Suppose that you are a new therapist in an adult day-treatment center that serves older adults who are at risk of losing their ability to live independently. You have a client, Mrs. Jones, who appears to have the physical capacity to engage actively in daily life activities, including basic self-care as well as instrumental activities such as shopping at the drugstore near her home. Though Mrs. Jones dresses herself, her grooming is inadequate; she rarely bathes, watches television for most of the day, and never leaves her home except to come, reluctantly, to the day-treatment program (arranged through her family members’ efforts). Although Mrs. Jones has not been evaluated for or diagnosed with depression, you think you should be addressing depression as a possible condition while carrying out your evaluation and intervention planning procedures. Minimal participation in daily life activities is one of the factors that contributes to your impression that Mrs. Jones possibly is depressed. You recognize that you need evidence to verify the validity of your impression, so you create the following clinical question to guide your search for evidence: Among elderly men and women without depression, do those with depression have restricted participation in daily life activities compared with those who are not?

The purpose of this question is to seek information that describes factors associated with depression. The act of creating this question demonstrates a high level of clinical reasoning because you have recognized that your impression may or may not be well-founded. After formulating the question, you search for evidence, tapping the current research literature as one source (preferably the literature within the past 10 years). You find several reported studies that appear to provide relevant evidence.

In this imaginary scenario, you, as the new therapist, have now completed the first two of five steps involved in evidence-based occupational therapy (Tickle-Degnen, 2000). This installment of the “Evidence-Based Practice Forum” describes the third step: evaluating the gathered evidence to determine which of it is the best for answering the clinical question (Sackett, Richardson, Rosenberg, & Haynes, 1997). My focus is on evaluating evidence from the research literature. Future installments will describe the last two steps: communicating about the evidence to clients and colleagues and conducting ongoing evaluation of the implementation of evidence-based evaluation and intervention procedures as they are implemented with clients.

Messy Evidence

It is highly likely that the search for evidence turns up reported studies that (a) almost but do not quite answer the specific clinical question and (b) do not replicate one another exactly. It may feel as if the evidence is too “messy” to be of use, that no coherent answer can be derived from it. For example, in relation to the descriptive question formulated for Mrs. Jones, some retrieved studies may report activity participation in elderly men and women with depression but do not compare it with the participation of elderly men and women without depression. Another study may not measure the variable of interest (in this case, activity participation) but rather a different but relevant variable, such as activity preferences. One study may be of persons 50 years to 65 years of age, whereas another may be of persons 65 years to 80 years of age. A few studies may report finding no differences in activity participation when comparing those who are depressed with those who are not, whereas others may report finding important differences.

This apparent messiness is the characteristic of any body of research on human beings, especially when that body of research was not conducted with a singular clinical question and client in mind. Human beings are complex, and research design and implementation is fallible. Despite this complexity and fallibility, all is not lost. An evidence-based therapist can find the best possible answers, if not an exactly single correct one, in an apparently messy body of evidence. The therapist must first recognize that it is highly unlikely that any given study alone will provide enough evidence to answer a specific clinical question. More than likely, the best evidence will be the combined findings of more than one study.
Evaluating the Contribution of a Single Study to the Body of Evidence

Having retrieved reported studies through a thorough search of the literature, the therapist evaluates whether each study can make a meaningful contribution to the body of evidence that will be used in deriving possible answers to the clinical question. I have outlined three sets of ideal standards against which to review each study. These standards are derived from quantitative research methodologies and, as a result, cannot be easily applied to evaluating evidence from studies conducted using qualitative methodologies. (I invite readers who have expertise in qualitative research standards for evidence-based practice reviews of the literature to submit manuscripts to this forum or The American Journal of Occupational Therapy [AJOT].) The first set of standards listed here is for evaluating descriptive studies, that is, those carried out to answer descriptive questions. The second is standards for evaluating assessment studies, and the third set is standards for evaluating intervention studies. It is not expected that very many, if any, of the retrieved studies will meet all of the standards because the science of rehabilitation and occupational therapy is in a relatively formative stage. Those few studies that meet all of the standards provide the strongest contribution to the body of evidence around a clinical question. Studies that meet one or two standards provide the next strongest contribution, whereas studies that almost but not quite meet any of the standards provide the weakest evidence.

Standards for Evaluating Descriptive Studies

The descriptive study standards listed here are meant to provide general guidance and do not outline standards in the same detail as would a research methods textbook (or as in the recent “Quantitative Research Series,” in AJOT introduced by Davies and Case-Smith [1998]). However, they can be used as the beginning step in the evaluation of a single reported study:

1. The study investigated a variable that is relevant to the specific occupation, occupational performance, or component of occupational performance variable in the evidence-based clinical question. In the descriptive question of our imaginary scenario, this variable is “participation in daily life activities.” The therapist must determine whether a given study provides information that is relevant to the kinds of activities that Mrs. Jones, given her age, gender, and home and community environment, would possibly undertake if she were not restricted from participation because of a health condition.

2. The study assessed this occupational variable in a manner that is valid and reliable. An assessment is valid if it reflects the participant’s actual behavior, thoughts, or feelings in nonstudy contexts. In addition, the most valid assessments reflect the complex, multifaceted nature of occupation. An assessment is reliable if it does not fluctuate in an erratic, unexplained manner across different periods or across different persons who are making the assessment.

3. The study participants included members of the population identified in the clinical question. Among the study participants there should be persons who are similar enough to Mrs. Jones so that the therapist feels somewhat confident that the study findings can be generalized to her. In other words, the participants were likely to respond in a way that would be similar to Mrs. Jones had she participated in the study. Ideally, all study participants should be similar to Mrs. Jones on key attributes that may relate to activity participation (e.g., age, socioeconomic status, physical condition), not including depression. If the study includes participants with depression, those persons should have a level of depression that is relatively similar to Mrs. Jones’s suspected level of depression.

4. The study was designed in such a manner as to answer the entire clinical question. A study may have been designed to answer a part of the therapist’s clinical question, for example, about activity participation of elderly persons without depression or among those with depression, but not both groups at the same time. A study that answers part but not all of the question may contribute to the body of evidence; however, the evidence will not be as strong as that from a study that involves all important comparisons.

5. The reported results show how participants within a particular group varied among themselves. Because every person is unique, there are likely to be differences among persons, even when they are similar on key attributes that may be related to the occupational variable. The strongest evidence would come from studies that either involved only participants similar to Mrs. Jones or that report activity participation results for subgroups of participants (i.e., separate results for different levels of depression, for men and women, for different ages). The therapist can then look at the results that were most relevant to Mrs. Jones. Even if results are not presented according to subgroups, usually standard deviations or ranges of responses are given. These measures can be used to judge the likelihood of Mrs. Jones’s activity participation falling within either the depressed or nondepressed group.

After reviewing and evaluating single studies, the therapist evaluates the body of evidence as a whole, giving extra weight to the single studies with the strongest evidence. The “best” evidence is the best that can be found, not in the sense of meeting all of the standards. The best evidence may give us a high degree of confidence about possible answers to the clinical question, a moderate degree of confidence, or a low degree of confidence.

For our imaginary scenario, let us assume that the research evidence, in combination with other sources of evidence (e.g., interviews with the client and family members), suggests that Mrs. Jones could indeed be depressed. The compiled evidence can now be communicated to the client, family members, and colleagues to determine whether a formal evaluation for depression should be recommended.

Standards for Evaluating Assessment Studies

Now that the therapist has informed him-
The study investigated an assessment question. What are the most reliable and valid methods for assessing occupation and occupational performance among elderly women with depression who live in the community?

A set of standards to guide the evaluation of a single study retrieved to answer an assessment question is as follows:

1. The study investigated an assessment method that was designed to be used for clinical purposes. Many instrument validation and reliability studies reported in the literature have investigated instruments that were designed for research and not for clinical purposes. If the measure was designed for research, its purpose might be to compare large groups of individuals to one another, not to draw conclusions about a single individual.

2. The study investigated a method designed to assess the specific occupation, occupational performance, or component of occupational performance attribute outlined in the clinical question.

3. The study adequately tested the validity or reliability of an assessment method. There are different procedures used for testing various forms of reliability and validity. The correct procedure depends on the purpose of the assessment method. For example, an assessment that will be used for documenting a client’s progress or decline over time should be tested for its evaluative validity, and at least, its test–retest reliability. An observational assessment used to decide whether a client’s behavior falls within or outside the normal range of behavior should be tested for its discriminative validity, and at least, its interrater reliability.

4. The study participants included members of the population identified in the clinical question. An assessment that was tested with participants who are unlike Mrs. Jones may not have adequate validity or reliability when used with her.

5. The reported results show how participants within a particular group varied among themselves. This is related to the previous standard. The therapist should examine results for subgroups of individuals to determine whether a given assessment method is equally valid and reliable for different kinds of persons. For example, is the method of equal validity for men and women? This information is important for helping the therapist select which assessment methods are the most appropriate to use with Mrs. Jones.

As with descriptive evidence, after reviewing and evaluating single assessment studies, the therapist evaluates the body of assessment evidence as a whole, giving extra weight to the single studies with the strongest evidence. Does the body of evidence suggest possible answers to the assessment question? If so, what degree of confidence does the body of evidence give us that the answers are correct?

Standards for Evaluating Intervention Studies

After assessment methods are chosen and completed, the therapist may want to seek information about effective interventions for increasing Mrs. Jones’s participation in daily life activities. The therapist might write the following intervention clinical question: What are the most effective intervention methods for increasing participation in satisfying daily life activities among elderly women with depression who live in the community?

A set of standards to guide the evaluation of single studies retrieved to answer an intervention question is listed here. All of these standards, except for the fourth, are similar to the ones discussed previously. Refer to those discussions as needed.

1. The study investigated an outcome variable that is relevant to the specific occupational outcome variable in the clinical question.

2. The study assessed this outcome variable in a manner that is valid and reliable. For intervention studies, which usually are documenting change (increases or decreases) in an outcome variable, it is important that the outcome variable was measured with an assessment that had demonstrated evaluative validity and test–retest reliability. If the outcome was measured by observation, interrater reliability should also be demonstrated.

3. The study participants included members of the population identified in the clinical question.

4. The study was designed to rule out nonintervention explanations of the study’s outcomes. The strongest evidence comes from studies that controlled all factors that might affect the measured outcomes. This control involves randomly assigning participants to either an intervention condition or a control intervention, making sure that the control condition is similar to the intervention condition in all respects except for the important intervention elements (e.g., making sure that participants in both conditions receive equal and warm attention from researchers and clinicians) and making sure that the research assistants, clinicians, and participants in both conditions have
equally high expectations for positive outcomes (often accomplished by keeping them "blind" to hypotheses and administering measures and intervention in an unbiased manner). Many other factors can be controlled as well. The studies with the strongest evidence usually involve large numbers of participants (more than 60), but equally strong evidence can come from smaller studies in which participants received both conditions in randomly assigned differing sequences. The purpose of control, whether it be in large or small studies, is to assure that any differences found between the outcomes of the two conditions are due to having received versus not received the intervention (Cook & Campbell, 1979).

5. The reported results show how the outcomes of participants within a particular group varied among themselves. The therapist can find valuable information from studies that broke down the results into the subgroups. For example, the results may show that women responded less favorably to the intervention than did men.

As the therapist reviews the entire body of intervention evidence found in response to the written clinical question, he or she may find that a particular intervention has a strong positive body of evidence for its effectiveness, yet perhaps it is too costly for the client, or the resources are not available for offering such an intervention in the client’s locality. Another intervention may have a weaker body of evidence, perhaps with positive findings from studies that exercised limited control over nonintervention explanations for the outcomes, yet the intervention may be suitable for implementing with the client. What is important is that the therapist, through a search and review of the research literature, has become highly informed about the quality of the research evidence for the effectiveness of various interventions.

Conclusion
In the case of Mrs. Jones, the therapist has become highly informed, through a search for the best available descriptive, assessment, and intervention evidence, about (a) activity participation in elderly women, (b) available and relevant assessments of occupational performance, and (c) possible interventions for increasing Mrs. Jones’s participation in satisfying daily life activities. In the spirit of a collaborative model of service, the therapist must pass along this information to Mrs. Jones and her family members, enabling all decision makers to be well-informed. To communicate effectively, the therapist must give a coherent, simple summary either of his or her own synthesis of the research literature or, if available, from practice guidelines that are based on a systematic synthesis (such as a meta-analysis) of the research literature. Fortunately, the therapist can communicate coherently even when the research literature reports conflicting findings. The means for achieving this coherence is the subject of a future installment of this forum.

References

