethical issues force the use of specific research designs. In EBM/EBP, research designs are one key part of appraising study quality.

While all clinical social workers are introduced to research methods as part of their required course work, most do not make much use of this knowledge after graduation. EBP, however, will require clinical social workers and other mental health professionals to make greater use of their research knowledge in evaluating research for practice. Therefore, we will review several types of research designs in considerable detail.

Research designs are so important to EBM/EBP that this chapter will focus on them exclusively. (As noted, we will examine other very important–and very closely related–aspects of research methods will be examined in the next chapter.) Our goal is to provide a useful refresher and reference for clinical social workers. For readers who have a basic grasp of research designs and methods, this chapter can serve as a brief review and resource. Some terminology, drawn from medicine, will no doubt be unfamiliar. For other readers, who need only an update on research methods, this chapter offers it. We will identify many excellent follow-up resources in each section of the chapter. We also remind readers that there is an extensive glossary at the end of this book.

Research Designs

This review of research designs has three main purposes. First, it will introduce the variety of terminology used in EBP research, which is often drawn from medical research. Much of this terminology differs from the terminology used in most social work research texts that draw on social sciences research terminology. Second, the strengths and limitations of each research design are examined and compared. Third, the research designs are rank ordered from 'strongest' to 'weakest' following the EBM/EBP research hierarchy. This allows readers to quickly understand why some research designs are favored in the EBM/EBP literature.

Thyer (2011) states, quite accurately, that the EBP practice decision-making process does not include any hierarchy of research designs. This is indeed correct. The EBP practice decision-making process states that clinicians should use the best available evidence. It does not state that only the results of research with certain types of research designs are to be valued. That is, it is entirely appropriate to use the results of case study research, or even practice wisdom when no better evidence is available. Yet many organizations and institutions make quite explicit that there

is a *de facto* hierarchy of evidence within EBP. This hierarchy is even clearly stated in the early writing of Dr. Archie Cochrane (1972), who promoted the use of experimental research knowledge to inform contemporary practice decision making. Littell (2011) notes that the Cochrane Collaboration publishes, 'empty reviews' that report no research results deemed to be of sufficient quality to guide practice decision making. This practice contradicts the idea of identifying the best available evidence. In effect, the best available evidence is limited to evidence generated by experimental research designs. This practice creates confusion about what constitutes the best available evidence for clinicians, policy planners, and researchers. Such authors do not report the best available evidence, but instead report only experimental evidence that they deem worthy of guiding practice. They make this choice because only well-designed experiments allow attribution of causal relationships; to say that an intervention caused observed changes with minimal error. Still, this practice represents some academic and economic politics within EBP research summaries. As discussed in Chapter 2, there are good arguments for and against this position, but it is not entirely consistent with the stated EBM/EBP practice decision-making model. Clinical social workers should be aware that this difference in viewpoints about the importance of research design quality is not always clearly stated in the EBP literature. Critical, and well-informed, thinking by the clinician is always necessary.

Research designs differ markedly. They have different purposes, strengths, and limitations. Some seek to explore and clarify new disorders or concerns and to illustrate innovative practices. Others seek to describe the characteristics of client populations. Some track changes in clients over time. Still others seek to determine if a specific intervention caused a specific change. While we agree that the EBP practice decision-making process states that clinicians should use the best available evidence, and not solely evidence derived from experimental results, we will present research designs in a widely used hierarchy drawn from the Oxford University Centre for Evidence-based Medicine (2009, 2011). This hierarchy does very clearly give greater weight to experimental, randomized controlled trial (RCT) research results. It should be seen as representing a specific point of view, applied for specific purposes. At the same time, such research designs do provide a strong basis for arguing that a treatment caused any changes found, so long as the measures are appropriate, valid and reliable and the sample tested is of adequate size and variety. Due to the strong interval validity offered by experimental research designs, results based on RCTs design are often privileged in EBM/EBP reports.

We will begin this listing with the experimental research designs that allow causal attribution. We will then progress from experiments, to quasi-experiments, and then move to observational or descriptive research and end with case studies. The organization of this section follows the research evidence hierarchy of the Oxford University Centre for Evidence-based Medicine (2009, 2011).

Types of Clinical Studies

Part I: Experimental Studies

EBP researchers view properly conceptualized and executed experimental studies (also called randomized controlled trials or RCTs) as providing very strong empirical evidence of treatment effectiveness. They are prospective in nature as they start at the beginning of treatment and follow changes over time (Anastas 1999). Random assignment of participants symmetrically distributes potential confounding variables to each group in the study. Probability samples provide a suitable foundation for most statistical analytic procedures.

The key benefit of an experimental research design is that it minimizes threats to internal validity (Campbell and Stanley 1963). This means the conclusions of well-done experiments allow researchers to say an intervention caused the observed changes. This is why experiments are so highly regarded in the EBM/EBP model. The main limitation of experiments is their high cost in money, effort, and time. Further, they may be ethically unsuitable for some studies where random assignment is not appropriate. A final disadvantage is that volunteers willing to participate may not reflect clinical populations well. This may lead to bias in external validity, or how well results from controlled experiments can be generalized to less controlled practice settings (Centre for Evidence Based Medicine 2009).

In the European medical literature, experiments and quasi-experiments may be called *analytic studies*. This is to distinguish them from descriptive studies that, as the name implies, simply *describe* clinical populations. Analytic studies are those that quantify the relationship between identified variables. Such analytic studies fit well with the P.I.C.O. or P.I.C.O.T. treatment decision-making model (Center for Evidence Based-Medicine, undated).

A. Randomized Controlled Trial

A prospective, group-based, quantitative, experimental study based on primary data from the clinical environment (Solomon et al. 2009). Researchers randomly assign individuals with the same disorder or problem at the start to one of two (or more) groups and the outcomes for each group are compared at the completion of treatment. Since researchers create the groups by random assignment to generate very similar groups, the RCT is sometimes called a *parallel group design*. Usually one group is treated and the other is used as an untreated control group. Researchers sometime use placebo interventions with the control group. However, researchers may alternately design experiments comparing two or more different treatments where one has been previously demonstrated to produce significantly better results than were found for an untreated control group. Pre- to post- comparisons demonstrate the changes for each group. Comparison of post- scores across the treated groups allows for demonstration of greater improvement due to the treatment.

Follow-up comparisons may also be undertaken, but this is not a requirement of an experiment.

The experiment or RCT can be summarized graphically as:

$$\begin{array}{llll} \mathbf{R} & \mathbf{O}_1 & \mathbf{X} & \mathbf{O}_2 \\ \mathbf{R} & \mathbf{O}_1 & & \mathbf{O}_2 \end{array}$$

where **R** stands for random assignment of participants, $\mathbf{O}_1$ stands for the pretest assessment (most often done using a standardized measure), **X** represents the intervention, and $\mathbf{O}_2$ stands for the post-test, done after treatment, using the same measure (Kazdin 2002). There may also be additional follow-up post-tests to document how results vary over time after treatment ends. These would be represented as $\mathbf{O}_3$, $\mathbf{O}_4$, etc. Frequently, more than one measure of outcome is used in the same experiment.

In medical studies, particularly of medications or devices, it is possible to "*blind*" participants, clinicians, and even researchers to their experimental group assignments. The goal is to reduce differences in expectancies that might lead to different outcomes. In effect, either conscious or unconscious bias is limited to strengthen the validity of the study results. A *"double blind"* experiential study keeps all group assignments unknown to participants and to the treating clinicians. *"Single blind"* experiments keep only the participants unaware of group assignments. Blinding is more possible where placebo pills or devices can be used to hide the nature of the intervention. Blinding is much more difficult in mental health and social service research where interactions between clients and providers over time are common. In mental health research, interactions between client and provider make double blinding very difficult.

While blinding is common in EBM studies of medications and devices, it is rare in mental health research. There is, however, research that shows that clinical practitioners and researchers may act consciously or unconsciously to favor treatment theories and models that they support (Dana and Loewenstein 2003). This phenomenon is known as "*attribution bias*," in which people invested in a particular theory or treatment see it more positively than they do other approaches. Attribution bias may work consciously or unconsciously to influence study implementation and results. In turn, it is stronger research evidence if clinicians and researchers who do outcome studies are not the originators or promoters of the treatment under study. The American Psychological Association standards for empirically supported treatments (ESTs) require that persons other than the originators of a treatment do some of the outcome studies used to designate an EST (Chambless and Hollon 1998). That is, at least one study not done by the originator of a treatment is required for the EST label. How clinician and researcher biases are assessed in the EBM/EBP model is less clear. Similarly, Cochrane Collaboration and Campbell Collaboration systematic reviews do assess and evaluate the potential for bias when the originators of treatments are the only sources of outcome research on their treatments (Higgins and Green 2011; Littell et al. 2008). All Cochrane and Campbell Collaboration systematic reviews must include a statement of potential conflicts of interest by each of the authors.

It is important to keep in mind that experiments based on small samples may have serious limitations despite their use of a 'strong' research design. Sample size is one such issue. Many clinical studies compare small groups (under 20 people in a group). Studies using small samples may lack the statistical power to identify differences across the groups correctly and fully. That is, for group differences to be identified, a specific sample size is required. The use of an experimental research design alone does not mean that the results will always be valid and meaningful. (We will examine issue beyond research design that impact of research quality later in the Chapter 7.) Still, done carefully, the experimental research design or RCT has many merits in allowing cause-effect attribution.

The CONSORT Statement (2010) established standards for the reporting of RCTs. CONSORT is an acronym for "CONsolidated Standards of Reporting Trials." The people who make up the CONSORT group are an international organization of physicians, researchers, methodologists and publishers. To aid in the reporting of RCTs, CONSORT provides a free 37 item checklist for reporting or assessing the quality of RCTs online at www.consort-statement.org/index.aspx?o=2964. The CONSORT group also provides a free template for a flow chart of the RCT process and statement online at www.consort-statement.org/index.aspx?o=2966. These tools can be very helpful to the consumer of experimental research since they serve as guides for assessing the quality of RCTs. The CONSORT flow chart is often found in published reports of recent RCTs.

B. Randomized Cross-Over Clinical Trial

A prospective, group-based, quantitative, experimental research design based on primary data from the clinical environment. Individuals with the same disorder, most often of a chronic or long-term type, are randomly assigned to one of two groups and treatment is begun for both groups. After a designated period of treatment sufficient to show positive results, groups are assessed and a "wash-out" phase is begun in which all treatments are withheld. After the washout period is completed, the treatments for the groups are then switched so that each group receives both treatments. After the second course of treatment is completed, a second assessment is undertaken. Comparison of outcomes for each treatment at both end points allows for determination of treatment effectiveness on the same groups of patients/clients for both treatments. A comparison of active treatment outcomes for all patients is possible. However, if the washout period is not sufficient, there may be carry over effects from the initial treatment that in turn undermines the validity of the second comparison. Used with many medications, there are often lab tests that allow determination of effective washout periods. Secondary effects, such as learning or behavior changes that occur during the initial treatment may continue and not 'washout.' Similarly, it may not be possible to washout learned or internalized cognitions, skills, attitudes or behaviors.

The merit of cross-over designs is that each participant serves as his or her own control which reduces variance due to individual differences among participants. This may also allow use of smaller sample sizes while generating a large enough sample to demonstrate differences. This is known as statistical power. All participants receive both treatments. Random assignment provides a solid foundation for statistical tests. Disadvantages of cross-over studies include that all participants receive a placebo or less effective treatment at some point which does not benefit them immediately. Further, washout periods can be lengthy and curtail active treatment for the washout period. Finally, cross-over designs cannot be used where the effects of treatment are permanent, such as in educational programs or surgeries.

Cross-over trials may also be undertaken with single cases rather than groups of participants. These are called *single case cross-over trials.* The basic plan of the single case cross-over trial mimics that used for groups of clients, but is used with just a single case. The cross over trial may be represented graphically as:

$$\mathbf{A}_1 \quad \mathbf{B}_1 \quad \mathbf{A}_2 \quad \mathbf{B}_2 \quad \mathbf{A}_3$$

where $\mathbf{A}_1$ stands for the initial assessment, $\mathbf{B}_1$ represents the first intervention given, $\mathbf{A}_2$ represents the next assessment which is made at the end of the first intervention and $\mathbf{B}_2$ stands for second type of intervention or the cross-over. Finally, $\mathbf{A}_3$ represents the assessment of the second intervention done when it is completed. Note that a wash out period is not specifically included in this design, but could be included. Comparison of treatment outcomes for each intervention with the initial baseline assessment allows determination of the intervention effects. More than one measure may be used in the same study.

Since random assignment is not possible with single cases, the results of single case cross over studies are often viewed as 'weaker' than are group study results. However, each individual, each case, serves as its own control. Since the same person is studied, there is usually little reason to assume confounding variables arise due to physiologic changes or social circumstances.

It is possible to aggregate the results of single case designs. This is done by closely matching participants and replicating the single case study over a number of different participants and settings. This model is known as *replication logic*, in which similar outcomes over many cases builds confidence in the results (Anastas 1999). It is in contrast to *sampling logic* used in experiments in which potentially confounding variables are assumed to be equally distributed across the study groups through random assignment of participants. In replication logic, repetition over many cases is assumed to include potentially confounding variables. If treatment outcomes are positive over many cases, the general effectiveness of the treatment may be inferred.

In EBM single case studies are not usually designated as providing strong research evidence. Yet consistent findings from more than 10 single case study outcomes are rated as strong evidence in the American Psychological Association's designation of empirically supported treatments (ESTs) (Chambless and Hollon 1998).

C. Randomized Controlled Laboratory Study

A prospective, group, quantitative, experimental study based on laboratory rather than direct clinical data. These are called *analog studies* since the lab situation is a good, but not necessarily perfect, replication of the clinical situation. Laboratory studies are widely used in the so-called 'basic' research since all other variables or influences except the one under study can be controlled or identified. This allows testing of single variables, but is unlike the inherent variation found in clinical settings. Randomized controlled laboratory studies are often conducted on animals, where genetics can be controlled or held constant. Ethical issues, of course, limit laboratory tests on humans. Applying the results of laboratory studies in clinical practice has some limitations, as single, 'pure' forms of disorders or problems are infrequent and contextual factors can impact of treatment delivery and outcome.

Effectiveness Versus Efficacy Studies: Experiments Done in Different Settings

In mental health research, a distinction is drawn between clinical research done in real-world clinical settings and that done much more selectively for research purposes. Experimental studies done in everyday clinical practice setting are called *"effectiveness studies."* Such studies have some potentially serious limitations in that they often include people with comorbid disorders and researchers may not be able to ensure that treatments are provided fully and consistently to all clients. This reduces the interval validity of effectiveness studies for research purposes. On the other hand, using real-world settings enhances their external and ecological validity, meaning that the results fit with real-world practice and generalize to everyday clients and settings quite well. In contrast, more carefully controlled studies that ensure experimental study of just a single disorder are known as "*efficacy studies."* Efficacy studies carefully document that a fully applied treatment for a single, carefully screened disorder are effective (or are not effective).

One well-known example of a clinical efficacy study is the NIMH Cross-site Study of Depression (Elkin et al. 1989). This study rigorously compared medication and two forms of psychotherapy for depression. Strict exclusion criteria targeted only people with depression and no other comorbid disorders. Medication washouts were required of all participants. Such efficacy studies emphasize internal validity; they focus on showing that the treatment alone caused any change. The limitations of applying efficacy studies results are that real-world practice settings may not be able to take the time and effort needed to identify only clients with a single disorder. Such efforts might make treatment unavailable to people with comorbid disorders, which may not be practical or ethical in many clinical settings. Further, the careful monitoring of treatment fidelity required in efficacy studies may not be possible to

provide in many clinical settings. This is often for practical reasons of funding, staffing, and time.

Efficacy studies are somewhat like laboratory research, but the similarity is not quite exact since they are done in clinical settings, just with extra care. Efficacy studies add an extra measure of rigor to clinical research. They do show with great precision that a treatment works for a specific disorder. However, results of efficacy studies may be very difficult to apply fully in everyday clinical work given the ethical, funding, and practical limitations of clinical work.

D. The Quasi-Experimental Study or Cohort Study

In studies of clinical practice in mental health, it is sometimes unethical or impractical to randomly assign participants to treated and control groups. For example, policy makers may only fund a new type of therapy or a new prevention program for a single community, or only certain types of insurance pay for the new therapy. In such situations, researchers use existing groups to examine the impact of interventions. Where pre- and post-comparisons are done on both groups, such a research design is called a quasi-experiment. The key difference from a true experiment is the lack of random assignment of participants to the treated and control groups.

The quasi-experiment can be summarized graphically as:

$$\begin{array}{ccc} \mathbf{O}_1 & \mathbf{X} & \mathbf{O}_2 \\ \mathbf{O}_1 & & \mathbf{O}_2 \end{array}$$

Once again, $\mathbf{O}_1$ stands for the pretest assessment (most often with a standardized measure), **X** represents the intervention, and $\mathbf{O}_2$ stands for the post-test, done after treatment using the same measure (Kazdin 2002). There may also be additional follow-up post-tests to document how results vary over time after treatment ends. More than one measure may be used in the same quasi-experiment. Note carefully that the key difference from a true experiment is only the lack of random assignment of participants.

The lack of random assignment in a quasi-experiment introduces some threats to the internal validity of the study. That is, it may introduce unknown differences across the groups that ultimately affect study outcomes. The purpose of random assignment is to distribute unknown variables or influences to each group as equally as possible. Without random assignment, the two groups may have important differences that are not equally distributed. Say, for example, that positive social supports interact with a treatment to enhance its outcome. Without random assignment, the treated group might be biased in that it includes more people with strong social supports than does the control group. The interaction of the treatment with the impact of social supports might make the results appear better than they might have been if random assignment had been used. Thus in some EBM/EBP hierarchies of research evidence, quasi-experimental study results are rated as ‘weaker’ than are the results of true experiments. That said,

quasi-experiments are still very useful sources of knowledge. They are often the best available research evidence available for some treatments and service programs. To reduce potential assignment bias, quasi-experimental studies use "matching" in which as many characteristics of participants in each group are matched as closely as possible. Of course matching is only possible where the relevant variables are fully known at the start of the study.

Advantages of cohort studies include their ethical appropriateness in which participants are not assigned to groups. Participants can also make their own treatment choices on an informed basis. Cohort studies are usually less expensive in cost than are true experiments, though they may both be costly. Disadvantages of cohort studies are potentially confounding variables may be operative but unknown. Further, comparison groups can be difficult to identify. For rare disorders, large samples are required which can be difficult to obtain and may take a long time to complete.

E. The "All or None" Study

The Center for EBM at Oxford University (2009) includes in its rating of evidence the "All or None" research design. This is a research design in which, in very desperate circumstances, clinicians give an intervention to a group of people at high risk, usually of dying. If essentially all the people who received the intervention survive, while those who do not receive it die, the inference is that the intervention caused the life saving change. This is actually an observational research design, but the all or none results are viewed as strong evidence that the treatment caused the change. However, given their very important life saving effects, such research results are highly valued—so long as all or a large fraction of people who receive the intervention survive after treatment. Such designs fit crisis medical issues much better than most mental health issues, so all or none design are extremely rare in the mental health literature. They do have a valuable role in informing practice in some situations.

Part II: Observational Studies, Including Non-interventive Studies

Not all practice research is intended to show that an intervention *causes* a change. While EBM/EBP hierarchies of research evidence rank most highly those research designs that can show an intervention causes a change, even these studies stand on a foundation built from the results of other types of research. In the EBM/EBP hierarchy, clinicians are reminded that exploratory and descriptive research may not be the best evidence on which to make practice decisions. At the same time, exploratory and descriptive research designs are essential in setting the stage for

rigorous and relevant experimental research. These types of studies may also be the best available evidence for EBP if experiments are lacking or are of poor quality. Critical thinking is crucial to determining just what constitutes the best available evidence in any clinical situation.

F. Observational Studies

Are prospective, longitudinal, usually quantitative, tracking studies of groups of individuals with a single disorder or problem (Kazdin 2002). Researchers follow participants over time to assess the course (progression) of symptoms. Participants may be either untreated or treated with a specified treatment. People are not randomly assigned to treated or control groups. Because participants may differ on unknown or unidentified variables, observational studies have potential for bias due to the impact of these other variables. That is, certain variables such as genetic influences or nutrition or positive social support may lead to different outcomes for participants receiving the same treatment (or even no treatment). Some scholars view observational studies as a form of descriptive clinical research that is very helpful in preparing the way for more rigorous experimental studies.

G. Cohort Study (Also Called Longitudinal Study or Incidence Study)

A prospective, quantitative and/or qualitative, observational study ideally based on primary data, tracking a group in which members have, or will have, exposure or involvement with specific variables. For example, researchers might track behavioral problems among people following a specific natural disaster or the development of children living in communities with high levels of street violence. In medicine, researchers might track people exposed to the SARs virus. Researchers use such studies to determine the incidence of specific responses to the initial variable. While such variables are often stressors, cohort studies may also be used to track responses to positive events, such as inoculation programs or depression screen programs.

Graphically a cohort study can be represented as:

$$\begin{array}{cccccccc} \mathbf{X} & \mathbf{O}_1 & \mathbf{O}_2 & \mathbf{O}_3 & \mathbf{O}_4 & \mathbf{O}_5 & \mathbf{O}_6 & \mathbf{OR} \\ \mathbf{O}_1 & \mathbf{O}_1 & \mathbf{O}_2 & \mathbf{O}_3 & \mathbf{O}_4 & \mathbf{O}_5 & \mathbf{O}_6 & \end{array}$$

Here the **X** stands for exposure to a risk factor and **O** stands for each assessment. The exposure or event **X** may either mark the start of the study or may occur while assessments are ongoing. Participants are not randomly assigned which may introduce biases. Note, too, that there is no control or comparison group though studies of other people without the target exposure can serve as rough comparison groups.

In contrast to experimental studies with random assignment, participants in cohort studies may be selected with unknown strengths or challenges that, over time, affect the study results. Thus confounding variables can influence cohort study results. Over time, loss of participants may also bias study results. For instance, if

the more stressed participants drop out of a study, their loss may make the study results appear more positive than they would be if all participants continued to the study's conclusion. Because cohort studies are prospective in design, rather than retrospective, they are often viewed as stronger than are case–control studies. Cohort studies do not demonstrate cause and effect relationships, but can provide strong correlational evidence.

H. Case-Control Study

A retrospective, usually quantitative, observational study based on secondary data (or data already collected, often for different initial purposes). Looking back in time, case–control studies compare the proportion of cases with a potential risk or resiliency factor against the proportion of controls who do not have the same factor. For example, people who have very poor treatment outcomes for their anxiety disorder might be compared with a closely matched group of people who had very positive outcomes. A careful look at their demographic characteristics, medical histories and mental health histories might identify risk factors that distinguish most people in the two groups. Rare but important differences in risk or resiliency factors are often identified by such studies. Case–control studies are relatively inexpensive but are subject to multiple sources of bias if used to attribute cause to the risk or resiliency factors they identify.

I. Cross-Sectional Study (or Prevalence Study)

A descriptive, quantitative, study of the relationship between disorders or problems and other factors at a single point in time (Kazdin 2002). Cross-sectional studies are used descriptively in epidemiology. They can provide useful baseline information on the incidence of disorders in specific geographic areas. Cross-sectional studies are very valuable in a descriptive manner to policy planning, but do not demonstrate cause and effect relationships. An example of a cross-sectional study would be to look at the rate of poverty in a community during one month of the year. It is simply a snap shot picture of how many individuals would be classified as living in poverty during that month of the study. Comparing the number of persons in poverty with the total population of the community gives *prevalence rate* for poverty in this community.

J. Case Series

A descriptive, observational study of a series of cases, typically describing the manifestations, clinical course, and prognosis of a condition. Both qualitative and quantitative data are commonly included. Case series can be used as exploratory research to identify the features and progression of a new or poorly understood

disorder. They can be very useful in identifying culture-bound or context-specific aspects of mental health problems. Case series are inherently descriptive in nature. They are most often based on small and non-random samples. The results of case series may not generalize to all potential patients/clients. Despite its limitations, scholars point out that the case series is the most common study type in the clinical literature. It may be the type of study closest to real-world practice and the type of study practitioners can undertake easily.

In some EBM/EBP research design hierarchies, the case series are among the least valued form of clinical evidence, as they do not demonstrate that an intervention caused a specific outcome. They nonetheless offer a valuable method for making innovative information about new disorders or problems and new treatment methods available at an exploratory and descriptive level. One example of this type of research design in the Nurses' Health Study (Colditz et al. 1997). This study examined female nurses who worked at Brigham and Women's Hospital in Boston and who completed a detailed questionnaire every two years on their lifestyle, hormones, exercise, and more. Researchers did not intervene with these women in any way beyond the survey, but used the information compiled over several decades to identify trends in women's health. These results can then be cautiously generalized to other women or used to provide information on health trends that could be explored further through more intervention-based research (Colditz et al. 1997).

K. Case Study (or Case Report)

This design centers on use of descriptive evidence drawn from a single case (Kazdin 2002). Case studies may be the best research design for the identification of new clinical disorders or problems. They can be very useful forms of exploratory clinical research. They usually include the description of a single case, highlight the manifestations of the disorder, its clinical course, and outcomes of intervention (if any). Because case studies draw on the experiences of a single case, and often a single clinician, some researchers call them "anecdotal." This term is used to differentiate evidence based on multiple cases from that based on just a single case. It also implies that case study reports often lack the systematic pre- and post-assessment found in single case research designs. The main, and often major, limitation of the case study is that the characteristics of the single case may, or may not, be similar to other cases in different people and circumstances. Another key limitation is that reporting of symptoms, interventions, course of the problem, and outcomes may be piecemeal. This may be because the disorder is unfamiliar or unique in some way (making it worth publishing about). Yet since there are few widely accepted standards for case studies authors provide very different kinds and quality of information to readers (Spence 1982).

Case studies offer a valuable method for generating innovative information about new disorders or problems, and about new treatment methods on an

exploratory or formative basis. These ideas may become the starting point for future experimental studies.

We note again that case studies may be best available evidence found in an EBP search. If research based on other designs is not available, case study research may be used to guide practice decision-making.

L. Expert Opinion or Practice Wisdom

The EBM/EBP research design hierarchy reminds clinicians that expert opinion may not (necessarily) have a strong evidence base. This is not to say that the experience of supervisors, consultants, and talented colleagues has no valuable role in practice. It is simply to point out that they are not always systematic and may not work well for all clients in all situations. As research evidence, unwritten expert opinion lacks systematic testing and control for potential biases. This is why it is the least valued form of evidence in most EBM/EBP evidence hierarchies. Such studies may still be quite useful and informative to clinicians in specific circumstances. They serve to point out new ways of thinking and intervening that may be useful to specific clinical situations and settings.

Research Designs: Section Summary

This chapter has reviewed the range of research designs used in clinical research. The different types of research designs have different purposes and different strengths. These purposes range from exploratory, discovery-oriented purposes for the least structured designs like case studies, to allowing attribution of cause and effect relationships for highly structured experimental designs. This chapter has also explored the research design terminology used in EBM/EBP. Some of this terminology draws heavily on medical research and may be unfamiliar to persons trained in social work or social science research. Still, most key research design concepts can be identified despite differences in terminology.

The EBM/EBP research design hierarchy places great emphasis on research designs that can document that a specific treatment caused the changes found after treatment. This is an important step in determining the effectiveness or efficacy of a treatment. Many documents portray experiments, or RCTs, as the best form of evidence upon which to base practice decisions. Critical consumers of research should pay close attention to the kind of research designs used in the studies they examine.

Key reviews of outcome research on a specific topic, such as those from the Cochrane Collaboration and the Campbell Collaboration, use research design as a key selection criterion for defining high quality research results. That is, where little or no experimental or RCT research is available, the research summary may indicate there is inadequate research knowledge to point to effective treatments.

Some so-called 'empty' summaries pointing to *no* high quality research evidence on some disorders are found in the Cochrane Review database. This reflects their high standards and careful review. It also fails to state just what constitutes the best available evidence for practice. Empty reviews do not aid clinicians and clients in practice decision-making. They simply indicate that clinicians should undertake an article-by-article review of research evidence on their clinical topic. Clinicians must bear in mind that the EBP practice decision-making process promotes use of the best available evidence. If such evidence is not based on experimental research, it should be used, but used with caution. Still, it is entirely appropriate in the EBP framework to look for descriptive or case study research when there is no experimental evidence available on a specific disorder or concern.

Even when experimental or RCT research designs set the framework for establishing cause and effect relationships, a number of related methodological choices are also important to making valid knowledge claims. These include the quality of sampling, the quality of outcome measures, the definitions of the treatments used, and the careful use of the correct statistical tests. Adequate sample size and representativeness is important to generalizing study results to other similar people and settings. Appropriately conceptualized, valid, reliable, and sensitive outcome measures document baseline status and any changes. How treatments are defined and delivered will have a major impact on the merit and worth of study results. Statistics serve as a decision-making tool to determine if the results are unlikely to have happened by chance alone. All these methods work in combination to yield valid and rigorous results. These issues will be explored in the next chapter on appraising some additional methodological issues in practice research.