8

The Ethical Research Consumer Assesses Needs and Evaluates Improvement

This chapter discusses the context in which the process of discovering evidence that matters takes place and focuses on three topics: identifying clients' needs, determining if programs have met those needs, and applying ethical standards when practicing research.

Discovering evidence that matters depends upon integrating the best research with clients' values and needs to make decisions about the effectiveness and appropriateness of programs and practices. This chapter discusses techniques that can be used to **assess the needs** and priorities of institutions, communities, and society. The techniques include using key informant methods, public or community forums, focus groups, the nominal group process, the Delphi technique, the RAND/UCLA Appropriateness Method, surveys, and consensus development conferences.

An important step in evidence-based practice is to evaluate one's own effectiveness and efficiency and seek ways to improve both. The chapter therefore discusses improvement evaluation, which is designed to study and enhance the effectiveness of already established evidence-based programs and practices. Did the evidence-based process meet the needs and conform to the values of clients? Is improvement needed in the process?

This chapter also focuses on the **ethical concerns** associated with assessing needs, conducting research with human participants, and practicing research.

After reading this chapter, you will be able to

- Distinguish among social, behavioral, administrative, environmental, communal, physical, and educational needs that affect the choice of evidence-based programs and standards for selecting evidence that matters
- Identify the features of techniques for assessing needs and priorities, such as key informant techniques, public or community forums, focus groups, the nominal group process, the Delphi technique, the RAND/UCLA Appropriateness Method, surveys, and consensus development conferences
- Compare the characteristics of improvement and effectiveness evaluations or evaluation research
- Identify the characteristics of research ethics and research misconduct in doing all types of evaluation research
- Identify the limitations of evidence-based practice and their ethical implications
- Distinguish among the ethical concerns that may occur when conducting research
- Distinguish among the ethical concerns that may occur when deciding if evidence matters

Figure 8.1 shows your location on the way to discovering evidence that matters.

Identifying Needs, Preferences, and Values

Evidence-based practitioners count on the experimental method to provide evidence that matters and agree on the need to incorporate users' needs, preferences, and values and expectations into treatments, practices, and programs. A systematic effort to identify user needs and provide a context for them is called a **needs assessment**.

Needs occur when gaps in services and programs lead to unsolved problems affecting health, education, and social well-being. Needs are intertwined with values and preferences, which together affect priorities. For example, a community may need better social services for its elderly but may prefer to spend most of its resources on preventing violence in teens.

Needs can be arranged into six categories: social, communal or epidemiological, behavioral, environmental, educational, and administrative.

Social needs usually refer to the community's perceptions of its problems. For example, one community may see gang warfare or teen

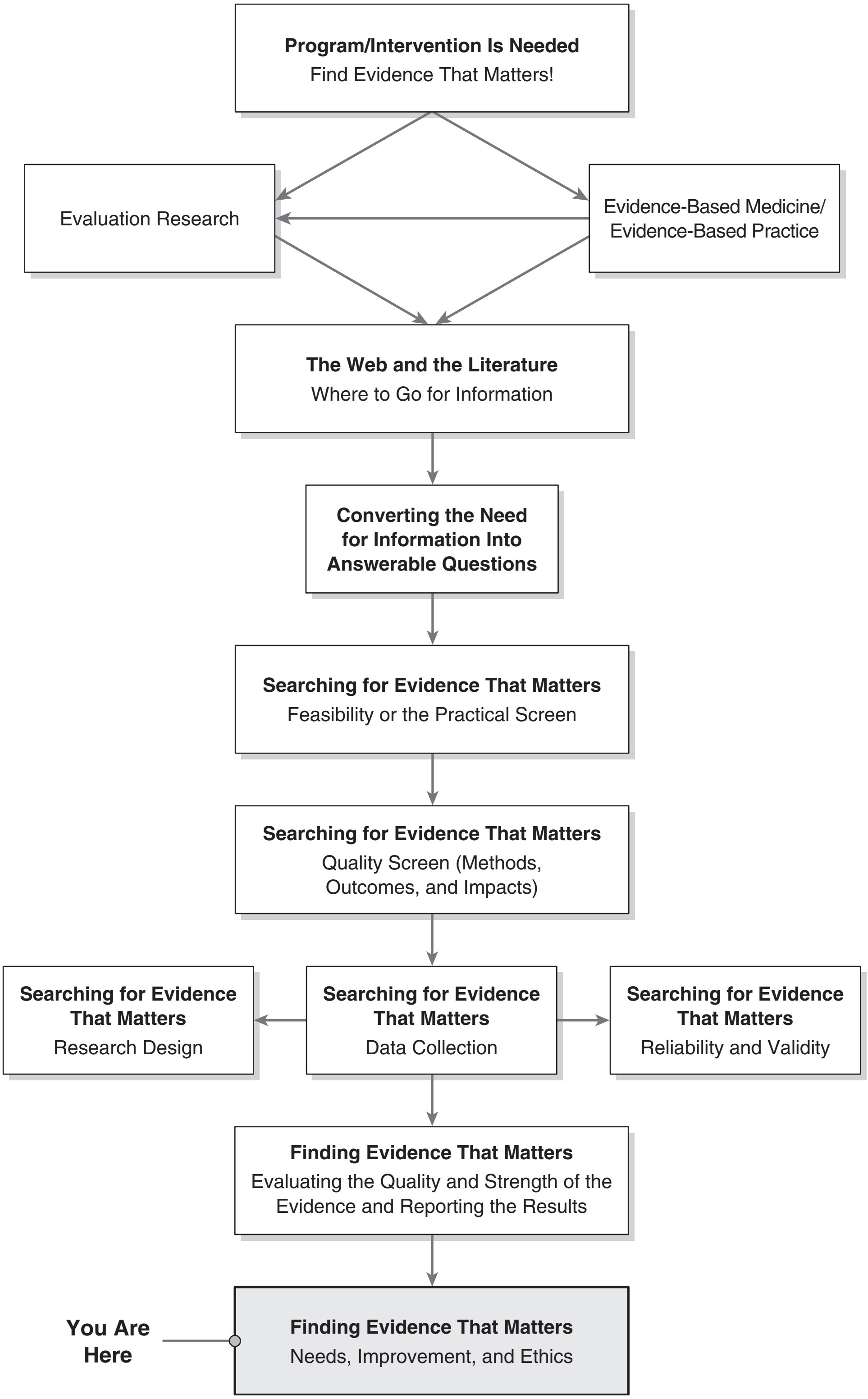
Program/Intervention Is Needed
Find Evidence That Matters!
Evaluation Research
Evidence-Based Medicine/ Evidence-Based Practice
The Web and the Literature
Where to Go for Information
Converting the Need for Information Into Answerable Questions
Searching for Evidence That Matters
Feasibility or the Practical Screen
Searching for Evidence That Matters
Quality Screen (Methods, Outcomes, and Impacts)
Searching for Evidence That Matters
Research Design
Searching for Evidence That Matters
Data Collection
Searching for Evidence That Matters
Reliability and Validity
Finding Evidence That Matters
Evaluating the Quality and Strength of the Evidence and Reporting the Results
You Are Here
Finding Evidence That Matters
Needs, Improvement, and Ethics

Figure 8.1 Location on the Way to Discovering Evidence That Matters

violence as its major problem, while another may regard the unemployment of its youth as the most pressing need.

Communal or epidemiological needs refer to problems that can be documented to affect a large number of people in the community. For example, school records may reveal inadequacies in meeting the needs of special education students in a school district, and a review of the state's statistics on low-weight births may reveal higher than state averages for three counties. Data on communal or epidemiological needs usually come from school or medical records, administrative documents, and other databases like vital statistics and local and national surveys of the public's health, educational status, and welfare.

Behavioral needs refer to individual and communal lifestyles and beliefs that affect a community's well-being. Abundant evidence exists, for example, that some communities rely on diets that are high in fat and that this contributes to high rates of obesity and to concomitant illnesses in those communities. As another example, prenatal care may be viewed by some in Community A as a necessity, while in Community B it may be seen as an attempt to make a medical problem out of a natural process.

Physical needs refer to social or physical factors that are external to an individual or a community. If access to nutritious food is limited in a community, for instance, then it will be difficult to implement a program to instill good eating practices.

Educational needs refer to individual and community knowledge, attitudes, skills, and self-efficacy beliefs. Some communities are interested in the social and political process and have knowledge about how the "system" works. Others are less interested in or knowledgeable about these things.

Administrative needs refer to policies and resources that exist in the organizations and institutions (e.g., schools, hospitals, businesses, nongovernmental organizations) that might facilitate or hinder the adoption of a new program. Evaluation of these needs helps to answer questions like these: What are the barriers to implementation (e.g., lack of staff commitment, lack of space)? What policies should be changed to remove the barriers?

Table 8.1 summarizes each of the six needs that can be explored through a needs assessment.

Techniques for Assessing Needs

At least seven methods are commonly used in assessing individual and public needs. Each has its advantages and limitations.

Table 8.1 Exploring Six Needs

Need	*Explanation*	*Question/Comment*
Social	People's perception of their own needs	What are the community's needs and preferences? Does the community have the resources to solve problems? How readily can the community implement programs?
Communal/ Epidemiological	Determination by researchers or practitioners of which problems are important for specific groups in a community	Information comes from analyses of school records, national databases, and administrative databases
Behavioral	Determination of individual and community lifestyles or behaviors that contribute to existing needs	For instance, these may include dietary preferences that lead to obesity and diabetes; customs regarding receipt of prenatal care
Environmental	External social and physical factors	For instance, how healthy are the foods served to children in school cafeterias? How accessible is prenatal care? Fresh fruits and vegetables?
Educational	Individual and community knowledge, attitudes, skills, and self-efficacy beliefs	A major question is how these factors interact to assure the implementation of new programs and practices
Administrative	Refers to policies and resources prevailing in the organizational context that might facilitate or hinder program implementation	What are the barriers to implementation (e.g., staff commitment, lack of space)?

Key Informant

The purpose of the **key informant method** is to collect information about a community's needs by interviewing community leaders who are likely to be in a position to know what the needs are. Key informants in most communities include religious leaders, physicians, teachers, selected members of public service organizations, the mayor, public safety administrators, business owners, and so on.

The key informant technique is relatively inexpensive to implement because the interviews are almost always conducted locally, and no travel costs are incurred. Because only a relatively few people are interviewed (say, 25), analyzing the results is relatively simple. Further, the technique is useful for getting a variety of perspectives on any single set of problems and an "insider's" view of events.

The main disadvantage of the key informant technique is that you cannot be sure that you have interviewed all relevant people, and, if you did not, the results may not be representative or correct. Also, the leaders who are willing and able to participate may not always be the "voice" of the people. If you want input from the community itself, a technique like the public forum may be a better choice.

Public or Community Forum

A community forum consists of a group of people who meet together to discuss a common problem. The meeting is open to all members of the community.

The purpose of convening one or more public forums is to obtain information from a relatively large number of people in the community at one time. The community may share a geographic area (e.g., a school district) or a special interest (e.g., parents whose children are threatened by school gangs). Usually, a single agency—a community college, a homeowners' association, the local Parent Teacher Association—sponsors the forum and drafts the initial questions for participants. The questions focus on the needs that should have highest priorities in the community, what is being done currently to meet the needs, what should be done in the future, and where resources might come from.

Public or community forums are useful in enabling a large number of interested participants with diverse perspectives to have their say. If done well, a forum can provide a relatively quick look at the community's present needs. Good leadership and advance preparation are crucial if only to make sure that good community debate does not turn into pointless argument.

Focus Groups

A focus group is designed to collect information from "insiders" or "people in the know." The group usually consists of about 10 carefully selected participants and a trained moderator. The session lasts about two hours and is centered on getting answers to four or five carefully constructed questions. Focus group participants almost always receive financial rewards for participation.

The advantage of a well-conducted focus group is that it can produce answers to difficult questions in a short period of time. **Focus groups** are, however, among the most costly needs assessment methods. The questions asked of the group must be skillfully assembled, and an experienced moderator is needed to conduct the discussion. Often, many focus groups are deemed necessary to get a full complement of community views, so, if you add in the incentives, the costs can become pretty steep.

Nominal Group Process

In the nominal group technique, participants are brought together for a discussion session led by a moderator. After the topic of concern has been presented to session participants and they have had an opportunity to ask questions or briefly discuss the scope of the topic, they are asked to take a few minutes to think about and write down their responses. The session moderator will then ask each participant to read, and elaborate on, one of his or her responses. These are noted on a flipchart. Once everyone has given a response, participants will be asked for a second or third response, until all of their answers have been noted on flipchart sheets posted around the room.

Once duplications are eliminated, each response is assigned a letter or number. Session participants are then asked to choose up to 10 responses that they feel are the most important and rank them according to their relative importance. These rankings are collected from all participants and aggregated. Here is an example:

Response	*Participant 1*	*Participant 2*	*Participant 3*	*Columns Inserted Here for Participants 4–12*	*Relative Importance of Each Response*
A	ranked 1st	ranked 2nd	ranked 2nd		5 participants ranked A 1st
B	ranked 3rd	ranked 1st	ranked 3rd		7 participants ranked B 3rd
C	ranked 2nd	ranked 3rd	ranked 1st		6 participants ranked C 2nd
D	ranked 4th	ranked 4th	ranked 4th		12 participants ranked D 4th

Sometimes, the results are given back to the participants in order to stimulate further discussion, and perhaps a readjustment in the overall rankings will be assigned to the various responses. This is done only when group consensus regarding priorities is important.

The **nominal group process** can be used in a wide variety of settings. For example, a nominal group process was used to collate information for the development of a mental health program for victims of drought in rural Australia (Sartore). Twenty-three participants were recruited in consultation with rural mental health organizations. They were asked questions about the best mental health service strategies to minimize and respond to the mental health impact of drought. Three general strategies emerged: community-building and education about the physical, financial, and mental health effects of drought; cooperation between and coordination among agencies in delivering mental health and other drought support; and continuity and planning of improved mental health services.

Delphi Technique

The **Delphi technique** is a structured method of determining the degree of agreement on a topic, selecting alternatives, or setting priorities. Delphi techniques use questionnaires that are completed by participants on their own, in groups, or both. The questionnaires are structured to ask people to rate or rank the importance or validity of certain ideas. For example, Delphi participants might be asked to rate the importance of a particular program objective (1 = *definitely important* to 5 = *definitely not important*) as well as the likelihood that it might be achieved in a particular institution (1 = *definitely likely* to 5 = *definitely not likely*). The results of the ratings (round 1) are sent back to the respondents who are asked to review them and re-rate the items (round 2).

In "mailed" Delphi's (regular mail, e-mail, or on the Web), the participants are usually not known to one another. "Anonymity" is thought to encourage people to focus on the issues rather than on each other. In a Delphi variation in which round 1 is mailed but round 2 is a face-to-face meeting, the participants are known to each other, but their individual ratings are not. The idea behind a face-to-face discussion of the first round's results is that the dialogue increases attention to the subtleties of the issues and introduces new views into the rating process.

The following example (Example 8.1) describes an actual use of the Delphi method to identify the essential characteristics of cognitive behavioral treatment manuals.

Example 8.1 Using Delphi by E-Mail: What Do Experts Agree Should Be in a Cognitive Behavioral Treatment Manual?

Who Were the Delphi Participants?

All participants were experts who were defined as individuals who have published treatment manuals or have used them in published research. Potential participants' names were gathered primarily through an electronic search of the online literature.

How Was Anonymity Guaranteed?

Twenty-nine prospective participants were e-mailed a pre-notification letter inviting them to participate in the study. A positive response was viewed as informed consent. The e-mail process enabled the mass mailing of all correspondence without individuals' knowledge of each other.

What Was the Study Plan?

Round 1. An e-mail, with attachment, was sent to the consenting sample. In the attachment, participants were asked to list their preferences for the contents of a good cognitive behavioral therapy treatment manual.

Round 2. Items generated from round 1 were thematically analyzed by the investigators and a colleague experienced in using treatment manuals. When initial disagreement regarding the categorization of items occurred, discussion took place until the investigator and his colleague reached agreement. Following analysis, a questionnaire was designed for the subsequent rounds. In order to facilitate completion, a 3-point rating scale was generated for each item:

E = Essential. Each manual must contain this item

D = Desirable. Inclusion of this item enhances the manual

I = Inappropriate. Not applicable to the manual

Round 3. The results of round 2 were collated, and the percentage agreement for each category was placed next to each item. The returned questionnaire included the participants' original responses, and participants were given the opportunity to amend their selections (if desired) in response to viewing the overall feedback.

How Was Consensus Defined?

As there is no agreement concerning the required degree of consensus in a Delphi study, the investigators set consensus levels at two-thirds of the responses.

What Did the Delphi Find to Be Essential?

Only 11 (13%) of the generated items were rated as essential:

(Continued)

Example 8.1 (Continued)

General characteristics of treatment manuals

Appropriate for the problem addressed
Coherent and focused
Based on a clear theoretical model

General information that should be found in treatment manuals

A clear specification of what the intervention aims to do
A statement of the aims and objectives of each session
A detailed description of the problem for which the manual has been designed

Intervention strategies/chapters

Rationale of therapy should be linked to intervention
Treatment procedures should be detailed
Treatment procedures should be illustrated with realistic clinical case examples

Specific content of patient led treatment manuals

User friendly
Give hope that therapy will work

A unique feature of the Delphi technique is the anonymity of participants or responses. If participants are not known to one another, the method can be used to obtain agreement among groups and individuals that are normally hostile to one other. But Delphi participants may not be representative of the very group whose needs are being assessed, and this is a limitation. Participation requires the completion of written questionnaires through at least two rounds. Not everyone is survey savvy and can spend the required time. If the results are seen as nonrepresentative, then they might not be taken seriously. Further, no established definition of agreement or consensus exists, and this alone can make Delphi findings appear arbitrary for those who would use a different definition. Finally, the method tends to encourage a middle-of-the-road view, especially if average ratings or majority ratings are used.

The RAND/UCLA Appropriateness Method

The **RAND/UCLA Appropriateness Method (RUAM)** is a method for determining the extent of agreement on controversial topics and those for which the research base is poor or ambiguous (a mixture of positive and negative findings).

The RUAM was originally created to determine the appropriateness of certain medical procedures and surgical operations, such as gallbladder removals or coronary artery bypass surgery. Because the method has proven to be flexible, reliable, and valid (Shekelle, 2004), it has been adapted for use in a variety of other health and mental health contexts, including the creation of indicators of quality of care for children with ADHD, conduct disorder, and major depression (Zima et al., 2005) and the identification of indicators of quality care for elderly patients undergoing surgery (McGory, Shekelle, Rubenstein, Fink, & Ko, 2005).

The RUAM incorporates elements of the NIH consensus development process, the Delphi, and the nominal group techniques. It has the following characteristics:

- Six to fourteen panelists are assembled. The panelists are well known in their fields and differ in their expertise.
- The study team compiles a state of the art review of the literature for the panelists.
- The panelists participate in a two-round rating process using a 9-point scale.
- Round 1 is usually done by each panelist independently, before a group meeting.
- A highly skilled moderator conducts a meeting to discuss the ratings and clarify individual concerns. After discussion, the panelists do their ratings a second time; this constitutes the second round.
- The second and final ratings are used for the statistical analysis. Do the panelists agree? Disagree?

How is the RUAM used in identifying needs? Example 8.2 shows how the RUAM was used in developing a manual to address the intermediate and long-term mental health needs of students and staff after incidents of school-related violence. In this RUAM, 10 panelists were given a literature review and asked to rate scenarios in a two-round process. The first round of ratings was conducted by each panelist in his or her office while the second was done in a telephone group session.

Surveys

Surveys are usually used to gather information from large numbers of people. Several types are possible. A face-to-face interview may

Example 8.2 Using the RAND-UCLA Appropriateness Method (RUAM): School-Related Violence and Intermediate and Long-Term Mental Health Needs

Problem: Many American communities have suffered and continue to suffer from school-related violence. School administrators and teachers have found that the mental health recovery needs of students and staff often span months or even years beyond the initial date of the tragic event. What are the needs of administrators and teachers in dealing with this problem?

Method:

Panel Composition

Ten teachers, administrators, school police, and school mental health personnel who represented urban, suburban, and rural school districts from geographic regions across the United States agreed to participate.

Round 1

The panelists were mailed a review of the literature on immediate, intermediate, and long-term mental health services after school-related violence. At the same time, they were asked to review five scenarios, based on actual events. They were also asked to respond to 34 questions about each scenario. The questions focused on which mental health services would be appropriate after each event in four different time frames: in the first week after a violent event, between one week and one month after the event, between one and six months after the event, and more than six months after the event.

One of the five scenarios was as follows:

Sample Scenario Used for a RUAM: School-Related Violence

At 10 a.m. on Thursday morning, students and teachers in several classrooms at South Suburban High School heard a loud noise. No one was sure what it was, but, a few seconds later, there was another, similar noise, this time followed by screams. Teachers followed school crisis procedures by locking classroom doors and making the students get down on the floor, away from the windows. Cellular phones from throughout the school peppered the 911 operators with calls.

Police, firefighters, and paramedics were at the scene within minutes, and they cautiously began evacuating the school. Gradually they discovered the classroom from which the gun had been fired. Its door was locked, the teacher lying bleeding, but still alive, in the hallway. Most of the students had escaped, but there were still seven inside, being held hostage by Tom Georgian, their fifteen-year-old classmate, who was armed with a semi-automatic pistol.

Hostage negotiators tried without success to engage Georgian in conversation. Yelling and crying could be heard from the classroom, and then screaming, and more shots. A window opened, and a sobbing girl started trying to climb out, despite the fact that the classroom was on the second floor. Emergency personnel yelled to her to stop, and she began screaming, "He's blocking the door! I can't get out! There's blood everywhere, and I can't get out!"

Firefighters brought a ladder up to the window, and helped the screaming girl and three other terrified but unharmed students down to the ground. Inside the classroom, one of the hostages was dead and two others were wounded. Tom Georgian had shot himself in the head, and his body was blocking the classroom door.

Mr. and Mrs. Georgian had seen no signs of trouble in their son. He had left no clues to his intentions. The surviving hostages said he had been incoherent but that he seemed to be saying something about the school play, which had been scheduled to begin rehearsals the next day.

School administrators closed the school temporarily and met to determine how to proceed.

One of the 34 questions about the scenario was this:

Question Asked in Connection With the Scenario

For the **first week** following this event:

Teachers should provide trauma-specific mental health services to students in the classroom.

1	2	3	4	5	6	7	8	9

Very inappropriate Uncertain Very appropriate

Panelists were asked to rate the appropriateness of services on a scale of 1 to 9, with 1 being the least appropriate and 9 being the most appropriate. A rating of 7, 8, or 9 indicated that the panelist thought that a service was "definitely appropriate" and that the benefits of the service exceeded the costs or risks by a sufficient margin, that is, that the service was definitely worth providing. A service was deemed "definitely inappropriate" if the costs or risks exceed the benefits (ratings 1, 2, or 3). If the benefits and costs were determined to be about equal, or if there was not enough information to make a definite decision, then panelists were asked to assign a rating of 4, 5, or 6.

The results were compiled and summarized into a re-rating form. This form was very similar to the original questionnaire, but it indicated how many panelists had selected each answer, identified the panelist's original answer, and provided a graphic median and inter-quartile range for each question. This is an example of the summary of the ratings.

(Continued)

Example 8.2 (Continued)

Example of the Ratings for Round 1 Summarized

1	*2*	*2*				*2*	*1*	*2*
1	2	*3*	4	5	6	7	8*	9

Very inappropriate Uncertain Very appropriate

In this example, one (*1*) panelist rated the appropriateness as 1, two (*2*) rated it at 2, two (*2*) at 3, one (*1*) at 6, two (*2*) at 7, and two (*2*) at 9. This particular panelist rated the appropriateness as 8 because the 8 has an asterisk (*).

Disagreement/Agreement

Did the panelists agree with one another? A standard definition for disagreement is that it occurs if 3 or more panelists rate a statement in the high range (7–9) AND 3 or more rate it in the low range (1–3). Any other combination of ratings is considered agreement. In the example above, the panelists disagreed.

Round 2

Panelists were given the summary of the round 1 ratings. They then participated in a two-hour telephone conference to discuss the ratings, with particular emphasis on disagreement. After discussion, the panelists re-rated the scenarios. Needs for services were recommended if, after the second round of ratings, a service received a rating of 7 or higher (very appropriate) *and* there was no disagreement.

result in in-depth information but requires a skilled interviewer. Telephone surveys also require skilled interviewers, but it has become increasingly difficult to get people to agree to participate. People hanging up, the need to call back, and messages left on voice mail are costly.

E-mail, Web-based, or electronically distributed surveys can reach large numbers of people. However, technical expertise is needed to design the survey, and you have to make certain that the people who are selected to participate can access the survey, and complete it.

Mail surveys are in some ways the simplest because they do not require staff training (as do interviews) or technical expertise. The number of people who return questionnaires that are mailed to them, however, is often extremely low unless they are given an

Example 8.3 A Survey of the Needs of Women With Diabetes

Purpose: The purpose of this survey was to explore food purchasing, preparation, and consumption among black women with type 2 diabetes mellitus in an urban setting to assess barriers to medical nutrition therapy recommendations.

Methods: A telephone survey asked about shopping habits, the use of community resources for food supplementation, use of restaurant/fast-food establishments, dining habits, food purchasing and consumption, and food preparation methods. The survey contained 38 items.

Results: The respondents identified ways in which their participation in a culturally competent intervention of diabetes care and education helped them to change their dietary behaviors. The most common areas of change included purchasing, preparation, and portion size. The most commonly cited barriers to medical nutrition therapy included low income, time constraints, competing demands, and knowledge deficits.

Conclusions: Culturally sensitive diabetes interventions are an effective way to overcome some of the barriers to medical nutrition therapy. The results of the survey suggest that identification of more affordable healthy food resources in the community is needed. In addition, transportation to grocery stores should be on the public policy agenda. Finally, alternate sites for nutrition education, such as a supermarket forum, warrant further investigation.

SOURCE: From Galasso P, Amend A, Melkus GD, Nelson GT. Barriers to medical nutrition therapy in black women with type 2 diabetes mellitus. *Diabetes Educ.* Sep-Oct 2005;31(5):719-725. Used with permission of Sage Publications, Inc.

incentive to do so. The use of incentives in surveys may be costly and may boost the number of responses only slightly. Writing survey questionnaires requires a great deal of skill. Poorly designed questionnaires result in low response rates. Low response rates mean invalid survey results.

An illustrative use of a survey to identify needs is given in Example 8.3 (Galasso, Amend, Melkus, & Nelson, 2005) . The purpose of the survey was to assess the nutritional needs of black women with type 2 diabetes as well as the barriers they encounter in meeting these needs.

Table 8.2 summarizes the objectives, characteristics, advantages, and limitations of some of the main needs assessment methods. They should be considered as part of the research consumer's store of techniques for finding out about individual and community needs and about the place of new programs and practices in meeting those needs.

Table 8.2 Needs Assessment Methods

Name	*Objectives*	*Implementation*	*Advantages*	*Limitations*
Key Informant	Gather views from members of community who know the community from the "inside" because of their training or affiliations	Identify key informants (e.g., elected officials, religious and public services leaders, professionals, teachers, lawyers) Interview or survey them	Relatively inexpensive Provides perspectives from different viewpoints Can get "insider's" view	Limit on how many people can be included and so results may not be representative Cannot be sure that all "key" people are included
Public or Community Forum	Get views from a wide range of community residents in public meetings	A sponsoring agency puts on public forums A list of discussion questions is generated by the sponsors	A large number of views can be heard Enables people to participate in generating ideas about their community Relatively easy to implement Provides a quick view of community needs and preferences	Requires good leadership and advance organization Only some people might attend and so the views expressed may not be representative May generate more questions than answers
Focus Group	A small group technique with up to 10 people The purpose is to answer no more than about 5 important questions in two hours	A trained moderator leads a discussion of each question in a "permissive" atmosphere (that is, "anything goes")	Provides answers to important questions in a relatively short period of time	Need a trained and experienced moderator Questions must be formulated so they are important and answerable by the group within the time allotted May be costly because you need several groups and each person must be compensated financially

Name	*Objectives*	*Implementation*	*Advantages*	*Limitations*
Nominal Group Process	A small group process in which all members' views are heard The purpose is to make decisions on competing alternatives	A moderator and a group of people are assembled Ideas are put on paper Discussion leads to listing ideas on a board Ideas are voted on	Leads to definite conclusions Gives all participants an equal opportunity to express their views	Need an experienced moderator who makes sure everyone is heard Results may not be representative because group size is relatively small
Delphi	A method for generating ideas and also coming to consensus	Respondents complete a questionnaire The results are summarized and sent to participants Participants review the summary and complete the questionnaire a second time Process is repeated until general agreement is reached	Participants are unknown to one another, which encourages frankness Conducive to independent thinking Can be used to get agreement among groups that are hostile to each other because of the anonymity of participation	Participants may not be representative Tends to encourage a middle-of-the-road view Can be time consuming because method requires several "rounds" Definition of agreement may be considered arbitrary
RAND/UCLA Appropriateness Method (RUAM)	A panel method to identify areas of agreement, disagreement, and uncertainty (not enough evidence to make a decision)	Panelists are given a review of the literature and sent a set of items to rate On their own, they rate the appropriateness of each item using a standardized scale, where 1 = definitely not appropriate and 9 = definitely appropriate	Provides an opportunity to combine existing research with expert opinion Statistics on reliability and validity of the process are known and acceptable	An expert leader is essential The process is expensive because it is relatively lengthy Potential bias in choosing panelists whose views are likely to be accepted

(Continued)

Table 8.2 (Continued)

Name	*Objectives*	*Implementation*	*Advantages*	*Limitations*
		The ratings are summarized and given to the panelists The panelists meet face-to-face in a one to two day session to clarify ratings and discuss them Panelists rate each of the items once again Extent of agreement, disagreement, and uncertainty is identified	Process is flexible and can be used in a variety of settings and for a variety of topics	Does not force consensus and allows for uncertainty as well as providing a reliable portrait of current knowledge
Surveys	To obtain information from a large number of participants	A survey questionnaire (interview, mail, online) is developed The sample is selected The survey is conducted	Can reach a relatively large number of people with diverse perspectives	Requires expertise in survey research including questionnaire development, sample selection, and data analysis (including statistical analysis) Can be costly to reach some people who do not like written surveys (mail or Web)

Facing Uncertainty and Coming to Consensus

Research consumers may find themselves confronted with the need to make decisions on topics for which data are inconsistent, of poor quality, or even missing. The need to "do something" in situations of uncertainty is certainly characteristic of the health professions. In response, the U.S. National Institutes has supported a program since 1977 whose aim is to produce "state-of-the-science" reports (www.consensus.nih.gov). Although not yet widely implemented in the other helping professions, **consensus development conferences** are an integral part of evidence-based health care, and their methods can be usefully adapted to other fields. In fact, as can be seen from the sample topics in Table 8.3, many of the NIH reports are applicable to a wide variety of disciplines other than health care.

NIH Consensus and State-of-the-Science statements are prepared by independent panels of health professionals and public representatives on the basis of (1) the results of a systematic literature review, (2) presentations by investigators working in areas relevant to the conference questions during a two-day public session, (3) questions and statements from conference attendees during open discussion periods that are part of the public session, and (4) closed deliberations and a production of a report by the panel during the remainder of the second day and morning of the third. Consensus statements are independent reports of the panel; they are not policy statements of the NIH or of the federal government of the United States. Each statement reflects the panel's assessment of knowledge available at the

Table 8.3 National Institutes of Health Consensus Development Conferences: A Small Sample

Manifestations and Management of Chronic Insomnia in Adults
Preventing Violence and Related Health-Risking Social Behaviors in Adolescents
Symptom Management in Cancer: Pain, Depression, and Fatigue
Diagnosis and Treatment of Attention Deficit Hyperactivity Disorder
Interventions to Prevent HIV Risk Behaviors
Effective Medical Treatment of Opiate Addiction
Sunlight, Ultraviolet Radiation, and the Skin
Treatment of Destructive Behaviors in Persons With Developmental Disabilities
The Health Benefits of Pets
Pain, Discomfort, and Humanitarian Care

time the statement was written. Thus, it provides a “snapshot in time” of the state of knowledge on the conference topic.

One consensus development conference, Preventing Violence and Related Health-Risking Social Behaviors in Adolescents, was a two-and-a-half-day conference at the U.S. National Institutes of Health. The conference participants examined and assessed the current state of knowledge regarding adolescent violence and related health-risking social behavior and identified directions for future research.

Twenty-one experts presented the latest research findings on risk and protective factors involved in the development of adolescent violence and related behaviors and on interventions to reduce those behaviors. The presenters had expertise in public policy, social learning and development, pediatrics, psychiatry, community development, psychology, social medicine, violence prevention, sociology, and nursing.

After a day and a half of presentations and public discussion, an independent panel of 13 people weighed the available evidence and drafted a statement addressing the following questions:

1. What are the factors that contribute to violence and associated adverse health outcomes in childhood and adolescence?
2. What are the patterns of co-occurrence of these factors?
3. What evidence exists on the safety and effectiveness of interventions for violence?
4. Where evidence of safety and effectiveness exists, are there other outcomes beyond reducing violence? If so, what is known about effectiveness by age, sex, and race/ethnicity?
5. What are the commonalities among interventions that are effective and those that are ineffective?
6. What are the priorities for future research?

The panelists had expertise in pediatrics, psychiatry, law, sociology, nursing, research methods, adolescent health, and social work. In their statement, they gave specific examples of programs that effectively reduced arrests or precursors to violence. They also discussed the characteristics shared by these programs, such as being derived from sound theoretical rationales, addressing strong risk factors, involving long-term treatments (lasting a year and sometimes much longer), working intensively with those targeted for treatment and often using a clinical approach, following a cognitive/behavioral strategy, being multimodal and multicontextual, focusing on improving social competency and other skill development strategies for targeted youth and their families, being developmentally appropriate, not being delivered in coercive institutional settings, and having the capacity to be delivered with fidelity.

The panelists also named specific programs that are not effective even though they are in use (and in some cases widely) and cited reasons for the lack of success of these programs. These reasons included implementation protocols that are not clearly articulated, staff that are not well supervised or held accountable for outcomes, programs limited to scare tactics, programs limited to toughness strategies, and programs that consist largely of adults lecturing at youth.

The panelists concluded (among other things) that some interventions have been shown by rigorous research to reduce violence precursors, violence, and arrest. However, many interventions aimed at reducing violence have not been sufficiently evaluated or proven effective, and a few widely implemented programs have been shown to be ineffective and perhaps harmful. The panelists also recommended funding sufficient to promote the dissemination of violence prevention programs that have been shown to be effective through rigorous RCT (randomized controlled trial) research. Funding, they said, must include support for research, and monitoring must continue as these programs are more widely implemented.

NIH Consensus Development Conferences are the gold standard for developing agreement. And the NIH literature reviews that form the research base for these panel reviews are world class. Also, the speakers and panelists participating in NIH consensus development are internationally renowned. Of course, these characteristics encourage acceptance of the conference statements.

Improvement Evaluations: How Are We Doing? Can We Do Better?

The fifth step of evidence-based medicine or evidence-based practice calls for practitioners to evaluate their own effectiveness and efficiency and seek ways to improve both next time (Chapter 1). 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 evidence-based methods made a difference in the processes or outcomes of care.

Evaluation has a long history as a mechanism for improvement. The American Evaluation Association states that purposes of evaluation include bettering 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 search for evidence that matters is almost completed when the evidence about needed programs has been compiled. If a program is to be adopted or a campaign to change practices is to be implemented, an evaluation of the effectiveness of the entire process and its outcomes is in order. The purposes of the evaluation would be to (1) examine the extent to which the newly adopted programs met clients' needs and (2) identify where improvements should be made to improve future efforts to find evidence that matters. The type of evaluation that is used to accomplish these purposes is called an **improvement evaluation**.

The ideal improvement evaluation takes place as follows:

1. The evaluators identify the services, programs, or outcomes that should be improved based on a needs assessment. This can be done by reviewing existing records (e.g., school or case records); observing behavior (e.g., in a classroom, a village); or surveying members of the community (e.g., politicians, leaders) and practitioners (e.g., nurses, social workers). Requests or demands for change can also come from political pressure or legislation.
2. The evaluators set performance standards. These are the benchmarks against which improvement is measured, and, to the extent possible, they should come from evidence-based sources. In other words, they should be selected because evidence exists that, if the standards are met, beneficial results are more likely to occur than otherwise.

In the health field, performance standards are derived from research findings and the consensus of experts. The American Diabetes Society, for example, has issued clinical practice recommendations for the diagnosis and treatment of diabetes, while the Joint National Committee on the Prevention, Detection, Evaluation, and Treatment of High Blood Pressure has released recommendations for the prevention and treatment of high blood pressure. Proponents of evidence-based health care often use published recommendations or guidelines as the basis for setting performance standards in improvement evaluations. For instance, here in Example 8.4 is a very small portion of the practice recommendations issued by the American Diabetes Association:

Example 8.4 Psychosocial Assessment and Care Recommendations

- Preliminary assessment of psychological and social status should be included as part of the medical management of diabetes.
- Psychosocial screening should include but is not limited to attitudes about the illness, expectations for medical management and outcomes, affect/mood, general and diabetes-related quality of life, resources (financial, social, and emotional), and psychiatric history.
- Screening for psychosocial problems such as depression, eating disorders, and cognitive impairment is needed when adherence to the medical regimen is poor.
- It is preferable to incorporate psychological treatment into routine care rather than wait for identification of a specific problem or deterioration in psychological status.

SOURCE: Derived from the American Diabetes Association (2006).

3. An evidence-based program is selected and adapted for implementation. Implementing programs is an extremely complex activity. Effective strategies for translating solutions from one setting to another are not readily available. Also, introducing a new “solution” may pose new problems if the organization is not prepared or is resistant to it.

4. The program is implemented. The program’s implementation should be monitored to make certain all of its components are being put in place as planned. If any component is difficult to implement or proves to be unsatisfactory to participants, revision may be necessary.

5. The evaluators assess performance using the agreed upon standards and measures.

6. The evaluators provide data on the results to all participants so that they can review their progress toward meeting the identified need.

7. If the need has been met, participants decide if they want to continue with the program. If the need has not been met, participants decide if they want to continue or revise their activities until the standards are achieved.

Implementing and evaluating programs to improve the quality of services and care is an emerging discipline. EBM researchers have urged caution in adopting programs without a solid evidence base and careful evaluation (Auerbach, Landefeld, & Shojania, 2007).

Improvement and Effectiveness Evaluations: Two Purposes for One Discipline

What are the differences between evaluations that are conducted primarily to improve provision of services to meet specific needs and evaluation research that is designed to provide evidence of effectiveness? Listen in on a conversation between an improvement and an effectiveness evaluator (Table 8.4).

Table 8.4 A Conversation Between an Improvement and an Effectiveness Evaluator

Improvement evaluator: We are planning to find out if an improvement evaluation will improve the quality of care we give to patients with diabetes or high blood pressure.

Effectiveness evaluator: Where are the patients located?

Improvement evaluator: In a rural health clinic. We have three physicians in the clinic and two full-time registered nurses.

Effectiveness evaluator: Why did you decide to focus on diabetes and high blood pressure?

Improvement evaluator: These are the two most common medical problems treated in the clinic. We looked at a sample of medical records and found that many patients did not receive recommended care.

Effectiveness evaluator: What do you mean by recommended care?

Improvement evaluator: Recommended care are the standards—processes and practices—that are likely to result in optimal outcomes based on the conclusions of experts like the American Diabetes Association and the Joint National Committee on the Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. The practices are based on clinical expertise and research-based evidence.

Effectiveness evaluator: You say the evaluation aims to improve the quality of care. Are you interested only in the process of care, that is, what is done to and for the patient, or are you also interested in the outcomes?

Improvement evaluator: We are interested in both.

Effectiveness evaluator: So what will the clinic do to improve care?

Improvement evaluator: The physicians and nursing staff have created a program that they are fairly sure will work for them. It includes providing quarterly feedback to physicians, empowering nurses to remind patients of the essentials of their diabetes and high blood pressure care, and flagging medical records to remind physicians of the care due for each patient. The program was chosen because a recent analysis of 25 studies showed that physician reminders improved the quality of care for diabetic patients. Patients will be given educational materials so that they can see the standards we were aiming for. Two recent studies have shown that patients who are informed of the standards of care are more likely to be compliant with treatment than those who are not.

Effectiveness evaluator: How will you decide which patients are eligible to be in the evaluation?

Improvement evaluator: We have decided that they must be 18 years of age or older and have diabetes and/or high blood pressure. We have about 5,000 patients in the clinic, and we anticipate that, by the end of the study's first 12-month period, we will have about 250 diabetics and 650 patients with high blood pressure enrolled in the evaluation.

Effectiveness evaluator: Do you have a comparison group?

Improvement evaluator: Not one that is created especially for this evaluation. The clinic is interested in improving its performance to meet established standards. In essence, the standards are what an ideal control group who is receiving "perfect" care would receive. Education experts might call this a **criterion-referenced** design.

Effectiveness evaluator: How will you evaluate whether or not improvement takes place?

Improvement evaluator: The standards we plan to use are evidence based and state that all diabetic patients should have an eye exam, be vaccinated for pneumococcal pneumonia, and have a cholesterol screening test. The standards also state that patients with high blood pressure should achieve a measure of less that 140/90mm hg and should be taking a "baby" aspirin (81mg) daily. We have done extensive needs assessments and have found out the proportion of patients who currently receive the recommended care. To find out if the standards are met, we will compare the proportions achieving the criterion before we begin the intervention and in two months. Our aim is not just to increase the proportion, but to ensure that every patient receives recommended care to meet his or her particular, individual needs.

Effectiveness evaluator: How will you keep track of the patients and the changes in care?

Improvement evaluator: We have an electronic medical record system and plan to hire a programmer and statistician to help us. However, before we do the hiring, I am going to check out some ordinary data management programs to determine if they can handle the data management and the statistics.

Effectiveness evaluator: We are both evaluators, but my perspective is different from yours. I do evaluation research, which means that I rely upon experiments to test the comparative outcomes, impact, and costs of programs. You, on the other hand, evaluate whether programs designed to meet specific needs live up to evidence-based standards.

Table 8.5 compares effectiveness and improvement evaluations.

The effectiveness and improvement evaluators' perspectives are compared in Table 8.6. Effectiveness evaluators aim to determine if a program is successful when compared to an alternative. Improvement evaluators want to find out if a successful program is specifically beneficial in their setting.

Table 8.5 Effectiveness and Improvement Evaluations: Compare and Contrast

	Effectiveness Evaluation	*Improvement Evaluation*
Study Objective	Hypotheses or research questions are derived from reviews of previous research. Purpose is to find out if a program "works": Did it accomplish its objectives? What were its outcomes, impact, and costs?	The aim is to determine if evidence-based programs and practices have progressed to satisfying a public need. If not, can the newly adopted program be improved? Was the process of choosing the project meticulous?
Main Outcomes	A program is effective if its outcomes compare favorably to comparable alternative programs or to any new program and are clinically or practically meaningful.	Improvement is manifest if evidence-based standards are achieved; comparisons to alternative programs are sometimes made to determine the extent to which the new and the alternative meet justifiable standards. Is the new program more effective? Cost-effective?
Programs/ Interventions	The choice of programs to evaluate depends upon which ones have a sound theoretical base and promising evidence of effectiveness based on previous research. The implementation of the program is carefully monitored to make certain it is implemented as planned.	Recommended programs should have sound research evidence suggesting a high likelihood of success. Consultation with clients or the public can ease the new program's transition into the existing culture of the community, organization, etc. The implementation of the program is carefully monitored to make certain it is implemented as planned.

	Effectiveness Evaluation	*Improvement Evaluation*
Research Design	Research designs are rigorous (e.g., randomized controlled trial).	Research designs are not always a critical concern; however, unless rigorous designs are used, the evaluation results may only apply to a particular setting. If control groups are used, the comparability of the groups at baseline must be established and the control program must be a justifiable alternative to the experimental one.
Sampling	Sampling methods may be complex; inclusion and exclusion criteria are very specific and relatively restrictive; sample size is a concern: It must be large enough to detect a true difference.	Sampling may not be a major concern because everyone who is eligible is invited to participate in the improvement evaluation. Inclusiveness is often a goal.
Data Collection	Data collection relies on demonstrably reliable and valid measures.	Data collection is focused on measuring progress toward meeting needs; measures are often adapted for the local setting and must be tested to ensure that they “fit in” and still produce valid data.
Data Analysis and Data Management	Data management and analysis are often multi-faceted; expert statistical knowledge is essential.	Although expert statistical assistance is recommended, it may not always be necessary. Ability to maintain a database and use inferential statistics is almost always required.

Table 8.6 The Effectiveness and Improvement Evaluators' Perspectives

The Effectiveness Evaluator's Perspective: "Here is a program that is grounded in theory and is likely to be effective, based on findings from previous research in a similar population and setting. I plan to do a randomized controlled trial to find out if the program achieves its objectives in this population and setting and if it is associated with beneficial outcomes when compared to an appropriate alternative program."

The Improvement Evaluator's Perspective: "Here is a program with evidence that matters. It has been selected after an extensive review of the literature and consultation with experts and other decision makers. We held public meetings to discuss the program and received positive reactions. We have pilot tested the program in our organization, with favorable results. Now, I am interested in finding out if the program increases the number of people being served and improves the quality and outcomes of their services. I plan to use a comparison group to help us decide if any observed differences in numbers and quality are truly due to the program."

Research and Ethics: An Indomitable Connection

Research with human participants raises ethical concerns because people accept risks and inconvenience in order to contribute new knowledge and provide benefits to others. Research consumers are not responsible for ensuring that research has been conducted in an ethical manner, but they can benefit greatly if they understand the characteristics of ethical research. Consumers are responsible for the use of research results, so it is important for them to learn about the ethical consequences of applying evidence-based practices. In recognition of the strong link between research and ethics, many medical journals require that authors state that their study protocol was reviewed and approved by an **ethics committee** or **institutional review board (IRB)**.

Research and the Institutional Review Board

An **institutional review board** (**IRB**) or ethics committee is an administrative body whose purpose is to protect the rights and welfare of **human research subjects** who are recruited to participate in research activities. Research is defined by the U.S. Department of Health and Human Services (DHHS) as systematic investigation (including research development, testing, and evaluation) designed to develop or contribute to *generalizable knowledge*. The key point here is that knowledge resulting from research must be presumed in advance to apply to other people in other settings. Thus, using the DHHS definitions, **effectiveness evaluations** are research whereas **improvement evaluations** are usually not. (The exception, which we discuss below, occurs when the results of improvement evaluations are considered generalizable and made public through publication.)

According to the DHHS a **human subject** is a living individual about whom an investigator (whether a professional or a student) conducting research obtains (1) data through intervention or interaction with the individual (e.g., in a counseling session or a classroom) or (2) identifiable private information (e.g., birth date or school record number). (For more information about this and other definitions, see the U.S. Department of Health and Human Services, 2005.) The IRB is in charge of determining if the research is structured to guarantee that each participant's privacy and rights are protected. If it is, the research can proceed. If it is not, the IRB will not allow any data collection. All major and reputable social, health, and welfare agencies (school districts, departments of mental health and social services, health departments, and so on) have ethics committees and protection requirements for **human subjects**. Research that receives any U.S. government support (e.g., from the National Institutes of Health, the National Science Foundation, the U.S. Department of Education) must be formally approved by an IRB or ethics committee that itself has been approved by the U.S. Office for Human Research Protections (OHRP: http://www.hhs.gov/ohrp/). Many other countries are equally rigorous as the United States in applying human subject protection, and most of the principles are similar if not identical.

Three Guiding Principles

According to the U.S. government, all IRB activities related to human subjects research should be guided by the ethical principles in *The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research* (www.ohrp.osophs.dhhs.gov). *The Belmont Report* was prepared by the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research in 1979 and is still the foundation for ethical research. Three major principles come from the Belmont Report:

Respect for Persons. Respect for persons requires investigators to obtain **informed consent** from research participants, to protect participants with impaired decision-making capabilities, and to maintain confidentiality.

Beneficence. This principle requires that research design be scientifically sound and that the risks of the research be acceptable in relation to the likely benefits. The principle of beneficence also means that persons are treated in an ethical manner not only by respecting their decisions and protecting them from harm, but also by actively making efforts to secure their well-being.

Justice. Justice refers to the balance between receiving the benefits of research and bearing its burdens. For example, to ensure justice, the selection of research participants needs to be scrutinized in order to determine whether some classes (e.g., welfare recipients, persons in institutions) are being systematically selected simply because of their easy availability rather than for reasons directly related to the problems being studied.

U.S. government policy also mandates that an IRB must have at least five members, with varying backgrounds. When selecting members, the IRB must take into account racial and cultural heritage and be sensitive to community attitudes. In addition to possessing the professional competence necessary to review specific research activities, the IRB members must also be able to ascertain the acceptability of proposed research in terms of institutional commitments and regulations, applicable law, and standards of professional conduct and practice.

U.S. government policy requires that, if an IRB regularly reviews research that involves a vulnerable category of participants (such as children, prisoners, pregnant women, or handicapped or mentally disabled persons), it must consider the inclusion of one or more individuals who are knowledgeable about and experienced in working with these participants. Also, the IRB must make every nondiscriminatory effort to ensure that it does not consist entirely of men or entirely of women.

Table 8.7 lists the major criteria used by IRBs and ethics committees in approving research protocols.

Obtaining Informed Consent

The informed consent process requires researchers to disclose information that will be relevant to the potential participant's decision about whether to participate in the research. Disclosure means answering questions such as these: Why is the research being done? What will participants do? What are the risks and benefits of participating?

Informed consent is usually obtained in writing. If written consent cannot be obtained (participant is blind or cannot write), then the researchers must provide evidence that consent was administered (say, on the phone) and understood. The consent form is designed to

Table 8.7 Criteria Used by an Institutional Review Board (IRB) in Approving Research Protocols

- **Study Design:** Many experts agree that an IRB should approve only research that is both valid and of value. The thinking is that a poorly designed study will necessarily lead to misleading results. Study design includes subject recruitment, selection, and assignment to groups; measure or instrument reliability and validity; and data analysis.
- **Risks and Benefits:** IRBs evaluate whether the risks to participants are reasonable in relation to the anticipated benefits, if any, to the participants, and they asses the importance of the knowledge reasonably expected to result from the research.
- **Equipoise:** The ethical basis for assigning treatment by randomization is the judgment that current evidence does not favor the superiority of the experimental over the control program.
- **Equitable Selection of Participants:** The IRB usually considers the purpose of the research and the setting of the research and closely examines studies involving vulnerable populations, such as children, prisoners, participants with cognitive disorders, or economically or educationally disadvantaged people.
- **Identification of Participants and Confidentiality:** The IRB is required to review the method for prospective identification of research participants. IRB members examine the researchers' means of identifying and contacting potential participants and the methods for ensuring the participants' privacy and confidentiality.
- **Participant Payment:** Many medical and health-related studies provide financial and other incentives to study participants to compensate them for their time. Ethical concerns arise if the payment is high or too low. If the payment is high, some participants may be induced to take risks against their better judgment. If the payment is too low, some participants may not believe the study is worth their time.
- **Qualifications:** The IRB examines the qualifications of the evaluator and the evaluation team. In addition, the IRB considers the facilities and equipment used to conduct the research and maintain the rights and welfare of the participants.
- **The Informed Consent Process: Informed consent** means that participants who agree to participate in the research are knowledgeable about the risks and benefits of participation and the activities that comprise participation. They also agree to the terms of participation and are knowledgeable about their rights as research subjects.

protect all parties: the participant, the researcher, and the institution. Therefore, it is important that consent forms present information in an organized and easily understood format.

In some studies, researchers design separate informed consent forms for parents and assent (verbal) forms for children.

Table 8.8 contains the contents of an informed consent form that should be discussed with research participants.

Table 8.8 The Contents of Informed Consent

1. **The Characteristics of the Research.** The participant should be told directly that research is being conducted, what the research's purpose is, and how participants were chosen.
2. **The Study Procedures.** Participants should be told what they are going to be doing in the project, how much time will be needed, and when participation will begin and end. Alternative procedures should be discussed as should blinding or randomization.
3. **Potential Risks and Benefits.** Although some research may pose significant risks (a new drug therapy, for example) other research may not. Almost all research results in some discomfort (e.g., feeling embarrassed by questions about drug or alcohol use).

   Also, participation in research is sometimes mistakenly assumed to mean benefit, especially for people who know they are in the experimental group (not blinded). Researchers should state clearly that they do not know if the experimental treatment is better than the control or an alternative.
4. **Assurance That Participation Is Voluntary.** Participants should be told that they are free not to enroll or to enroll and drop out.
5. **Procedures to Maintain Confidentiality.** These include coding research data, storing it in locked computers and file cabinets, and limiting the number of people who have access to it. These measures should be discussed with research participants.

The Special Case of Evaluations That Are Exempt From IRB Approval

A program evaluation is considered to be research by many IRBs when the evaluator intends to create generalizable knowledge that will be shared outside of the program being evaluated in professional presentations, reports, or published articles. In all likelihood, process and implementation and improvement evaluations will not be considered research. Process and implementation evaluation data are used, typically, to assess progress and better understand operations within a program, while improvement evaluations are designed to assess quality within an institution. The results of these evaluations are almost always not designed for publication. If the evaluators do not intend to generalize or publicize the results, their studies may be exempt from IRB scrutiny. Consider Example 8.5.

Limits to Confidentiality

Depending on the research's aims, there may be limits to the investigator's promise of confidentiality to the subject. An example would be if a participant reveals information about child or elder abuse, and you were required by law to report this information.

Example 8.5 An Evaluation That Is Exempt From IRB Approval

The Health Center wants to improve its influenza vaccination rate. An automatic e-mail system is set up to remind physicians of their patients who are due for the vaccination. An evaluation is conducted of the effectiveness of the e-mail reminder system, and data are collected each year for two years. Information from the evaluation will not be shared with anyone outside the Health Center.

Comment: This evaluation is probably exempt from being reviewed by the ethics committee because the findings are going to be used only by the Health Center. It does not conform to the definition of human subjects' research, which results in generalizable information and may be published.

Research Misconduct

Research misconduct includes such factors as fabrication, falsification, and plagiarism. Fabrication means making up results and recording or reporting them. Falsification includes changing or omitting data or results. Plagiarism means taking another person's ideas, results, or work without giving due credit.

Research misconduct is becoming an increasingly important concern throughout the world. The following (Table 8.9) are problematic behaviors and definitions of misconduct that may apply to many situations in which evaluations are conducted.

Faking the data is a clear example of research misconduct. More subtle examples include

- Exaggerating findings to support the researcher's point of view
- Changing the research protocol or method of implementing the program without informing the IRB before doing so
- Failing to maintain adequate documentation of the research methods (such as preparing a code book or operations manual)
- Releasing participant information without permission to do so
- Having insufficient resources to complete the research as promised
- Having financial or other interests in the funders or supporters of the evaluation (conflict of interest)

Many agencies and professional organizations provide guidelines for ethical research. Table 8.10 lists some of these organizations and their Web sites.

Table 8.9 Problematic Behaviors in Research Leading to Charges of Misconduct

Problematic Behavior	*Definition*
Misconduct	Fabrication, falsification, or plagiarism
Questionable research practices	Actions that violate values of research and may be detrimental to the research process but do not directly threaten the integrity of the research record Examples include failing to retain research records for a reasonable period or using inappropriate statistics to enhance findings
Other misconduct, not pertaining to scientific integrity	Unacceptable behaviors subject to generally applicable legal and social penalties but that are not unique to research Examples include sexual harassment, misuse of funds, or violations of federal regulations
Other misconduct, pertaining to scientific integrity	Unacceptable behavior that does not directly affect the integrity of the research process but is nevertheless directly associated with misconduct in science Examples include cover-ups of scientific misconduct or reprisals against whistleblowers
Sloppiness	Negligent or irregular research practices that risk distortion of the research record but that lack the intent to do so

Table 8.10 Agencies That Have Policies and Guidelines for Conducting Ethical Research

American Psychological Association (APA)

- Ethical Principles of Psychologists and Code of Conduct
  http://www.apa.org/ethics/code2002.html
  The current version of this document was adopted by the American Psychological Association Council of Representatives on June 1, 2003. It includes information about issues pertaining to privacy and confidentiality, therapy, publishing, and more.
- Ethical Principles of Psychologists and Code of Conduct (1992)
  http://www.apa.org/ethics/code1992.html
  Between 1992 and 2003, the APA was guided by this ethics code.
- Guidelines for Ethical Conduct in the Care and Use of Animals
  http://www.apa.org/science/anguide.html
  This is a set of guidelines developed by APA to be used by psychologists working with animals. The document covers areas such as housing of animals, experimental procedures, and educational use of animals.

American Public Health Association (APHA): Public Health Code of Ethics

http://www.apha.org/programs/education/progeduethicalguidelines.htm

The ethical guidelines can be accessed from this Web site highlight issues that are unique to the public health field.

American Statistical Association (ASA): Ethical Guidelines for Statistical Practice

http://www.amstat.org/profession/index.cfm?fuseaction=ethicalstatistics

ASA's Committee on Professional Ethics prepared these guidelines, and they were approved by their Board of Directors on August 7, 1999. This document contains two sections: the preamble and ethical guidelines.

Applied Research Ethics National Association (ARENA)

http://www.primr.org/membership/overview.html

ARENA is a national membership organization that deals with biomedical and behavioral research issues such as scientific misconduct, ethical decision making in health care, and the protection of human and animal subjects. The group was organized in 1986.

Association for Practical and Professional Ethics

http://www.indiana.edu/~appe/

The Association for Practical and Professional Ethics was founded in 1991 with the support of Indiana University and a Lilly Endowment. Its mission is to "encourage interdisciplinary scholarship and teaching of high quality in practical and professional ethics by educators and practitioners." This site includes association information, association activities, association publications, and electronic networking opportunities.

Association of University Professors (AAUP): Statement on Professional Ethics

http://www.aaup.org/AAUP/pubsres/policydocs/statementonprofessionalethics.htm

The statement that appears at this site is a revised version of one that originally appeared in 1966. In 1987, the AAUP adopted this current document that was endorsed at its seventy-third annual meeting.

Center for Academic Integrity (CAI)

http://www.academicintegrity.org

The Center for Academic Integrity is affiliated with the Rutland Institute for Ethics at Clemson University. CAI's mission is "to identify and affirm the values of academic integrity and to promote their achievement in practice."

Council on Undergraduate Research (CUR)

http://www.cur.org/conferences/responsibility/ResRespons.html

The mission of CUR is to "support and promote high-quality undergraduate student-faculty collaborative research and scholarship." In June 2002, CUR held a major symposium titled *Research Responsibility and Undergraduates*. Manuscripts, post-conference workshop summaries, and news on guidelines related to responsible research are merely a few of the resources posted at this site.

Creating a Code of Ethics for Your Organization

http://www.ethicsweb.ca/codes/

Chris MacDonald, PhD, Philosophy Department, St. Mary's University (Halifax, Canada) has put together this site with links to resources to assist individuals and groups

---

(Continued)

Table 8.10 (Continued)

in writing a code of ethics. He discusses why organizations and institutions should even have a code and provides guidance in writing one. He also provides links to essays on ethics, sample codes, and contacts for ethics consultants.

Federal Policy on Research Misconduct

http://www.ostp.gov/html/001207_3.html

The Office of Science and Technology Policy has posted this site, which includes information on issues such as requirements for findings of scientific misconduct, responsibilities of federal agencies and research institutions, and guidelines for fair and timely procedures and agency administrative actions.

Framework for Policies and Procedures to Deal With Research Fraud

http://www.aau.edu/reports/FrwkRschFraud.html

This Association of American Universities document grew out of the belief that universities should be held responsible for the actions of their faculty and staff, not research sponsors. As a result of this belief, an interagency group got together to develop this "framework" in 1988. Areas such as "Definition of Research Fraud" and "Process for Handling Allegations of Research Fraud" are covered in this document.

Illinois Institute of Technology Codes of Ethics Online

http://ethics.iit.edu/codes/coe.html

Illinois Institute of Technology's Center for the Study of Ethics in the Professions (CSEP) developed this online collection of over 850 codes of ethics. CSEP received a grant from NSF in 1996 to put its collection of codes on the Web, a collection that grew out of CSEP's paper archive of codes. In addition to the codes, resources for authoring a code, case studies, and other information can be found at this site.

Office of Human Research Protections (OHRP)

http://www.hhs.gov/ohrp/

This OHRP site, part of the U.S. Department of Health and Human Services, provides links to IRB registration and filing information, policy guidelines, compliance oversights, educational materials, and upcoming workshop events.

Office of Research Integrity (ORI)

http://ori.dhhs.gov

The goal of the ORI is to "promote integrity in biomedical and behavioral research supported by the Public Health Service (PHS)." This site has links to resources like breaking news stories, tips for handling misconduct, publications, and policies, regulations, and statutes.

Scientific Freedom, Responsibility & Law

http://www.aaas.org/spp/sfrl/

This program is part of the AAAS Directorate for Science & Policy, and it focuses on the ethical, legal, and social issues associated with the conduct of research and with the advances in science and technology. Information and links to projects and activities, publications, and access to the PER newsletter (http://www.aaas.org/spp/sfrl/per/per.htm) can be found at this site.

University of California, San Diego: Office of Graduate Studies and Research

http://ogs.ucsd.edu/

The Office of Graduate Studies and Research at the University of California, San Diego (UCSD) has posted policies that are applicable to those doing research.

Training in research ethics and the proper conduct of research can be done online. Many institutions require that researchers complete such training before doing research with human subjects. (See, for example, the OHRP site *Human Subject Assurance Training*: http://ohrp-ed.od.nih.gov/CBTs/Assurance/login.asp.)

Practicing Research and Ethics

From an ethical perspective, the strongest arguments in support of evidence-based practice are that it allows the best evaluated programs (and useless or harmful ones) to be identified and that it facilitates informed decisions. But all may not be well because the body of research is often incomplete, methodologically flawed, or unresponsive to important social and cultural needs. Further, although evidence is emerging in health care that research-based interventions produce better outcomes, little evidence is available for other fields (although those studies are being done). Therefore, consumers who adopt evidence-based methods should do so cautiously, keeping in mind the ethical implications of the emerging field of evidence-based practice.

Table 8.11 lists some of the major ethical concerns associated with the use of evidence-based practices.

Table 8.11 Ethics and Evidence-Based Practices

Ethical Concern	*Explanation*
Many important outcomes cannot be measured.	Evidence-based practice aims to provide a simple, logical process for reasoning and decision making. But to make balanced decisions, all the relevant consequences of an action must be considered. Current measures of some outcomes (such as pain) are inadequate, while others (such as justice) may not be measurable. Further, other complex outcomes (such as quality of life) may not even be adequately definable (across cultures, generations, and over time). Often, researchers “settle” for imprecise measures or proximate outcomes.
Stakeholders’ needs may differ markedly from those of researchers and policy makers.	The community—recipients of evidence-based practices—has relatively little influence over the priorities and funding of research.

(Continued)

Table 8.11 (Continued)

Ethical Concern	*Explanation*
Because the large quantities of data required to meet the standards of evidence-based practice are available for relatively few interventions, a systematic bias may be inevitable toward those interventions.	The bias may ultimately result in the allocation of resources to those interventions for which there is rigorous evidence of effectiveness or toward those for which there are funds available to show effectiveness. This may be at the expense of other areas where rigorous evidence does not currently exist or is not attainable (such as palliative care services). Allocating resources on the basis of evidence may therefore involve implicit value judgments, which is at odds with evidence-based practices that emphasize explicit "objective" criteria.
The application of cost-effectiveness measures to decisions about who does or does not receive services may adversely affect the position of the weaker groups in our society.	People who are expected to benefit only slightly from particular programs, such as the elderly and the disabled, may be excluded from access to such programs particularly when they are expensive. Many vulnerable people have been excluded from large-scale research because of the perceived (or real) difficulty of retaining and caring for them.
Use of the term "evidence-based" may be misleading (Steinberg & Luce, 2005). Research consumers, policy makers, and others acting on the basis of recommendations labeled as being "evidence-based" should not blindly assume that the label truly applies.	EBP methods are often not applied consistently or interpreted properly. The potential exists for great variation in the validity of decisions and recommendations that claim to be "evidence-based." In addition, evidence may be available for some but not all issues related to a decision or recommendation that has to be made, or the evidence that is available may not be directly relevant to the situation to which it is being applied.

Acknowledging the limitations of evidence-based practice and its associated ethical problems should not deter research consumers from searching for the best evidence when making decisions about the selection of programs and interventions to improve the public's health and welfare. Research consumers who understand the complexity of the issues have more data to work with when making decisions than those who unthinkingly accept or reject the methods and findings of evidence-based practice. Informed consumers may even decide to use their knowledge to demand that researchers devise more publicly responsive methods of collecting evidence—so that it truly matters.

Summary of Chapter 8: The Ethical Research Consumer Assesses Needs and Evaluates Improvement

Words to Remember

administrative needs; behavioral needs; beneficence; communal or epidemiological needs; consensus development conferences; Delphi technique; educational needs; effectiveness evaluation; equipoise; ethics committee; evidence that matters; focus groups; human subject; improvement evaluation; informed consent; institutional review board; justice; key informant; needs assessment; nominal group process; physical needs; public or community forum; the RAND/UCLA Appropriateness Method; research misconduct; respect for persons; social needs; surveys

Evidence-based practitioners count on the experimental method to provide evidence that matters, and they agree on the need to incorporate users' needs, preferences and values, and expectations into treatments, practices, and programs. A systematic effort to identify user needs and provide a context for them is called a **needs assessment.**

Needs can be arranged into six categories: social, communal or epidemiological, behavioral, environmental, educational, and administrative.

At least seven methods are commonly used to determine individual and public needs:

1. The purpose of **the key informant method** is to collect information about a community's needs by interviewing community leaders who are likely to be in a position to know what the needs are.
2. A **community forum** consists of a group of people who meet together to discuss a common problem. The meeting is open to all members of the community.
3. A **focus group** is designed to collect information from "insiders" or "people in the know." The group usually consists of about 10 carefully selected participants and a trained moderator. The session lasts about two hours and is centered on getting answers to four or five carefully constructed questions.

4. In **the nominal group technique**, participants are brought together for a discussion session led by a moderator. After the topic of concern has been presented to session participants and they have had an opportunity to ask questions or briefly discuss the scope of the topic, they are asked to take a few minutes to think about and write down their responses. The session moderator will then ask each participant to read, and elaborate on, one of his or her responses. These are noted on a flipchart. Once everyone has given a response, participants will be asked for a second or third response, until all of their answers have been noted on flipchart sheets posted around the room.
5. **The Delphi technique** is a structured method of determining the degree of agreement on a topic, selecting alternatives, or setting priorities. Delphi techniques use questionnaires that are completed by participants on their own, in groups, or both. The questionnaires are structured to ask people to rate or rank the importance or validity of certain ideas.
6. **The RAND/UCLA Appropriateness Method** (RUAM) is a method for determining the extent of agreement on controversial topics and on subjects for which the research base is poor or ambiguous (a mixture of positive and negative findings).
7. **Surveys** are usually used to gather information from large numbers of people. Several types are possible. A face-to-face interview may result in in-depth information but requires a skilled interviewer. Telephone surveys also require skilled interviewers, and it has become increasingly difficult to get people to agree to participate. People hanging up, the need to call back, and messages left on voice mail are costly. E-mail, online, and mailed surveys can reach large numbers of people.

Although not yet widely implemented in the other helping professions, **consensus development conferences** are an integral part of evidence-based health care, and their methods can be usefully adapted to other fields. NIH Consensus and State-of-the-Science statements are prepared by independent panels of health professionals and public representatives on the basis of (1) the results of a systematic literature review, (2) presentations by investigators working in areas relevant to the conference questions during a two-day public session, (3) questions and statements from conference attendees during open discussion periods that are part of the public session, and (4) closed deliberations and the production of a report by the panel during the remainder of the second day and the morning of the third.

Some evaluations are designed to examine the extent to which the newly adopted programs met clients' needs and identify where improvements should be made to improve future efforts to find evidence that matters. The type of evaluation that is used to accomplish these purposes is called an **improvement evaluation**.

Effectiveness and improvement evaluators have differing goals. Effectiveness evaluators aim to determine if a program is successful when compared to an alternative. Improvement evaluators want to find out if a successful program is specifically beneficial in their setting.

Research with human participants raises **ethical** concerns because people accept risks and inconvenience in order to contribute new knowledge and provide benefits to others. Research consumers are not responsible for ensuring that research has been conducted in an ethical manner, but they can benefit greatly if they understand the characteristics of ethical research. Consumers are responsible for the use of research results, so it is important for them to learn about the limitations of research methods and evidence-based practices and how some of these limitations have ethical implications.

An **institutional review board** (IRB) or **ethics committee** is an administrative body whose purpose is to protect the rights and welfare of human research subjects who are recruited to participate in research activities. The IRB is in charge of determining if the research is structured to guarantee that each participant's privacy and rights are protected. If it is, the research can proceed. If it is not, the IRB will not allow any data collection.

Three major principles guide much of health research.

> ***Respect for Persons.*** Respect for persons requires investigators to obtain **informed consent** from research participants, to protect participants with impaired decision-making capabilities, and to maintain confidentiality.
>
> ***Beneficence.*** This principle requires that research design be scientifically sound and that the risks of the research be acceptable in relation to the likely benefits. The principle of **beneficence** also means that researchers treat people in an ethical manner, not only by respecting their decisions and protecting them from harm but also by actively making efforts to secure their well-being.

> ***Justice.*** Justice refers to the balance between receiving the benefits of research and bearing its burdens. For example, to ensure justice, the selection of research participants needs to be scrutinized in order to determine whether some classes (e.g., welfare recipients, persons in institutions) are being systematically selected simply because of their easy availability rather than for reasons directly related to the problems being studied.

The informed consent process requires researchers to disclose information that will be relevant to the potential participant's decision about whether to participate in the research. Disclosure means answering questions such as these: Why is the research being done? What will participants do? What are the risks and benefits of participating?

Research misconduct includes such factors as fabrication, falsification, and plagiarism. Fabrication means making up results and recording or reporting them. Falsification includes changing or omitting data or results. Plagiarism means taking another person's ideas, results, or work without giving due credit.

Research consumers should be aware of at least four ethical concerns in the application of evidence-based practices to their own settings. First, many important outcomes of treatment cannot be measured. Second, it may be impossible to decide between the competing claims of different stakeholders. Third, because the large quantities of data required to meet the standards of evidence-based practice are available for relatively few interventions, a systematic bias may be inevitable toward adapting those interventions. Finally, the application of cost-effectiveness measures to decisions about who does and who does not receive services may adversely affect the position of the weaker groups in our society.

Exercises

1. Name the assessment technique used in each of the following studies.

   a. **Home Injury Hazard Risks and Prevention Methods for Young Children**

   The Board requested a list of 5–7 injury hazards and 5–7 potential prevention behaviors and/or devices for children aged 1–5 years in each of the following areas of the home: bedroom/play area, kitchen/dining area, bathroom, living room, basement/garage (including other outdoor areas such as the driveway), pool, stairs/hallway, and multiple rooms/general safety. We asked participants to develop their lists of hazards by considering the *frequency, severity,* and *preventability* of the potential injury from each hazard, as well as the *efficacy* and *feasibility* of each prevention method. Efficacy was defined as the ability of the behavior or the device, if implemented, to eliminate the hazard and/or to prevent the injury. Feasibility was defined as the likelihood of implementation of the behavior or the device (depending on acquisition, installation, utilization, and maintenance).

   Round 2 asked participants to *rate* each hazard and behavior/device listed in the responses submitted to survey 1 using a scale of 1 to 3 (with 3 being highest priority). Participants could also assign a score of zero (0) if they believed that an item should not remain on the list. In rating each item, the participants were instructed to consider the same factors used in the first round (for example, children aged 1–5 years; frequency, severity, and preventability for the hazards; and efficacy and feasibility for the behaviors/devices). We calculated a mean score for each item by summing all ratings reported for a single item. Items were subsequently listed in descending order of priority.

   The 47 hazards and 52 prevention methods with the highest mean scores were selected for inclusion in survey 3 based upon natural clusters, rather than just choosing the top 50 of each.

   For the 99 selected items, the third round asked participants to *rate* each hazard using a Likert scale of 1 to 5 (with 5 being the most important) considering overall importance in an injury prevention program for preschool aged children, 3–5 years of age. This age group request differed from previous rounds as we sought to use the panel's findings for a future injury prevention program targeted at children aged 3–5 years.

SOURCE: Katcher, Meister, Sorkness, et al. (2006).

b. Impact of Smoke-Free Residence Hall Policies: The Views of Administrators at Three State Universities

Interviews with XXX aimed to (1) explore staff interpretation of trends and data, (2) assess observed changes in campus constituent with attitudes and behaviors resulting from the policy change, and (3) determine the impact of the policy change on personnel workload. We designed questions tailored to each department to elicit information and to enrich understanding of the policy's impact. As appropriate, the interviewer requested additional existing documentation during interviews.

In total, we contacted 47 personnel for interviews. Thirty campus personnel contributed to the study through telephone interviews, e-mail correspondences, providing data, or a combination thereof. We conducted 27 telephone interviews: 10 at URI, 7 at MSU, and 10 at OSU. Three additional xxx answered questions by e-mail correspondence. At MSU and URI, personnel from all identified departments, except admissions, participated in the interviews. At OSU, personnel from all 7 departments participated. In some cases, we interviewed multiple personnel from a single department. The interviewer took copious notes during the interviews and then compiled them along with e-mail correspondence into an interview report.

SOURCE: Gerson, Allard, and Towvim (2005).

2. Match each need with its appropriate definition.

Need	*Definition*
A. Social	1. Individual and communal lifestyles and beliefs that affect a community's well-being
B. Behavioral	2. Community's perceptions of its problems
C. Administrative	3. Problems that can be documented to affect a large number of people in the community
D. Communal/ Epidemiological	4. Social or physical factors that are external to an individual or a community
E. Physical	5. Policies and resources that exist in the organizations and institutions (e.g., school, hospital, business, nongovernmental organization) that might facilitate or hinder the adoption of a new program
F. Educational	6. Individual and community knowledge, attitudes, skills, and self-efficacy beliefs

3. Which of these should be applied to improvement evaluation? Check all that apply.
   - ☐ Evidence-based programs
   - ☐ Randomized controlled trials with blinded observers
   - ☐ Evidence-based performance standards
   - ☐ Very detailed study inclusion criteria

4. Which of these should be applied to effectiveness evaluations? Check all that apply.
   - ☐ Reliable and valid measures
   - ☐ Inclusive study eligibility criteria
   - ☐ Assessment of evidence-based programs
   - ☐ Flexible evaluation designs

5. Match the following statement with the concept that supports or defines it.

1. In our RCT, we went to great trouble to ensure that the alternative program has not been proven superior to the experimental.	a. Informed consent
2. Participants in the research are knowledgeable about the risks and benefits of participation and the activities that comprise participation. They also agree to the terms of participation and are knowledgeable about their rights as research subjects.	b. The ethics committee or institutional review board (IRB)
3. U.S. government policy also mandates that it must have at least five members, with varying backgrounds. When selecting members, the committee must take into account racial and cultural heritage and be sensitive to community attitudes.	c. Equipoise
4. The selection of research participants needs to be scrutinized in order to determine whether some classes (e.g., welfare recipients, persons in institutions) are being systematically selected simply because of their easy availability rather than for reasons directly related to the problems being studied.	d. Justice

6. Which of these is characteristic of research misconduct? Circle all that apply.

Plagiarism	a
Falsification	b
Conflict of interest	c
Fabrication	d

7. Which of these is a potential ethical concern when considering the adaptation of evidence-based practices? Circle all that apply.

The application of cost-effectiveness measures to decisions about who does or does not receive services may adversely affect the position of the weaker groups in our society.	a
Many important outcomes cannot be measured.	b
Because the large quantities of data required to meet the standards of evidence-based practice are available for relatively few interventions, a systematic bias may be inevitable toward those interventions.	c
It may be impossible to decide between competing claims of different stakeholders.	d

8. What is the primary reason that research with human participants raises ethical concerns? Circle one.

Participants are not always told why they are being asked to join a study.	a
Participants accept risks they might not otherwise agree to.	b
Participants frequently get paid for their participation.	c
Participants are rarely part of the ethics committee to approve the study protocol.	d

9. Which of these is a defining characteristic of the ethical principal of beneficence? Circle one.

Scientifically sound research design	a
A balance between benefits and risks	b
Informed consent	c