

1

The Evaluation Research and Evidence-Based Practice Partnership

Evaluation research relies on the scientific method to provide valid evidence on the outcomes, impact, and costs of programs to improve the public's health, education, and welfare. But valid evidence alone may be insufficient to meet the needs and expectations of clients, patients, and other users of the evidence. Evaluators have identified methods for determining those needs and using their expertise to incorporate user values and expectations into their study's purposes and methods.

Evidence-based medicine and its sibling evidence-based practice (EBM/EBP) are concerned with using experience and clinical judgment to integrate best evidence from research with patient values and experience. EBM/EBP practitioners have developed special methods for identifying research studies and grading the quality and strength of the evidence. By design, EBM/EBP combines the best research evidence and brings it together with clinical and professional expertise and patient or client values to make clinical decisions.

This chapter provides the **research consumer** with an overview of the theoretical and methodological foundations of evaluation research and EBM/EBP and discusses how to use them in identifying evidence that matters. Research consumers use research as the basis for making decisions about programs, practices, and policy. Consumers practice research and are concerned with the practical applications of research findings.

Chapter Objectives

After reading this chapter, you will be able to

- Define evaluation research and EBM/EBP
- Distinguish evaluation research from other types of evaluation
- Explain how the research consumer uses evaluation research and EBM
- Distinguish evaluation research from other types of social research
- Find the outcomes and hypotheses in selected evaluation studies
- Identify the five steps of EBM
- Compare the similarities and differences between evaluation research and EBM/EBP

Figure 1.1 shows your location on the way to discovering evidence that matters.

Evaluation Research: What It Is and What It Is Not

Human problems are sometimes solved by enlisting people's participation in programs to improve their health, quality of life, social well-being, and economic prospects. A **program** consists of activities and resources that have been specifically selected to achieve these beneficial **outcomes**. An example of a program is a 10-session cognitive behavioral intervention (the program) to reduce children's symptoms of posttraumatic stress disorder (the outcome) resulting from exposure to violence. Another example is a 15-minute Web-based education tool (the program) to teach older adults about the risks of alcohol drinking (the outcome).

Evaluation researchers use scientific methods to assess the process, outcomes, impact, or costs of programs and to provide new knowledge about social behavior. Program processes refer to the staff, activities, materials, and methods that are used to accomplish the outcomes. Assessing the process includes evaluating the characteristics of the program's leadership, the adequacy of the in-service training the staff receives, the appropriateness of newly configured classrooms and offices, the effectiveness of the length of each unit of instruction, and the technological and other resources available.

Evaluations that focus on the program (What happens? Who is responsible? What does the program look like?) rather than on the

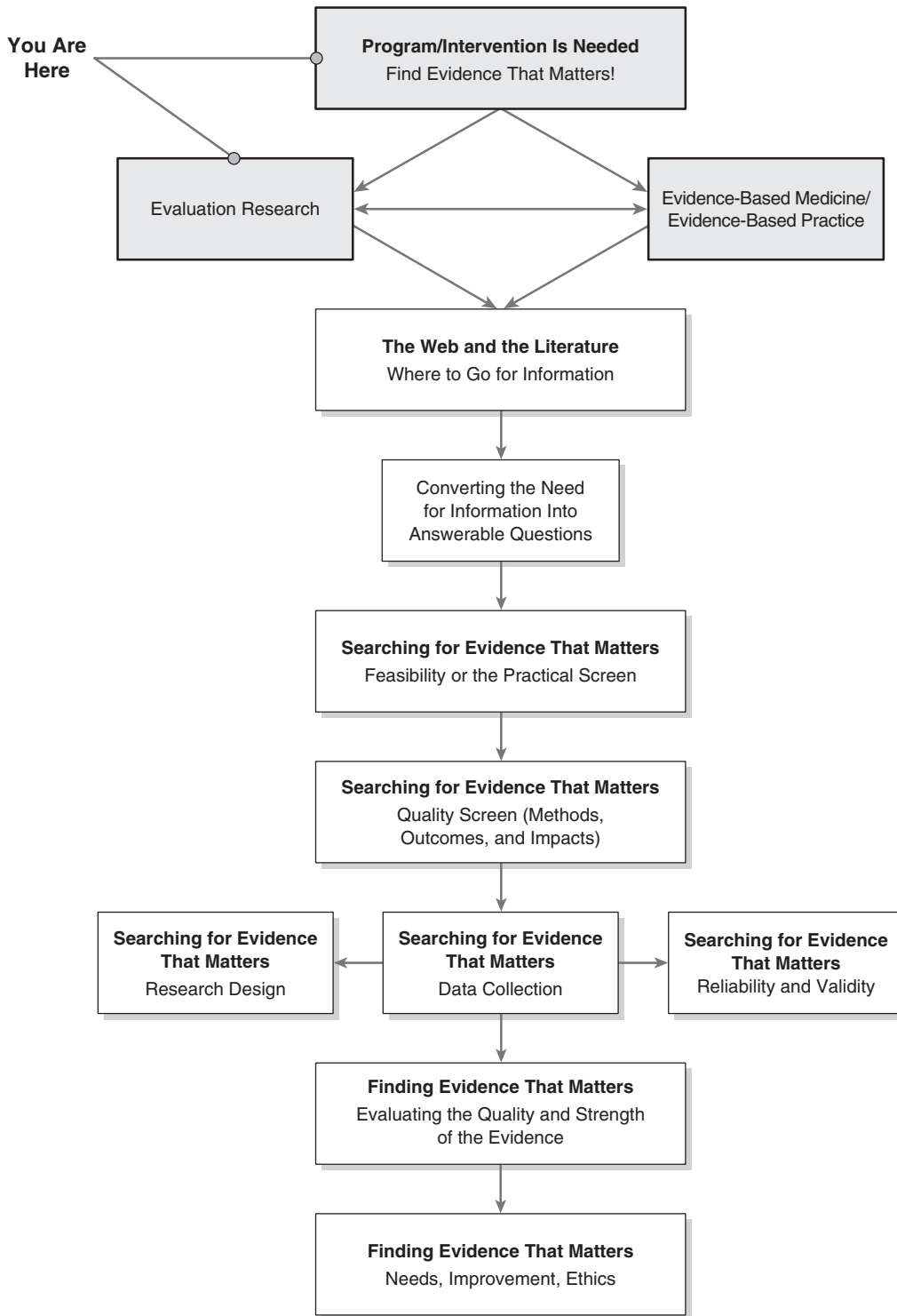


Figure 1.1 Location on the Way to Discovering Evidence That Matters

outcomes of participation are called by various names including **implementation**, **process**, and **formative** evaluations. In some professional fields (e.g., education, community research), formative evaluations are contrasted with **summative** evaluations, which examine outcomes and impact and are completed after a substantial portion (if not all) of the program's activities are completed.

A program's **impact** is its magnitude and duration. An evaluation researcher studying a cognitive behavioral therapy intervention's impact, for example, might assess the number of children (magnitude) who were beneficially affected (improvement in symptoms) and how long the benefits lasted (duration).

Evaluation research is a subdivision of the much larger field of program evaluation, which has been described by the American Evaluation Association or AEA (www.eval.org) as a profession composed of persons with varying interests, potentially encompassing but not limited to the evaluation of programs, products, personnel, policy, performance, proposals, technology, research, theory, and even of evaluation itself.

Evaluation research shares many of the purposes that are delineated by the AEA in connection with other forms of program evaluation. These include bettering products and practices, personnel, programs, organizations, governments, consumers, and the public interest; contributing to informed decision making and more enlightened change; precipitating needed change; and empowering all stakeholders by collecting data from them and engaging them in the evaluation process.

The main difference between the evaluation researcher and many other evaluators is that the researcher insists on obtaining the "best" evidence of program merit, which means strict adherence to the highest possible research standards. Many evaluators have a different focus than the research evaluator and tend to be more concerned with other aspects of the job, such as precipitating change or engaging others in the evaluation process. As you will see later, the research emphasis overlaps that of EBM's, which is defined by its explicit use of the "best" evidence in making decisions, evidence that comes from basic and clinical research.

When evaluators do **research**, they are participating in diligent and systematic processes of inquiry aimed at discovering, interpreting, and revising information about programs and interventions (terms that will be used here interchangeably). Research is also a term that is used to describe a collection of information about a particular subject, and it is associated with the **scientific method**. The scientific

method is a set of techniques for investigating phenomena and acquiring new knowledge of the natural and social worlds, based on observable, measurable evidence.

The scientific method is also characterized by the belief that a study's activities must be objective so that the scientist cannot bias the interpretation of the results or change the results outright. Another basic expectation is that the researcher will make available complete documentation of the data and the methodology for careful scrutiny by other scientists and researchers, thereby allowing them the opportunity to duplicate and verify the results. Enabling this replication of results is a scientific and an ethical imperative.

In fact, the field of **ethics**, also called moral philosophy, is directly associated with scientific research. Ethics involves systematizing, defending, and recommending concepts of right and wrong behavior. Because evaluations always include human participants, the evaluator must demonstrate that the study design attends to ethical principles and respects participants' privacy, ensures that the benefits of participation are maximized, and provides all participants with equal access to the benefits. The criteria for including and excluding participant must be justified, and there must be a sufficient number of participants so that a program has a chance to prove itself. Also, the data collection and analysis must be appropriate and valid. Research that is not sound is unethical in itself because it results in misleading or false conclusions that, when applied, may result in harm.

Evaluation researchers rely on the scientific method, a characteristic they share with all social researchers who strive for the "truth." The main difference is that evaluation researchers study the effects of programs and interventions on participants whereas other social scientists most typically do not. More often than not, other social researchers focus on studies that describe relationships and predict events.

Example 1.1 gives samples of evaluation and other types of social research.

Evaluation research can mimic other social research when it is designed to provide new knowledge about human behavior as well as to provide evidence of program effectiveness. Research objectives C and D in Example 1.1, for example, which are research objectives, might also be achieved in connection with an evaluation study. Objective C, for example, might be achieved in connection with a program to improve children's ability to cope with pain. The achievement of objective D might occur as part of an evaluation of a program to train nurses and mental health nurse assistants to deal with seriously mentally ill hospitalized patients.

Example 1.1 Evaluation Research: Yes? No?

A. Research Objective: To investigate the effectiveness of acupuncture compared with sham acupuncture and with no acupuncture in patients with migraine.

Is this objective likely to be consistent with evaluation research purposes?

Answer: Yes. The researchers compare three interventions: acupuncture, sham acupuncture, and no acupuncture. (No acupuncture is considered an “intervention” in this case because the absence of acupuncture does not mean the absence of anything at all. The no acupuncture group may be on medication or other forms of therapy.)

B. Research Objective: To determine the effectiveness of an abuse-prevention curriculum designed to empower women with mental retardation to become effective decision makers.

Is this objective likely to be consistent with evaluation research purposes?

Answer: Yes. The intervention in this study is an abuse-prevention *curriculum*.

C. Research Objective: To clarify the concepts of coping with pain and quality of life (QoL) and to present a literature review of the strategies that children with recurrent headaches use to cope with their pain, the impact of recurrent headaches on children’s QoL, and the influence of personal characteristics (such as age, family support) on headache, coping, and QoL in children.

Is this objective likely to be consistent with evaluation research purposes?

Answer: No. The researchers are not planning to analyze the process, outcomes, impact, or cost of a program or intervention.

D. Research Objective: To describe nurses’ and mental health nurse assistants’ perceptions of advantages and disadvantages of working on a psychiatric ward with a locked entrance door.

Is this objective likely to be consistent with evaluation research purposes?

Answer: No. The researchers are not planning to analyze the process, outcomes, impact, or cost of a program or intervention.

Although every discipline has its own specialized methodology and terminology, most evaluation researchers and other scientists begin their inquiries by asking **research questions** and proposing specific hypotheses as explanations of events. A research question is the objective of the evaluation, the uncertainty that the evaluator wants to diminish. Research questions often begin with a general concern. For example,

Can an educational program designed by and for teens reduce alcohol-related risks in adolescents?

Questions like these are usually refined even more so as to guide evaluation planning. For example,

How frequent is teen drinking in this county?

How willing are teens to participate in the design of a program to reduce alcohol-related risks?

How willing are teens to participate in an experimental study to reduce alcohol-related risks? The study's main activities will take place after school hours.

A research question is sometimes translated into one or more hypotheses. A hypothesis is a suggested explanation of a set of events, an explanation that is usually based on whether a relationship does not (or does) exist between these events. A typical evaluation hypothesis concerns the relationship between people who participate in a program and those who do not. Sample hypotheses for a program to reduce symptoms of depression among children in schools are given in Example 1.2.

Example 1.2 Two Sample Hypotheses for a Program to Reduce School Children's Symptoms of Depression

1. Students who receive a brief standardized program that is delivered by school mental health clinicians on school campuses will have significantly fewer self-reported symptoms of depression, and fewer reports of psychosocial dysfunction by parents at the 3-month assessment, than students who are randomly assigned to receive another program.
2. Improvement in symptoms of depression will not be translated into significant improvements in the classroom behavior as reported by teachers.

Hypotheses are tested in experimental studies. Does the relationship hold up under intense scrutiny? A controlled experiment generally compares the outcomes obtained from an experimental group against those from a control group, which is practically identical to the experimental group except for the one aspect (the experimental program) whose effect is being tested.



CAUTION: Research consumers should always check each study's research questions, objectives, and hypotheses to ensure that the outcomes and the study participants are relevant to their clients' needs. You may come upon evidence that supports the merits of a program that aims to achieve objectives like yours (e.g., reduce symptoms of depression) but that comes from a population (children) unlike yours (adults). It is also possible that claims of effectiveness are based on a population that interests you but that the outcomes (e.g., reduce symptoms) are not the ones of primary interest to you (e.g., provide knowledge about options for treatment).

A sample experimental study is outlined in Example 1.3. In the example, you will see that the experimental and control programs differ but that the objectives, outcomes, and participants do not. The idea of an experiment is that, if benefits are found over time in the experimental group, then it is fair to reason that they are due to the experimental program and not to group particularities (e.g., greater or lesser motivation to reduce risks).

Example 1.3 An Experimental Study to Reduce Alcohol-Related Risks: Comparing Experimental and Control Programs, Objectives, Participants, and Expected Outcomes

<i>Program</i>	<i>Objectives</i>	<i>Participants</i>	<i>Expected Outcomes</i>
Teen Alcohol Risk Reduction Program	Reduce risks of alcohol drinking	16 to 21 years of age Report drinking at least one alcoholic beverage in the past 4 weeks No history of alcohol or drug abuse	Risks are reduced as measured by Alcohol Risk Measure scores
Schools' Usual Alcohol Education	Reduce risks of alcohol drinking	16 to 21 years of age Report drinking at least one alcoholic beverage in the past 4 weeks No history of alcohol or drug abuse	Risks are reduced as measured by Alcohol Risk Measure scores



Not all evaluations are experimental. Some are **observational**. In observational research, the researcher takes a relatively passive role and does not introduce a new program but observes an already existing one. For example, a researcher who wants to find out why some teens are at risk for alcohol-related problems while others are not may interview students, some of whom participated in an alcohol risk reduction program about a year ago. The researcher would then analyze the interview data to uncover factors (e.g., participation in a program, age, family history of drinking) that predict risk. One question this researcher is likely to ask: “Do students who participated in the program have the same, greater, or lower risks than students who did not participate?” Observational studies are generally not as preferable as experimental ones because the researchers have no control over the criteria for participation in the program or other aspects of the program’s design or implementation.

Evaluations that use the scientific method are able to provide accurate answers to questions such as the following. Have all planned program outcomes been achieved? What are the resources that are needed for the program to achieve its outcomes? Have all program participants benefited from the achievement? How long must participants wait for benefits to occur? How long do the program’s effects last? Do the benefits of the program outweigh the risks to individuals? Do the benefits of program participation outweigh the costs to the public? What difference does the program make when compared to a logical alternative program?

Collecting information about the program and its costs and effects on participants is at the heart of evaluation. Data collection includes (1) identifying the **variables** that will be measured. A variable is a term given to a thing, or a certain value, that may change, such as the knowledge, performance, attitudes, or behaviors that describe the program participants’ characteristics before and after being in the program. Data collection also involves (2) selecting, adapting, or creating measures like surveys and tests; (3) demonstrating the reliability (consistency) and validity (accuracy) of the measures; (4) administering the measures so that the rights of “human subjects” are respected; (5) “scoring” or assigning a value to the results; and (6) interpreting the score. Evaluators are old hands at the science that underlies the validation of measures of important and often difficult to assess factors, such as quality of life and well-being.

Evaluations often result in voluminous amounts of data that require careful management. Data management includes data entry and storage and requires thinking about methods of ensuring the accuracy of entry and setting up a system for ensuring confidentiality.

Evaluation data analysis can be extremely complex and include **quantitative** and **qualitative** methods. Quantitative methods rely on mathematical and statistical models. Qualitative methods involve investigating participants' opinions, behaviors, and experiences from their point of view and using logical induction. Most evaluation *research* relies heavily on quantitative methods to answer research questions and test hypotheses.

Evaluations are almost always conducted in “real life” situations in which some eligible people may not participate in all of the program's activities, others may participate in competing activities, and still others may drop out. That is why evaluations are sometimes referred to as **effectiveness** rather than **efficacy** studies, which are done under ideal conditions. Inability to control the environment and implement perfect research designs have led evaluation researchers to find ways of shoring up study validity by developing and improving upon existing research methods (e.g., propensity score analysis to control for baseline differences in study groups). These research innovations are increasingly being used for the same reason by other social scientists and health researchers.

- Evaluators are always on the lookout for new research methods. They are not shy, and they often borrow from other fields without asking for permission. Do not be surprised if you find ideas in evaluation research that come from disciplines as diverse as psychology, health services research, clinical research, sociology, communications, economics, and epidemiology, to name a few. In fact, one of the skills research consumers should acquire is how to evaluate the research from other fields upon which evaluation studies rely. In Chapter 6, for example, we will review methods for assessing whether the outcome measures that evaluators choose are reliable and valid.

- Evaluation researchers sometimes look to **clinical trials** for ways of analyzing data. Clinical trials are evaluations of medical and surgical treatments, drugs, and other interventions that are conducted in health care settings like clinics and hospitals. The outcomes investigated in clinical trials may be medical (e.g., reductions in blood pressure), psychosocial (e.g., improvements in health-related quality of life), and economic (e.g., which of two equally effective programs costs less).

Example 1.4 provides two **abstracts** of evaluation studies. An abstract is an abbreviated version of the objectives, methods, findings, and conclusions of a much larger report. For now, you do not have

Example 1.4 Evaluation Research in the Abstract

1. An Evaluation of an Intervention to Reduce Alcohol-Related Risks and Problems in Older Adults (Fink, Elliott, Tsai, & Beck, 2005)

Objectives: To evaluate whether providing physicians and older patients with personalized reports of drinking risks and benefits and patient education reduces alcohol-related risks and problems.

Design: Prospective comparison study.

Setting: Community primary care.

Participants: Twenty-three physicians and 665 patients aged 65 and older.

Intervention: Combined report, in which 6 physicians and 212 patients received reports of patients' drinking classifications and patients also received education; patient report, in which 245 patients received reports and education, but their 5 physicians did not receive reports; and usual care.

Measurements: Assessments at baseline and 12 months later to determine patients' nonhazardous (no known risks), hazardous (risks for problems), or harmful (presence of problems) classifications using the Computerized Alcohol-Related Problems Survey (CARPS). The CARPS contains a scanned screening measure and scoring algorithms and automatically produces patient and physician reports and patient education.

Results: At baseline, 21% were harmful drinkers, and 26% were hazardous drinkers. The patient report and combined report interventions were each associated with greater odds of lower-risk drinking at follow-up than usual care (odds ratio=1.59 and 1.23, respectively, $P<.05$ for each). The patient report intervention significantly reduced harmful drinking at follow-up from an expected 21% in usual care to 16% and increased nonhazardous drinking from 52% expected in usual care to 58%. Patients in the combined report intervention experienced a significantly greater average decrease in quantity and frequency.

Conclusion: Older primary care patients can effectively reduce their alcohol consumption and other drinking risks when given personalized information about their drinking and health.

SOURCE: Fink, Elliott, Tsai, and Beck (2005).

2. A Psychological Intervention for Children With Symptoms of Posttraumatic Stress Disorder (Stein et al., 2003)

Context: No randomized controlled studies have been conducted to date on the effectiveness of psychological interventions for children with symptoms of posttraumatic stress disorder (PTSD) that has resulted from personally witnessing or being personally exposed to violence.

(Continued)

Example 1.4 (Continued)

Objective: To evaluate the effectiveness of a collaboratively designed school-based intervention for reducing children's symptoms of PTSD and depression that has resulted from exposure to violence.

Design: A randomized controlled trial conducted during the 2001–2002 academic year.

Setting and Participants: Sixth-grade students at 2 large middle schools in Los Angeles who reported exposure to violence and had clinical levels of symptoms of PTSD.

Intervention: Students were randomly assigned to a 10-session standardized cognitive-behavioral therapy (the Cognitive-Behavioral Intervention for Trauma in Schools) early intervention group ($n = 61$) or to a wait-list delayed intervention comparison group ($n = 65$) conducted by trained school mental health clinicians.

Main Outcome Measures: Students were assessed before the intervention and 3 months after the intervention on measures assessing child-reported symptoms of PTSD (Child PTSD Symptom Scale; range, 0–51 points) and depression (Child Depression Inventory; range, 0–52 points), parent-reported psychosocial dysfunction (Pediatric Symptom Checklist; range, 0–70 points), and teacher-reported classroom problems using the Teacher-Child Rating Scale (acting out, shyness or anxiousness, and learning problems; range of subscales, 6–30 points).

Results: Compared with the wait-list delayed intervention group (no intervention), after 3 months of intervention, students who were randomly assigned to the early intervention group had significantly lower scores on symptoms of PTSD (8.9 vs. 15.5, adjusted mean difference, -7.0 ; 95% confidence interval [CI], -10.8 to -3.2), depression (9.4 vs. 12.7, adjusted mean difference, -3.4 ; 95% CI, -6.5 to -0.4), and psychosocial dysfunction (12.5 vs. 16.5, adjusted mean difference, -6.4 ; 95% CI, -10.4 to -2.3). Adjusted mean differences between the 2 groups at 3 months did not show significant differences for teacher-reported classroom problems in acting out (-1.0 ; 95% CI, -2.5 to 0.5), shyness/anxiousness (0.1 ; 95% CI, -1.5 to 1.7), and learning (-1.1 , 95% CI, -2.9 to 0.8). At 6 months, after both groups had received the intervention, the differences between the 2 groups were not significantly different for symptoms of PTSD and depression; showed similar ratings for psychosocial function; and teachers did not report significant differences in classroom behaviors.

Conclusion: A standardized 10-session cognitive-behavioral group intervention can significantly decrease symptoms of PTSD and depression in students who are exposed to violence and can be effectively delivered on school campuses by trained school-based mental health clinicians.

SOURCE: Stein, Jaycox, Kataoka., et al. (2003).

to worry about the validity of the measurements, the adequacy of the research design, or the interpretation of the statistics. (That will come later on!) The idea is to show you how the separate parts of evaluation research are assembled to describe programs and evidence of their effects.

The data analytic methods used in evaluation research almost always produce an estimate of the program's effect by contrasting outcomes among the experimental and control participants. The researcher starts off with the assumption that no difference exists between the groups (the **null hypothesis**), and uses statistical tests to challenge the assumption. When the research data are analyzed, the statistical tests determine the *P* value, the probability of seeing an effect as big as or bigger than that occurring in the study by chance, if the null hypothesis were true. The null hypothesis is rejected in favor of its alternative if the *P* value is less than some predetermined level, traditionally 1% (0.01) or 5% (0.05). This predetermined level is called α (alpha) or the level of statistical significance.

For example, suppose a study compares programs A and B in terms of their ability to improve functional status in older adults. The researcher starts off with the assumption that the two programs are equally effective and sets alpha at 0.05. That means the researcher has set 5% as the maximum chance of incorrectly rejecting the null hypothesis (that there is no difference). This 5% is the level of reasonable doubt that the researcher is willing to accept. If the null hypothesis is falsely rejected (there actually is no difference and the null is correct), this is called a **Type I error**. Failing to reject the null hypothesis (there really is a difference) is called a **Type II error** and referred to as β (beta).

There are many reasons for Type I and Type II errors; among them are unreliable measures of the outcomes, inadequate research designs, and inappropriate data analysis. That is, Program A may actually be different or better than B (or the other way around), but a poorly designed study with inaccurate measures will not be able to detect the difference!

As you can see in the second evaluation in Example 1.4, evaluation researchers sometimes report results in terms of **confidence intervals** (CI) rather than *P* value. A CI is a measure of the uncertainty around the main finding of a statistical analysis. The **odds** are a way of expressing the chance of an event, calculated by dividing the number of individuals in a sample who experienced the event by the number for whom it did not occur. For example, if, in a sample of 100, 20 people did not improve and 80 people improved, the odds of improving are 20/80, which equals $\frac{1}{4}$, 0.25, or 1:4.

An **odds ratio** is the ratio of the odds of an event in one group (e.g., the experimental group or “cases”) to the odds of an event in another group (e.g., the control group or “controls”).

To calculate the odds ratio, you count the number in each group that experiences an event or has a “risk factor.” You then divide the odds of having the risk factor among people of interest—traditionally called “the cases”—by the odds of having the risk factor among the controls. (Risk factors are variables that increase the likelihood of disease or other bad outcomes. For instance, smoking is a risk factor for heart disease. The vocabulary—terms such as *cases*, *controls*, and *risk factors*—comes from studies of public health problems.)

The formula for calculating the odds ratio is as follows:

The Formula for the Odds Ratio: AB/CD

<i>Risk Factor Present?</i>		<i>Control</i>	<i>Total</i>
Yes	A	B	A + B
No	C	D	C + D
Total	A + C	B + D	N

Here’s how researchers might use the odds ratio.

Suppose a researcher is interested in the relationship between not eating breakfast (the risk factor) and trouble adhering to a diet. The researcher asks this question: When compared to people who eat breakfast, what is the likelihood that people who do not will have trouble adhering to a weight-loss diet?

The researcher identifies 400 people who have trouble adhering to a diet and 400 who do not. She finds that among all people with problems, 100 did not eat breakfast and 300 did. Among people without problems, 50 did not eat breakfast. To compare the odds of problems between the two groups, the researcher put the data into a table.

Odds Ratio: Problems Keeping to a Diet

<i>No Breakfast</i>	<i>Case</i>	<i>Control</i>	<i>Total</i>
Yes	100	50	150
No	300	400	700
Total	400	400	800

The odds ratio is calculated as follows:

OR = AD/BC so that

$$100(A) \times 400(D) / 50(B) \times 300(C) = 4000/1500 = 2.67$$

The odds of being exposed to the risk factor (no breakfast) is 2.67 higher for people who have problems adhering to a weight-loss diet than for people who do not have such problems. The answer to the question is that people who do not eat breakfast are 2.67 times more likely to encounter problems adhering to a diet than people who eat breakfast.

Estimates of unknown quantities, such as the **odds ratio** (cited in the first evaluation) comparing an experimental program with a control, are usually presented as a point estimate (such as a -3.4 difference in scores between groups) and a 95% confidence interval (e.g., -6.5 to -0.4 point difference in scores). This means that if someone were to keep repeating a study in other samples from the same population, 95% of the confidence intervals from those studies would contain the true value of the unknown quantity (in this case, the difference in scores).

Alternatives to 95%, such as 90% and 99% confidence intervals, are sometimes used. Wider intervals indicate lower precision; narrower intervals, greater precision. For example, researchers could have a 99% confidence interval that would mean that they are 99% confident that the true value will be between 12 and 1, and they could also have a 99% confidence interval that would mean they are 99% certain that the true value will be between 1.2 and 1. The narrower interval (between 1.2 and 1) indicates greater precision.

Many medical and health journals prefer reports of confidence intervals to *P* values because they provide a plausible range for the true value of the difference between groups.

Research consumers should be concerned with practical or clinical significance as well as with statistical significance when looking for evidence that matters. Practical significance reflects how much of an effect a client, physician, patient, student, or other program participant sees or finds useful. Assessing practical significance takes into account factors such as the size of the program's effect, the severity of the need being addressed, and the cost.

The practical significance of a program or treatment is based on external standards provided by practitioners, clients, customers, patients, or researchers. Unfortunately, little consensus currently exists in almost any field as to the criteria for these standards.

Practitioners and researchers in evidence-based medicine use several different methods to come up with standards of clinical significance that are useful in all types of evaluation research. One method EBM practitioners use involves estimating the number of patients who would have to receive a particular treatment to prevent just one from having a bad outcome over a particular time period. This estimate is called the number needed to treat (NNT). If, for instance, the number of patients needed for a treatment is 10, then the practitioner or clinician would have to give the treatment to 10 patients to prevent 1 patient from having the bad outcome over the defined period, and each patient who received the treatment would have a 1 in 10 chance of being a beneficiary. The best NNT is 1: every person treated benefits. Unfortunately, an NNT of 1 is rarely achieved, and even an NNT of 20 or 40 may be significant (McQuay & Moore, 1997).

Other methods used by evidence-based health practitioners to determine clinical significance include calculating the extent to which an experimental treatment increases the probability of a good outcome and reduces the probability of a bad one.

Because of the pragmatic nature of almost all evaluation research, none is immune to methodological flaws. The evaluator's challenge is to design and implement a study that results in findings and conclusions that are more accurate than they would have been with another sampling strategy, research design, outcome measure, or analytic method. The evaluation's strengths must be demonstrably greater than its limitations. In short, the highest quality evaluations are those that result in the best evidence.

Table 1.1 contains a list of evaluation reports that will give you an introduction to the contents and format of typical evaluation studies. Later on, we will address methods for assessing their quality.

So the Evidence Is Valid, but Does It Matter? _____

Evidence that matters is meaningful to its users as well as scientifically valid. Who are the users of evaluation research? What makes evidence meaningful? Evaluation users are the sponsors and funders of research, the individuals and communities who participate in the research, and the policy makers and others who decide on the adoption of programs and practices. Evaluation users are often known collectively as **stakeholders** or **decision makers**. As a consumer of evaluation research, you are a stakeholder.

Table 1.1 Sample Evaluation Reports and Where to Find Them

- Belsky, J., Melhuish, E., et al. (2006). Effects of Sure Start local programmes on children and families: Early findings from a quasi-experimental, cross sectional study. *British Medical Journal*, 332(7556), 1476–1478.
- Ciaranello, A. L., Molitor, F., Leamon, M., Kuenneth, C., Tancredi, D., Diamant, A. L., et al. (2006). Providing health care services to the formerly homeless: A quasi-experimental evaluation. *Journal of Health Care for the Poor and Underserved*, 17(2), 441–461.
- Diamant, A. L., Brook, R. H., Fink, A., & Gelberg, L. (2001). Assessing use of primary health care services by very low-income adults in a managed care program. *Archives of Internal Medicine*, 161(9), 1222–1227.
- Fink, A., Elliott, M. N., Tsai, M., & Beck, J. C. (2005). An evaluation of an intervention to assist primary care physicians in screening and educating older patients who use alcohol. *Journal of the American Geriatrics Society*, 53(11), 1937–1943.
- Finn, J., Kerman, B., & LeCornec, J. (2005). Effects of a national indicated preventive intervention program. *Journal of Community Psychology*, 33(6), 705–725.
- Fishbein, M., Hall-Jamieson, K., Zimmer, E., von Haeften, I., & Nabi, R. (2002). Avoiding the boomerang: Testing the relative effectiveness of antidrug public service announcements before a national campaign. *American Journal of Public Health*, 92(2), 238–245.
- Fox, P. G., Rossetti, J., Burns, K. R., & Popovich, J. (2005). Southeast Asian refugee children: A school-based mental health intervention. *International Journal of Psychiatric Nursing Research*, 11(1), 1227–1236.
- Hay, J., LaBree, L., Luo, R., Clark, F., Carlson, M., Mandel, D., et al. (2002). Cost-effectiveness of preventive occupational therapy for independent-living older adults. *Journal of the American Geriatric Society*, 50(8), 1381–1388.
- Husler, G., Werlen, E., & Blakeney, R. (2005). Strengthening families with first-born children: Exploratory story of the outcomes of a home visiting intervention. *Research on Social Work Practice*, 15(5), 323–338.
- Parrish, A. R., Oliver, S., Jenkins, D., Ruscio, B., Green, J. B., & Colenda, C. (2005). A short medical school course on responding to bioterrorism and other disasters. *Academic Medicine*, 80(9), 820–823.
- Polaschek, D. L. L., Wilson, N. J., Townsend, M. R., et al. (2005). Cognitive-behavioral rehabilitation for high-risk violent offenders—An outcome evaluation of the violence prevention unit. *Journal of Interpersonal Violence*, 20(12), 1611–1627.
- Runyan, C. W., Gunther-Mohr, C., Orton, S., Umble, K., Martin, S. L., & Coyne-Beasley, T. (2005). Prevent A Program of the National Training Initiative on Injury and Violence Prevention. *American Journal of Preventative Medicine*, 29(5S2), 252–258.
- Rydholm, L., & Kirkhorn, S. R. (2005). A study of the impact and efficacy of health fairs for farmers. *Journal of Agricultural Safety and Health*, 11(4), 441–448.
- Sheeran, P., & Silverman, M. (2003). Evaluation of three interventions to promote workplace health and safety: Evidence for the utility of implementation intentions. *Social Science and Medicine*, 56(10), 2153–2163.
- Stein, B. A., Jacox, L. H., Kataoka, S. H., Wong, M., Tu, W., Elliott, M. N., & Fink, A. (2003). Mental health intervention for school children exposed to violence: A randomized controlled trial. *JAMA*, 290, 603–611.
- Wells, K. B., Sherbourne, C., Schoenbaum, M., Duan, N., Meredith, L., Unutzer, J., et al. (2000). Impact of disseminating quality improvement programs for depression in managed primary care: A randomized controlled trial. *JAMA*, 268(2), 212–220.

Evaluators have traditionally advocated including stakeholders as part of the study team, sometimes inviting them early in the process to define the focus of the research and later on to help make certain the data being collected are valid and useful. Often, special techniques, such as community forums or interviews with key members of the community, are relied upon to make sure that evaluation considers high priority, culturally relevant concerns. More recently, the process of **participatory evaluation** has been extended by some evaluators to include stakeholders as partners in setting the evaluation agenda, doing all phases of the research including analyzing and reporting on the results. Because evaluation almost always takes place within a political and organizational context, it requires group skills, management ability, political dexterity, and sensitivity to multiple stakeholders.

An important development in health evaluation research is the call to health researchers to conduct practical clinical trials (Tunis, Stryer, & Clancy, 2003) in which the hypothesis and study design are developed specifically to answer the questions faced by decision makers. According to proponents of practical clinical trials, there are widespread gaps in evidence-based knowledge, and the existence of these gaps suggests that systematic flaws can be found in the production of scientific evidence, in part because there is no consistent effort to conduct clinical trials designed to meet the needs of decision makers. To remedy this situation, these proponents stress that clinical and health policy decision makers—namely, research users—will need to become more involved in all aspects of clinical research, including priority setting, infrastructure development, and funding.

We will cover some of the most commonly used methods to elicit the views and expectations of stakeholders later on (Chapter 8). Evaluation research reviewers will find these methods useful because they ensure the meaningfulness to their clients of the recommended selection of programs and practices.

Evidence-Based Medicine (EBM): Some History and Definitions

The History

Evidence-based medicine (EBM) has become a crucial and topical issue in modern health throughout much of the world. EBM is strictly defined as the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual

patients. A more expansive definition includes health programs, practices, and policies in the decision-making process.

Testing medical interventions for safety and effectiveness in an experimental manner has probably existed for many hundreds of years. Among the first recorded evaluations is one that dates back to biblical times. Daniel of Judah compared the health effects of a vegetarian diet (the intervention) with those of the Royal Babylonian diet (control group) over a 10-day period. The Book of Daniel (1:15–16) records the findings:

At the end of the ten days their appearance was better and their bodies healthier than all the young men who had been eating the royal delicacies. So the warden removed their delicacies and the wine from their diet and gave them a diet of vegetables instead.

Leonardo da Vinci wrote in his *Notebooks* (1508–1518) that

Those who are enamored of practice without science are like a pilot who goes into a ship without rudder or compass and never has any certainty where he is going. Practice should always be based upon a sound knowledge of theory. (1.20)

According to Sackett, Straus, Richardson, Rosenberg, and Haynes (2000), five of the originators of EBM and the authors of an extremely influential textbook, the roots of EBM lie in Chinese medicine. In the reign of the Emperor Qianlong (1711–1799), a method known as “kaozheng” (practicing evidential research) was apparently practiced in relation to Confucian medical texts. Sackett, Straus, Richardson, Rosenberg, and Haynes also identify the ideas of EBM with postrevolutionary Paris clinicians, at least one of whom rejected the pronouncements of authorities that vivisection was good for cholera.

It was only in the twentieth century, however, that EBM really evolved to affect almost all fields of health care and policy. Professor Archie Cochrane, a Scottish epidemiologist, through his book *Effectiveness and Efficiency: Random Reflections on Health Services* (1972) and subsequent advocacy, was responsible for the increasing acceptance of the concepts behind evidence-based practice. The explicit methodologies used to determine “best evidence,” however, were largely established by the McMaster University research group led by David Sackett and Gordon Guyatt. The term “evidence-based medicine” first appeared in the medical literature in 1992 in a paper by Guyatt, Cairns, Churchill, et al.

The Definitions: Research Evidence, Clinical Expertise, and Patient Values

Sackett, Straus, Richardson, Rosenberg, and Haynes (2000) define EBM as the integration of best research evidence with clinical expertise and patient values.

The *best research evidence* is clinically relevant research, often from the basic sciences of medicine but especially from patient-centered clinical research. Patient-centered clinical research is analogous to evaluation research particularly in its advocacy and use of experimental methods to test effectiveness, impact, and cost. EBM is concerned not only with evidence that matters about interventions and programs but also addresses needs for information pertaining to diagnosis, prognosis, and prevention.

New evidence from clinical research both invalidates previously accepted tests and treatments and replaces them with new ones that are more powerful, more accurate, more efficacious, and safer. For example, stomach ulcers were once thought to be the result of stress or eating spicy foods. Generations of ulcer sufferers drank gallons of milk, avoided certain foods, and tried to stay calm. In 2005, two Australian physicians won a Nobel Prize for their work showing that most stomach ulcers and gastritis were caused by colonization with a bacterium called *helicobacter pylori* and not by stress or spicy food, as had been assumed before. Now, stomach ulcer patients are often treated with antibiotics (Marshall & Warren, 1984).

Giving up firmly held beliefs in the face of new evidence is challenging to say the least. As another example, consider one of the most unshakeable tenets of many health care practitioners, namely, that if patients are educated and made to be active partners in the treatment of their disease, their ability to take care of their disease should improve dramatically. Even this belief has been challenged by the evidence.

To evaluate the effectiveness of patient education in diabetic patients, a team of researchers (Sanchez et al., 2005) enrolled 200 patients who were treated at Duke University Hospital for acute coronary syndrome (ACS), a condition characterized by blockages in coronary arteries that prevent oxygen-rich blood from nourishing the heart, which can lead to chest pain and possibly heart attack. At enrollment, each patient took a standardized test that measured his or her knowledge related to diabetes. Patients were then ranked as either high-scoring or low-scoring.

The researchers found that, for diabetics, improved disease knowledge alone did not translate into improved blood sugar control, cholesterol levels, weight management, or mortality rates. They concluded that, while education may be important, other health care delivery

variables must be addressed to reduce the risks of diabetic patients dying of heart disease, the main cause of death for these patients.

Clinical expertise means the ability to use clinical skills and past experience to rapidly identify each patient's unique health state and diagnosis, his or her individual risks and benefits of potential interventions, and his or her personal values and expectations.

Patient values are the unique preferences, concerns, and expectations that each patient brings to a clinical encounter and that must be integrated into clinical decisions if they are to serve the patient. EBM supporters and consumers believe that, when these three elements are integrated, clinicians and patients form a diagnostic and therapeutic alliance that optimizes clinical outcomes and quality of life.

Figure 1.2 is a graphic representation of the evidence-based medicine paradigm.

The figure suggests that clinical decisions must include consideration of the patient's clinical and physical circumstances to establish what is wrong with his or her health and what treatment options are available. Next, the options must be tempered by research evidence concerning their effectiveness and efficiency. Third, given the likely consequences associated with each option, the clinician must consider

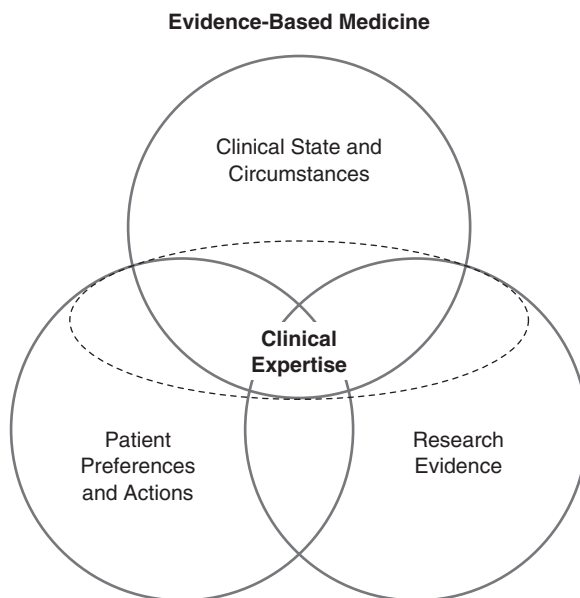


Figure 1.2 The Evidence-Based Medicine Paradigm

SOURCE: Sackett, D. L., Straus, S. E., Richardson, W. S., Rosenberg, W., & Haynes, R. B. (2000). *Evidence-based medicine: How to practice and teach EBM* (3rd ed.). New York: Churchill and Livingstone. Reprinted by permission of D. L. Sackett.

the patient's preferences and likely actions (in terms of which interventions she or he is ready and able to accept). Finally, clinical expertise is needed to bring these considerations together and recommend treatment that is agreeable to the patient. Put another way, EBM merges the science and art of medicine.

EBM has had a major impact on medicine, nursing, and other health professions in the United States and throughout the world. The application of the principles of EBM to all professions associated with health care, including purchasing and management, is referred to as evidence-based health care (Sackett et al., 2000). The principles have influenced thinking in nearly all the helping professions because of demands from the community and from program sponsors for evidence that matters. Outside of a strictly medical context, say in social work, criminology, criminal justice, education, or psychology, EBM is considered to be the parent discipline of evidence-based practice or EBP.

EBM practitioners have focused much of their intellectual effort on developing methods for grading the **quality** and rating the **strength** of a body of evidence (Lohr, 2004). The quality of each study's evidence depends on factors such as the characteristics of its research design, the adequacy of the sample size, the composition of the participants, and the validity of the outcomes. The strength of the evidence can only be determined if multiple studies are available. A strong body of evidence should meet three criteria: quality, quantity, and consistency. The quality of evidence is often a summation of the direct grading of the quality of individual studies. The quantity of evidence reflects the magnitude or impact of the effects (benefits and harms). The consistency of results reflects the extent to which studies report findings that reflect effects of similar magnitude and direction.

Because of the complexity of applying evidence-based research findings to clinical care, medical groups have created journals and online resources that provide practice guidelines, reviews of research, and bibliographies to help them practice EBM. Centers for EBM in a range of specialties exist throughout the world. The Centre for Evidence-Based Medicine (www.cebm.utoronto.ca) continually updates EBM resources, and the Agency for Healthcare Research and Quality (www.ahrq.gov) provides many useful tools. Despite these aids, however, many EBM practitioners find that they are like other consumers in needing to learn how to do their own research evaluation. Many medical schools anticipate this need and offer evidence-based medicine courses to teach students to identify individual and synthesized research studies and evaluate their quality. In fact, as the call for evidence increases, other fields, including nursing, education, and psychology, are introducing similar courses.

Evaluation researchers and EBM practitioners share similar objectives. They both count on the experimental method to provide evidence that matters and agree on the need to incorporate values and expectations into treatments, practices, and programs. Despite their many similarities, some EBM practitioners differ from evaluation researchers in at least one main use of their research results. In EBM, each physician's clinical expertise is used as the basis of judgments for applying research findings to the care of *individual* clients. That is, EBM physicians are presented with research findings on groups of people (such as diabetic patients or substance abusers), and they must translate the research into evidence-based care for individual patients. EBM physicians ask questions like these: Do the findings of this research apply to this individual? Is my patient so different from the study participants that the findings do not apply?

Evaluation consumers and other stakeholders who work outside of the direct clinical encounter concentrate on analyzing evidence to make decisions about the applicability of programs for *groups* of clients. They are not primarily concerned with translating research findings into options for individual care. Their job is to translate research that has included one group of people (such children exposed to violence or substance abusers) into recommendations regarding evidence-based programs and practices for another. Their questions are more likely to be like these: Do the findings of this research apply, *on average*, to the institutions and communities (such as schools, prisons, counties) in my setting? Are the people in my setting so different from participants in the study that the findings do not apply?

The techniques in this book are for stakeholders and other consumers of research who are mainly concerned with identifying programs to meet the needs of institutions, communities, and society. The book also advocates the acceptance of EBM's principle of incorporating client values into decisions about the choice of programs, and these values are represented by the common good.

Getting It Done: The Steps to EBM

EBM practice is comprised of five steps (Sackett et al., 2000) that are common to all evaluation consumption:

1. **Convert the need for information into answerable questions.** This step is often considered to be the hardest for EBM practitioners to accomplish. Any clinical situation can raise a very large number of questions, and selecting the most pertinent can be challenging. It is

also difficult to state questions simply because the concepts involved in clinical settings are often very complex. Well-formulated questions are essential, however, in devising strategies that are likely to direct practitioners to the data they need to get answers.

Evaluation researchers and consumers also need to learn how to convert their information needs into questions. Researchers must discover how to ask answerable questions about outcomes, impact, and costs. Consumers or evaluators of research must ask questions about the quality of the research and its relevance to clients.

2. Track down the best evidence to answer the questions. Evidence databases are emerging in many fields, and, when available, they can be a primary source of information for the program seeker. But they are not always available, comprehensive, and current. To be complete, evidence databases need to be regularly updated—an enormous and costly task. Because databases may be two or three years behind the published literature, it is not uncommon for the evaluation consumer to have to rely upon individual studies or syntheses of many studies that are published in journals or made public online. Thus, to be certain that you have all available information, you, as a research consumer, need to acquire skills for searching the literature and also for applying systems to grade the quality and rate the strength of evidence.

3. Critically appraise the evidence for validity and impact. Once one or more potentially suitable programs are identified through evidence databases or the literature or both, here are the next questions: How valid is the evidence for the effectiveness of these programs? Does the evidence support a large and important impact? An accurate evaluation of validity requires the consumer to become knowledgeable about the basic foundations of research in order to be able to grade a study's quality.

4. Integrate the critical appraisal with experience and understanding of values. A really great study may produce evidence that matters about a program that has had an impact on very small numbers of people. Program seekers must find ways to determine from stakeholders how large the impact must be. What proportion of the population should receive benefits? What is the acceptable amount of time for benefits to be manifest? To endure?

5. Evaluate one's own effectiveness and efficiency in executing steps 1–4 and seek ways to improve both next time. Evaluation may be as specific as analyzing performance with respect to asking questions or searching for evidence, or it may pertain to the extent to which the EBM process made a difference in the processes or outcomes of care.

EBM and EBP: Perfect or Not So?

Almost nothing in health or social science is perfect, so, in a way, this is an unfair question because you know the answer already: EBM is not perfect. Among the limitations of practicing EBM is the relative shortage of consistently high quality research that meets the needs of many consumers. With time, faith in the approach, and financial support, this situation may be corrected. Perhaps a more important limitation than the shortage of evidence is that evidence that matters is not available regarding the outcomes of practicing EBM and EBP. Sackett et al. (2000) point out that no investigative team or research granting agency has yet overcome the problems of sample size and follow-up that such an evaluation of EBM requires. Also, ethical concerns exist because to conduct a scientific evaluation means having a control group that would be denied access to evidence-based treatment (so that the outcomes could be compared to the experimental group that would be given the evidence-based treatment). Sackett et al. (2000) point out, however, that many other studies have found that patients who received evidence-based interventions have better outcomes than those who do not, so that is a good start.

Getting It Together: Evaluation Research, EBM/EBP, and the Research Consumer

The research consumer is someone who works in the helping professions such as social work, education, psychology, nursing, public health, or occupational health and who has been given the assignment to find **evidence-based** (or **research-based**) programs and practices that are likely to work with a given population in a particular setting. Evaluation research and EBM are comprised of principles and methods that, when used in combination, improve the likelihood that the program that research consumers select will be pertinent and trustworthy.

Evaluation research has been around for decades and has experience designing and assessing programs to improve the public's education and well-being. EBM has developed systems for identifying programs and evaluating their quality and rating the strength of their evidence. Both evaluation research and EBM emphasize the importance of placing the evidence in the context of client or patient values and preferences.

Table 1.2 shows the relationship among the evaluation researcher, EBM practitioner, and research consumer.

Table 1.2 Evaluation Research, EBM, and the Research Consumer: Compare and Contrast

	<i>Evaluation Research</i>	<i>Evidence-Based Medicine and Practice</i>	<i>Evaluating Research: The Consumer</i>
Objective	Produce valid evidence about the process, outcomes, impact, and costs of programs and interventions.	Identify evidence that matters (valid, meaningful, and with consistent impact) to provide the best possible clinical care.	Identify high quality evidence-based programs that meet client needs, values, and expectations.
Methods	Use scientific method to design studies, collect information, and analyze and interpret data. Add a participatory dimension to ensure that evidence obtained is evidence that matters (meets needs, values, and expectations of stakeholders).	Use prespecified and transparent systems for grading the quality and rating the strength of the evidence. Use the best evidence available to select options for treatment and incorporating patient values and expectations.	Identify clients' needs, values, and expectations. Identify high quality programs. Match clients' values and high quality programs.
Ethical Concerns	Respect participants' rights to privacy and to have an understanding of the risks and benefits of participation.	Respect participants' rights to privacy and to have an understanding of the risks and benefits of participation.	Make certain that ethical concerns have been addressed by the researchers. Understand the limits of evaluation research and systems for rating its quality when selecting programs to meet clients' needs.

Figure 1.3 is a graphic representation of the evaluation consumer's role. As you can see from the figure, research consumers and the public they serve constitute an alliance that is analogous to the EBM practitioners' with patients. The figure also suggests that, in selecting programs, research consumers should identify institutional, community, and societal needs first. Next, the choice of programs must be tempered by the best research evidence concerning their effectiveness.



Figure 1.3 Practicing Research: Discovering Evidence That Matters

The basis for analyzing the research evidence rests on principles derived from evaluation research and EBM. Third, the research consumer must consider the client's values and expectations in weighing the evidence. Finally, expertise in evaluating research is needed to bring these considerations together.

Summary of Chapter 1: Evaluation Research and Evidence-Based Practice Partnership

Words to Remember

abstracts, clinical trials, confidence interval, control group, controlled experiment, effectiveness, efficacy, ethics, evaluation research, evidence that matters, experimental group, experimental studies, formative evaluation, hypotheses, impact, implementation evaluation, null hypothesis, number needed to treat, observational research, odds ratio, outcomes, *P* value, participatory evaluation, practical clinical trials, practical significance, process, program, qualitative, quantitative, research, research consumer, research ethics, research-based, research questions, scientific method, study quality, study strength, summative evaluations, Type I error, Type II error, variables

Evaluation research is a systematic method of assessing the **process, outcomes, impact, and costs** of a **program or intervention**. Scientific evaluations produce the **best research evidence** about programs and new knowledge about social behavior. For research evidence to **matter**, it must be accurate and consistent with its users' values and expectations. **Evidence-based medicine (EBM)** and its siblings, **evidence-based health care** and **evidence-based practice (EBP)**, are exemplified by the integration of best research evidence, clinical expertise, and patient or client values in making clinical, programmatic, management, and policy decisions. Evaluation research and EBM overlap in their insistence on best evidence tempered by values and professional experience.

Evaluation researchers have expertise in analyzing the effectiveness and safety of programs to improve the public's health, education, and well-being and in working in organizational and political settings. EBM practitioners have developed explicit systems for locating and analyzing research findings and for grading their quality and strength. The **research consumer's** ability to obtain pertinent and reliable information that matters is strengthened by understanding and applying the methods and experiences that characterize both disciplines.

The Next Chapters

The next chapters build on evaluation research and EBM principles to focus on methods of identifying effective and useable programs, grading the quality and rating the strength of evidence that supports effectiveness, summarizing the results, and finding out about the values and expectations of research consumers and other users.

Chapter 2 discusses where to look for programs and how to ask the right questions.

Exercises

1. Explain whether each of these is an evaluation study or not.
 - a. **Research Objective:** The purpose of the study was to evaluate a randomized culturally tailored intervention to prevent high-HIV-risk sexual behaviors for Latina women residing in urban areas.
 - b. **Research Objective:** To determine the efficacy of a spit tobacco (ST) intervention designed to promote ST cessation and discourage ST initiation among male high school baseball athletes.
 - c. **Research Objective:** To study drivers' exposure to distractions, unobtrusive video camera units were installed in the vehicles of 70 volunteer drivers over 1-week time periods.
2. Read the abstract of an evaluation of a home visiting program to prevent child abuse and neglect. Then answer the questions below.

Abstract: Preventing Abuse and Neglect of Children

Objectives: To assess the impact of home visiting in preventing child abuse and neglect in the first 3 years of life in families identified as at risk of child abuse through population-based screening at the child's birth.

Methods: This experimental study focused on Hawaii Healthy Start Program (HSP) sites operated by three community-based agencies. From 11/1994 to 12/1995, 643 families were enrolled and randomly assigned to intervention and control groups. Child abuse and neglect were measured by observed and self-reported parenting behaviors, all hospitalizations for trauma and for conditions where hospitalization might have been avoided with adequate preventive care, maternal relinquishment of her role as primary caregiver, and substantiated CPS reports. Data were collected through annual maternal interviews (88% follow-up each year of all families with baseline interviews); observation of the home environment; and review of CPS, HSP, and pediatric medical records.

Results: HSP records rarely noted home visitor concern about possible abuse. The HSP and control groups were similar on most measures of

(Continued)

(Continued)

maltreatment. HSP group mothers were less likely to use common corporal or verbal punishment (AOR=.59, $p=.01$), but this was attributable to one agency's reduction in threatening to spank the child. HSP group mothers reported less neglectful behavior (AOR=.72, .02), which related to a trend toward decreased maternal preoccupation with problems and to improved access to medical care for intervention families at one agency.

Conclusions: The program did not prevent child abuse or promote use of nonviolent discipline; it had a modest impact in preventing neglect. Possible targets for improved effectiveness include the program's implementation system and model.

SOURCE: Duggan, McFarlane, Fuddy, et al. (2004).

Now that you have read the abstract, do the following:

- a. Describe the intervention
 - b. Name the main outcomes that were studied
 - c. Formulate at least two hypotheses based on a reading of the study's objectives
3. Define at least 6 characteristics of evaluation research.
 4. Define the main characteristics of EBM/EBP.
 5. In what ways are evaluation research and EBM similar? In what ways are they different?
 6. How does evaluation research differ from other social research?
 7. How is evaluation research like other social research?
 8. Explain how the research consumer combines methods from research and EBM/EBP.
 9. Compare these four definitions of evaluation.
 - Evaluation research is a systematic method of assessing the process, outcomes, impact, and costs of a program or intervention. Scientific evaluations produce the best research evidence about programs and new knowledge about social behavior. For research evidence to matter, it must be accurate and helpful to the evaluation's users.
 - The key to a successful program or project is evaluation. Evaluation provides formative feedback that helps guide a

program as it is being implemented. It also provides summative data that clearly demonstrate that the program is accomplishing its stated goals and objectives. Without effective evaluation, the program staff may fail to document important impacts the program has on its participants. It may also fail to recognize how different components in the program are affecting the participants or participating institutions. In an era of limited resources for educational programs, those programs that can document their success in having an impact on their participants and in using resources efficiently will be at an advantage for ongoing funding. (American Physiological Association, 2002)

- The purpose of evaluation is to produce information about the performance of a program in achieving its objectives. In general, most evaluations are conducted to answer two fundamental questions: Is the program working as intended, and why is this the case? Research methods are applied to answer these questions and to increase the accuracy and objectivity of judgments about the program's success in reaching its objectives. (Grembowski, 2001)
- The generic goal of most evaluations is to provide “useful feedback” to a variety of audiences including sponsors, donors, client-groups, administrators, staff, and other relevant constituencies. Most often, feedback is perceived as “useful” if it aids in decision-making. But the relationship between an evaluation and its impact is not a simple one—studies that seem critical sometimes fail to influence short-term decisions, and studies that initially seem to have no influence can have a delayed impact when more congenial conditions arise. Despite this, there is broad consensus that the major goal of evaluation should be to influence decision-making or policy formulation through the provision of empirically driven feedback. (Trochim, 2006)

Further Reading

Web Sites

Project ALERT is administered by the BEST Foundation for a Drug-Free Tomorrow—<http://www.projectalert.best.org>.

For more information or for training, contact

BEST Foundation

725 S. Figueroa St., Suite 1615

Los Angeles, CA 90017

800-ALERT-10

Online Evaluation Resource Library—<http://oerl.sri.com> (accessed August 2007)

The Online Evaluation Resource Library (OERL) is a library that has a collection of plans, instruments, and reports that have been used to conduct evaluations of projects funded by the Directorate for Education and Human Resources (EHR) of the National Science Foundation (NSF). OERL also contains glossaries of evaluation terminology, criteria for best practices, and scenarios illustrating how evaluation resources can be used or adapted.