

Chapter 6

Step 3 of Evidence-Based Practice: Part 1—Evaluating Research Designs



Once you have located some research 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 of this research. An initial inspection of materials should help differentiate those that are generally relevant for your purposes from those that are not. Relevance may be initially determined by examining the research question that each study addresses. Studies should have clear and relevant research questions, fitting your practice needs. Once these “apparently relevant” studies are identified, the appraisal shifts to issues of research methodology. Even studies that appear quite relevant initially may later on prove to have important limitations as the details of their methods are explored.

Evaluating the quality of research reports can be a complex process. It involves several components. We will begin by reviewing research designs used in EBP. While many of these designs should be familiar to social workers, they may be described using different terminologies in EBM and EBP research reports (Drisko, 2011). Chapter 7 will review several other methodological steps in appraising research (sampling, defining the treatment or other intervention, test and measures, and statistics). These provide the basis for examining meta-analysis and systematic reviews, two widely used methods for aggregating research results in EBM and EBP, examined in Chap. 8.

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.* Different research designs are used for research 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. Doing EBP, however, will require that clinical social workers and

other mental health professionals make greater use of their knowledge about evaluating research for practice.

Research designs are so important to EBM/EBP that this chapter will focus on them exclusively. Other very important—and very closely related—aspects of research methods will be examined in the following chapter (sampling, measures, definitions of treatments, and analysis). 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. Still, some terminology, drawn from medicine, will no doubt be unfamiliar. For others who need only an update, this chapter offers it. Many excellent follow-up resources are identified in each section of the chapter.

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. This terminology sometimes 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 design quality to guide practice decision-making. This practice contradicts the idea of identifying the best available evidence. In effect, the best available evidence is reduced to evidence generated by experimental research designs. This practice creates confusion about what constitutes the best available evidence for clinicians, policy planners, and researchers.

Some EBP/EBM authors do not report *all* the best available evidence, but instead report only the 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 Chap. 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’s Centre for Evidence-Based Medicine (2009, 2016). 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, then move to observational or descriptive research, and end with case studies. The organization of this section follows the format of the research evidence hierarchy created by Oxford University’s Centre for Evidence-Based Medicine (2009, 2100; 2016, 2018).

Types of Clinical Studies

Part 1: Experimental Studies or RCTs

EBP researchers view properly conceptualized and executed experimental studies. These are also called randomized controlled trials or RCTs. RCTs provide internally valid 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 and sources of error to each group. Probability samples further provide a suitable foundation for most statistical analytic procedures.

The key benefit of an experimental research design is that they minimize threats to internal validity (Campbell & 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 highly regarded in the EBM/EBP model. The main limitations of experiments are their high cost in money, participation, effort, and time. They may be ethically inappropriate for some studies where random assignment is inappropriate. 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 (Oxford Centre for Evidence-Based Medicine, 2019).

In the European medical literature, experiments and quasi-experiments may alternately 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 PICO or PICOT treatment decision-making model (Oxford Centre for Evidence-Based Medicine, 2019).

The Randomized Controlled Trial (RCT) or Classic Experiment

It is a quantitative, prospective, group-based study based on primary data from the clinical environment (Solomon, Cavanaugh, & Draine, 2009). Researchers randomly assign individuals who have the same disorder or problem at the start to one of two (or more) groups. Later, the outcomes for each group are compared at the completion of treatment. Since researchers create the two groups by random assignment to generate two 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 sometimes 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 does 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 any 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:

$$\begin{array}{cccc} R & O_1 & X & O_2 \\ R & O_1 & & O_2 \end{array}$$

where R stands for random assignment of participants, O_1 stands from the pretest assessment (most often with a standardized measure), X represents the intervention given to just one group, and O_2 stands for the posttest, done after treatment, but using the same measure. There may also be additional follow-up posttests to document how results vary over time. These would be represented as O_3 , O_4 , etc. There may be two or more groups under comparison in an RCT. Further, more than one measure of outcome may be 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 internal validity of the study results. A *double blind* RCT design keeps even 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.

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 & Loewenstein, 2003). This phenomenon is known as *attribution bias*, in which people invested in a particular theory or treatment model view it more positively than do others. 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. 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. However, most Cochrane 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 & Green, 2018; Littell, Corcoran, & Pillai, 2008). In addition, 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 may have serious limitations despite their use of a “strong” research design. Sample size is one such issue. Many clinical studies compare small groups (roughly under 20 people in a group). Studies using small samples may lack the statistical power to identify any 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 impacts research quality later in the next two chapters.) 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 <http://www.consort-statement.org/>. The CONSORT Statement is available in many different languages. The CONSORT group also provides a free template for a flow chart of the RCT process and statement. These tools can be very helpful to the consumer of experimental research since they serve as guides for assessing the quality of RCTs. A CONSORT flow chart (also called a Quorum chart) is often found in published reports of recent RCTs.

The Randomized Crossover Clinical Trial

It is a prospective, group-based, quantitative, experimental study 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 “washout” 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 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. This strengthens the internal validity of the study. 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 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 not wash out. Similarly, it may not be possible to wash out learned or internalized cognitions, skills, attitudes, or behaviors. This is a limitation of crossover research designs in mental health and social services.

The merit of crossover 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, known as statistical power. All participants receive both treatments, which benefits them. Random assignment provides a solid foundation for statistical tests. Disadvantage of crossover studies includes that all participants receive a placebo or less effective treatment at some point which may not benefit them immediately. Further, washout periods can be lengthy and curtail active treatment for the washout period. Finally, crossover designs cannot be used where the effects of treatment are permanent, such as in educational programs or surgeries.

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

$$A_1 \quad B_1 \quad A_2 \quad B_2 \quad A_3$$

where A_1 stands for the initial assessment, B_1 represents the first intervention given, the second A_2 represents the next assessment which is made at the end of the first intervention after washout, and B_2 stands for second type of intervention or the crossover. Finally, A_3 represents the assessment of the second intervention done when it is completed. Note that a washout period is not specifically included in this design but may be if the researchers chose to do so. 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 crossover study.

Since random assignment is not possible with single cases, the results of single-case crossover 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, personal history, 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 build confidence in the results (Anastas, 1999). It is in contrast to *sampling logic* used in group experimental designs 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 and address potentially confounding variables. If treatment outcomes are positive over many cases, treatment effectiveness may be inferred. In EBM, single-case studies are not designated as providing strong research evidence, but consistent findings from more than ten single-case study outcomes are rated as strong evidence in the American Psychological Association’s designation of empirically supported treatments (ESTs).

The Randomized Controlled Laboratory Study

It is 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 “basic” research since all other variables of 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 real-world 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 vs. Efficacy Studies: Experiments Done in Different Settings

In mental health research, a distinction is drawn between clinical research done in the 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 comorbid disorders and may not be able to ensure that treatments are provided fully and consistently. This reduces their internal validity for research purposes. On the other hand, using real-world settings enhances their external validity, meaning that the results are more likely to fit with actual practice with everyday clients and settings. 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 is effective (or are not effective).

One well-known example of a clinical efficacy study is the NIMH Cross-site Study of Depression (Elkin, Shea, Watkins, 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 (often for reasons of funding 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 steps. 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 practice (given its ethical, funding, and practical limitations).

Part 2: Quasi-experimental and Cohort Studies—Comparisons Without Random Participant Assignment

Random assignment of participants to treated versus control groups is a way to strengthen internal validity and to limit bias in research results. Random assignment ideally generates (two or more) equivalent groups for the comparison of treatment effects versus an untreated control group. Quasi-experimental research designs lack

random assignment but do seek to limit other threats to the internal validity of study results. They are often used where random assignment is unethical or is not feasible for practical reasons.

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 or control groups. For example, policy-makers may only fund a new type of therapy or a new prevention program for a single community or with payment by only certain types of insurance. In such situations, researchers use existing groups or available groups to examine the impact of interventions. The groups, settings, or communities to be compared are chosen to be as similar as possible in their key characteristics. The goal is to approximate the equivalent groups created by random assignment. Where pre- and post-comparisons are done on such similar 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 or control groups.

The quasi-experiment can be summarized graphically as:

$$\begin{array}{ccc} O_1 & X & O_2 \\ O_1 & & O_2 \end{array}$$

Once again, O_1 stands from the pretest assessment (most often with a standardized measure), X represents the intervention given to just one group, and O_2 stands for the posttest, done after treatment, but using the same measure. More than two groups may be included in a quasi-experimental study. There may also be additional follow-up posttests to document how results vary over time. More than one measure may be used in the same quasi-experiment. Note carefully that the key difference from a true experiment is 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 studied groups may have important differences that are not equally distributed across the groups. 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 was used. Thus in some EBM/EBP hierarchies of research evidence, quasi-experimental study results are rated as “weaker” than are results of true experiments or RCTs. That said, they are still useful sources of knowledge and are often the best available research

evidence 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 variables are fully known at the start of the study.

Advantages of quasi-experimental or cohort studies include their ethical appropriateness in that participants are not assigned to groups and can make their own personal treatment choices on an informed basis. Cohort studies are usually less expensive in cost than are true experiments, though they may both be financially 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.

The “All or None” Study

The Centre for Evidence-Based Medicine at Oxford University (2009, B13) includes in its rating of evidence the “All or None” research design. This is a research design in which, in very difficult circumstances, clinicians give an intervention to a group of people at high risk of serious harm or death. If essentially all the people who received the intervention improve or survive, while those who do not receive it continue to suffer or die, the inference is that the intervention caused the improvement. This is actually an observational research design, but given the nature of the groups compared, all or none results are viewed as strong evidence that the treatment caused the change. However, given their very important effects, such research results are highly valued so long as all or a large fraction of people who receive the intervention improve. Such designs fit crisis medical issues much better than most mental health issues, so all or none design is extremely rare in the mental health literature. They do have a valuable role in informing practice in some situations.

Part 3: Non-interventive Research Designs and Their Purposes

Not all practice research is intended to show an intervention *causes* a change. While EBM/EBP hierarchies of research evidence rank most highly, those research designs that do show an intervention cause 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.

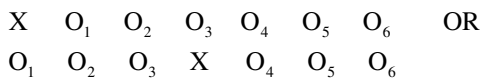
The Observational Study

It is a prospective, longitudinal, usually quantitative, tracking study of groups or of individuals with a single disorder or problem (Kazdin, 2010). 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.

The Longitudinal Study

It is a prospective, quantitative and/or qualitative, observational study ideally based on primary data, tracking a group in which members have had, or will have, exposure or involvement with specific variables. For example, researchers might track the development of 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. Longitudinal studies help identify the probability of occurrence of a given condition or need within a population over a set time period. 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 longitudinal study can be represented as:



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 longitudinal studies may be selected with unknown strengths or challenges that, over time, affect the study results. Thus, confounding variables can influence longitudinal study results. Over time, loss of participants may also bias study results. For instance, if the more stressed participants dropout 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 longitudinal studies are prospective in

design, rather than retrospective, they are often viewed as stronger than are case-control studies. Longitudinal studies do not demonstrate cause and effect relationships but can provide strong correlational evidence.

Case-Control Study

It is a retrospective, usually quantitative, observational study often 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 that do not have the same factor. For example, people who have very poor treatment outcomes for their anxiety disorder may 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 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.

Cross-Sectional Study or Incidence Study

These are descriptive, usually quantitative, studies of the relationship between disorders or problems and other factors at a single point in time. Incidence designs are used descriptively in epidemiology. They can be useful for learning baseline information on the incidence of disorders in specific areas. Cross-sectional studies are very valuable in a descriptive manner to policy planning, but do not demonstrate cause and effect relationships. They are not highly valued in the EBM/EBP research design hierarchy. An example of a cross-sectional study would be to look at the rate of poverty in a community during 1 month of the year. It is simply a snapshot 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 the *incidence rate* or *prevalence rate* for poverty.

The Case Series

It is 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,

but they are most often based on small and nonrandom samples. The results of case series may not generalize to all potential patients/clients.

Despite its limitations, many scholars point out that the case series is the most frequently found research design in the clinical literature. It may be the type of study most like real-world practice and is a 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 is the Nurses' Health Study (Colditz, Manson, & Hankinson, 1997). This is a study of female nurses who worked at Brigham and Women's Hospital in Boston and who completed a detailed questionnaire every second years on their lifestyle, hormones, exercise, and more. Researchers did not intervene with these women in any way but have used the information compiled by the study over several decades to identify trends in women's health. These results can then be 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).

The Case Study (or Case Report)

It is a research design using descriptive but "anecdotal" evidence drawn from a single case. The data may be qualitative and/or quantitative. 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, highlighting 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, they are often labeled "anecdotal." This differentiates evidence collected on multiple cases from that based on just a single case. Further, case study reports often lack the systematic pre-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), but since there are few widely accepted standards for case studies, authors provide very different kinds and quality of information to readers.

Case studies offer a valuable method for generating innovative information about new disorders or problems, even new treatment methods, available 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.

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 experiences of supervisors, consultants, and talented colleagues have 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 planned and 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 to new ways of thinking and intervening that may be valuable to specific clinical situations and settings.

Resources on Research Design in EBP

Many textbooks offer good introductions to research design issues and offer more illustrations than we do in this chapter. Note, however, that the terminology used in EBM/EBP studies and summaries may not be the same as is used in core social work textbooks. Resources addressing issues in research design are found in Table 6.1.

Table 6.1 More resources on research design

Anastas, J. W. (1999). <i>Research design for social work and the human services</i> (2nd ed.). New York: Columbia University Press.
Campbell, D., & Stanley, J. (1963). <i>Experimental and quasi-experimental designs for research</i> . New York: Wadsworth.
Creswell, J., & Creswell, J. D. (2018). <i>Research design: Qualitative, quantitative, and mixed methods approaches</i> . Thousand Oaks, CA: Sage.
Drisko, J. (2011). Researching Clinical Practice. In J. Brandell (Ed.). <i>Theory and practice in clinical social work</i> (2nd ed., pp 717–738). Thousand Oaks, CA: Sage.
Hulley, S., Cummings, S., Browner, W., Grady, D., & Newman T. (2018). <i>Designing clinical research</i> (4th ed.). Philadelphia, PA: Lippincott, Williams & Wilkins.
Kazdin, A. (2010). <i>Single-case research designs: Methods for clinical and applied settings</i> . (2nd ed.). New York: Oxford.
Rubin, A. (2008). <i>Practitioner’s guide to using research for evidence-based practice</i> . Hoboken, NJ: Wiley.
Soydan, H., & Palinkas, L. (2014). <i>Evidence-based practice in social work: Development of a new professional culture</i> . New York: Routledge.
Yin, R. K. (2018). <i>Case study research and applications: Design and methods</i> . Thousand Oaks, CA: Sage.

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 for practice application.

Key reviews of outcome research on a specific topic, such as those from the Cochrane Collaboration and 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. “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. 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 the use of “the best available evidence.” If such evidence is not based on experimental research, it should still be used, but used with caution. 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 also are important to making valid knowledge claims. These include the quality of sampling, the inclusion of diverse participants in the sample, the quality of the outcome measures used, the definitions of the treatments, and the careful use of the correct statistical tests. Adequate sample size and representativeness are important to generalizing study results to other similar people and settings. Appropriately conceptualized, valid, reliable, and sensitive outcome measures document 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

tandem to yield valid and rigorous results. These issues will be explored in the next two chapters on Step 3 of the EBP process, further appraising some additional methodological issues in practice research.

References

- Anastas, J. W. (1999). *Research design for social work and the human services* (2nd ed.). New York: Columbia University Press.
- Campbell, D., & Stanley, J. (1963). *Experimental and quasi-experimental designs for research*. New York: Wadsworth.
- Cochrane, A. (1972). *Effectiveness and efficiency: Random reflections on health services*. London: Nuffield Provincial Hospitals Trust.
- Colditz, G., Manson, J., & Hankinson, S. (1997). The Nurses' health study: 20-year contribution to the understanding of health among women. *Journal of Women's Health*, 6(1), 49–62.
- CONSORT Group. (2010). *The CONSORT statement*. Retrieved from <http://www.consort-statement.org/home/>
- Dana, J., & Loewenstein, G. (2003). A social science perspective on gifts to physicians from industry. *Journal of the American Medical Association*, 290(2), 252–255. <https://doi.org/10.1001/jama.290.2.252>
- Drisko, J. (2011). Researching clinical practice. In J. Brandell (Ed.), *Theory and practice in clinical social work* (2nd ed., pp. 717–738). Thousand Oaks, CA: Sage.
- Elkin, I., Shea, T., Watkins, J., et al. (1989). National Institute of Mental Health treatment of depression collaborative research program: General effectiveness of treatments. *Archives of General Psychiatry*, 46, 971–982.
- Higgins J., & Green, S. (Eds.). (2018). *Cochrane Handbook for Systematic Reviews of Interventions* Version 6.0 [updated September 2018]. The Cochrane Collaboration, 2011. Available from <http://handbook.cochrane.org>
- Kazdin, A. (2010). *Research design in clinical psychology* (5th ed.). New York: Pearson.
- Littell, J. (2011, January 15). *Evidence-based practice versus practice-based research*. Paper presented at the Society for Social Work and Research, Tampa, FL.
- Littell, J., Corcoran, J., & Pillai, V. (2008). *Systematic reviews and meta-analysis*. New York: Oxford University Press.
- Oxford University Centre for Evidence-based Medicine. (2009, April). *Asking focused questions*. Retrieved from <http://www.cebm.net/index.aspx?o=1036>
- Oxford University Centre for Evidence-based Medicine. (2016, May). *The Oxford levels of evidence 2.1*. Retrieved from <https://www.cebm.net/2016/05/ocebml-levels-of-evidence/>
- Oxford Centre for Evidence-Based Medicine. (2019). Study designs. Retrieved from <https://www.cebm.net/2014/04/study-designs/>
- Solomon, P., Cavanaugh, M., & Draine, J. (2009). *Randomized controlled trials*. New York: Oxford University Press.
- Thyer, B. (2011, January 15). *Evidence-based practice versus practice-based research*. Paper presented at the Society for Social Work and Research, Tampa, FL.