Evidence-Based Practice Forum

Heterarchy and Hierarchy: A Critical Appraisal of the "Levels of Evidence" as a Tool for Clinical Decision Making

Linda Tickle-Degnen,
Gary Bedell

One day the two authors of this paper were having a "quick conversation" that started with one of us (Linda) asking for advice about a research study that fell within the expertise of the other (Gary). As is often the case with quick conversations, the topic transformed and soon we were engaged in a more lengthy and gradually intensifying dialogue about a different topic: in this case, the role of levels of evidence in clinical reasoning. The term levels of evidence refers to a ranking scale that is used in evidence-based practice as a tool for determining the quality of information coming from research evidence (e.g., Holm, 2000; Law & Philp, 2002; Sackett, Strauss, Richardson, Rosenberg, & Haynes, 2000). We discovered that we both had worried about the inflexible ranking of research studies according to the standard levels of evidence model, and had worried even more about the possible effects of this model on clinical decision making in rehabilitation if the model were put into practice in an exclusive manner. As we sorted and clarified our thoughts, we came to the conclusion that the basis of these worries was that in actuality we as practitioners do not think in an inflexible or exclusionary manner about any source of information, nor should we given the complexity of human responses, the realities of practice resources, and the wide variety and quality of different forms of information available to us. If practitioners were to use exclusively the standard levels of evidence model in their selection and use of information, they would be making suboptimal decisions based upon less information than that which was relevant, valid, and available for making the decisions. This creates a paradox in that the very model that is intended to support the use of best evidence in decision making, in fact fails to support the formation of a best decision, and may even hinder it, if applied rigidly and exclusively.

By no means are we the only ones to worry about the problems of a model that restricts the use of information from a vast pool of important and available information. Writers across health practice disciplines have had similar worries. Sackett et al. (2000), in their recent revision of a foundational text of evidence-based medicine responded to some of these worries by revising their original description of evidence-based medicine, which had focused on using information from research studies as the primary source of information. The revision gave evidence-based medicine a more encompassing focus that included "the integration of best research evidence with clinical expertise and patient values." Under this new description, all three information elements have equal status in clinical decision making. Although this revision begins to address some of the concerns, it does not address specifically the problem of the rigid and exclusive use of the standard levels of evidence model for selecting and using evidence from research studies. In this respect, Law and Philp (2002) have voiced a common concern of occupational therapists that the standard levels help to rank evidence from quantitative research designs yet do not help to rank evidence from qualitative designs. This concern is an important one and applies not only to the failure of the levels with respect to qualitative designs, but also with respect to many quantitative designs.

The purpose of this paper is to argue for a method of including all relevant, valid, and available research evidence for making clinical decisions. We limit our discussion to research evidence because the levels of evidence model typically applies to research rather than other forms of information. Furthermore, we discuss only research evidence related to the goal of choosing appropriate therapeutic interventions for clients because the levels of evidence model typically used in occupational therapy is one that evaluates evidence with this goal in mind. However, we recognize that our discussion is relevant for understanding how all forms of information, not merely research evidence, come together into several types of clinical decisions, not merely ones related to choosing interventions.

It is our position that the standard levels of evidence model is not inherently wrong for certain purposes in evidence-based practice. Rather it is wrong when applied in a manner that excludes the use of all relevant, valid, and available research evidence for clinical decision making. This paper is offered with the intent to address a recent challenge from the Editor's Desk in The American Journal of Occupational Therapy (Ottenbacher, Tickle-Degnen, & Hasselkus, 2002). The editorial asked the following questions:

What is the best evidence for occupational therapy? Can best evidence
come not only from large randomized controlled trials, but also from other types of study designs, such as single-subject, quasi-experimental, correlational, narrative analysis, ethnographic, and phenomenological designs? (pp. 248–249)

We hope to reframe the argument about best evidence in occupational therapy and to invite the readership to join the discussion. We begin with a critical appraisal of the standard levels of evidence model used in rehabilitation and end with suggestions for an approach that moves away from overzealous exclusion of research studies and toward a more inclusionary approach. We favor a flexible, multifaceted approach to assessing, selecting, and using research evidence. Information goals in clinical practice are hierarchical in organization rather than hierarchical (see Austin & Vancouver, 1996, for a relevant review). There are multiple information goals that are better treated as forming a network. As opposed to the standard levels of evidence model in which a single overriding goal guides the retrieval and use of information, in a hierarchical approach there is a network of goals which function in parallel.

A Description of the Standard Levels of Evidence Model

The standard levels of evidence model is a ranking of research studies along two dimensions that are integrated into one ranking scale (e.g., Holm, 2000; Law & Philip, 2002; Sackett et al., 2000). One dimension is the internal validity of the research design for answering questions of causation (Cook & Campbell, 1979). For example, studies that are well designed randomized controlled trials (RCT) or meta-analytic syntheses of these types of trials have high internal validity because they have controlled for nonintervention factors that may affect the results of the study. The well-designed RCT provides a strong demonstration that research participants' outcomes are caused by having received or not received a tested intervention, not caused by other factors, such as biased assignment of participants to intervention and comparison groups. Studies that have some control over nonintervention factors in their designs, such as quasi-experiments or case-control studies, have intermediate levels of internal validity. Studies that have low internal validity are ones that have designs, such as case reports, in which there were no controls for factors other than the intervention that may affect the results. These studies provide less definitive, evidence for the intervention causing changes in the research participants, although, as we argue in a later section, they may provide valid evidence of another form.

The second dimension integrated into the ranking scale of the standard levels of evidence model is statistical conclusion validity (Cook & Campbell, 1979). Statistical conclusion validity has to do with confidence about probabilities of effectiveness in a population. Research studies that can show with high confidence the degree to which the intervention is likely to be effective in the larger clinical population have high statistical conclusion validity. These studies, which fall at the top of the ranking scale, have large sample sizes and reliable, valid measures. Studies that cannot provide a confident estimation of the degree to which the intervention is likely to be effective in the larger clinical population, because of small sample size or poor reliability of measurement or inappropriate data analysis, have low statistical conclusion validity and fall at the bottom of the ranking scale.

The standard levels of evidence model contains a single ranking system that integrates these two dimensions of study validity. Designs that have both high internal and statistical conclusion validity, such as RCTs with large sample sizes, receive the top ranking. Intermediate ranks are assigned to studies that have high validity on one dimension but compromised validity on another dimension (with a lowered internal validity driving the ranking down more heavily than a lowered statistical conclusion validity). And studies that have low validity on both dimensions receive the lowest ranking.

A Limitation of the Standard Levels of Evidence Model

The standard model is useful for selecting and using high quality research evidence for the clinical information goal of learning about probabilities of an intervention's causal effect in the clinical population as a whole. This type of information is very important in choosing and communicating about intervention evidence. If the only available studies are ones that fall lower in the standard ranking system, the practitioner must be cautious about her or his claims to others about intervention causing beneficial outcomes and the probability of that effect. If relevant top-ranking studies are available, then the practitioner can be more confident in making these claims to clients, managers, and funding agencies.

But there is more to choosing and communicating about an intervention than simply understanding its probability of causing a beneficial outcome in a population. Practitioners are concerned not only with causality and probability, but also with patterns and possibility when it comes to selecting an intervention. Quantitative and qualitative studies that do not provide information about causality or about the probability of effects occurring in the population as a whole are excluded or devalued by the standard levels of evidence model. However, these studies may provide relevant and valid intervention-related information about (a) patterns, in the form of associations and profiles, of client attributes, perceptions, occupations, and contexts, and (b) the range of possible intervention strategies and outcomes. Pattern and possibility information may come from quantitative descriptive, observational, and correlational designs, and qualitative narrative analysis, ethnographic, and phenomenological designs. The seeking of causality, probability, pattern, and possibility information are equal status information goals with respect to choosing an intervention. In addition, a single research study can have a higher quality status for providing valid evidence about one type of information (e.g., causality) while having a lower quality status for providing valid evidence about another type of information (e.g., pattern). The organization of research information into a single hierarchy of levels may provide a solution to the "best" evidence for answering a certain information need (e.g., causality information), yet will not provide a solution to the "best" intervention choice. A better model for organiz-
ing information is a heterarchy in which it is understood that goals are organized in a multidimensional web of linkages.

An Illustration

How might the search for, evaluation of, and use of research evidence be organized in a heterarchical model? In client-centered practice, this organization begins with the client’s needs and goals. The occupational therapy practitioner, who is the primary retriever and conduit of information, works with the client to (1) identify the research information needs of the client and the practitioner, (2) translate these needs into a search for relevant types of research evidence, (3) evaluate the quality of these different types of retrieved research evidence in relation to the specific information needs of the client and practitioner, (4) synthesize the research evidence with patient preferences and practitioner expertise, (5) make a decision about what to do, and (7) try out the intervention and revise accordingly.

We offer as illustration a hypothetical case in which a young man named John wants to get and retain a satisfying job. John is diagnosed with HIV/AIDS and has difficulty managing his symptoms, medications, and daily life routines. The occupational therapy practitioner works with John and identifies a set of research information needs and then translates them into clinical questions that guide the research retrieval process. John and the practitioner have identified that they need answers to the questions listed below. The questions are categorized broadly, and not necessarily mutually exclusively, into ones of possibility, pattern, causality, and probability:

1. What is possible? What is the range of interventions that have been used with people similar to John who would like to get and retain a job? What is the range of outcomes that people similar to John have experienced while participating in these interventions?

2. Is there a pattern? Are there patterns of satisfaction and dissatisfaction that people similar to John feel with certain intervention strategies, intensities, or contexts? Are there patterns of daily life routines in people similar to John that predict a successful versus unsuccessful response to intervention?

3. Is there a causality? Does a self-management intervention strategy (Gifford, Laurent, Gonzales, Chesney, & Lorig, 1998) lead to improvements in management of symptoms, medication, and daily life routines in people similar to John? Do improvements in management of symptoms, medication, and daily life routines lead to positive job outcomes in people similar to John?

4. What is the probability? How common is it for people similar to John to be able to learn to manage their symptoms, medications, and daily life routines? Is it likely that people similar to John will have positive job outcomes following a self-management intervention?

5. What is the information retrieval process? Can we learn from research studies that people similar to John feel with certain interventions? Do improvements in management of symptoms, medication, and daily life routines lead to positive job outcomes following a self-management intervention?

For answering question sets 1 and 2 above, both qualitative and quantitative studies can provide “best” evidence and must be judged according to the degree to which they provide meaningful and deep description of possibilities and patterns, as opposed to the degree to which they provide controls related to causal inference purposes. For answering question sets 3 and 4, quantitative studies are likely to involve designs and procedures that are relevant, however, it is not necessarily only the randomized control trial that provides the best evidence. For example, in set 3, single subject designs can offer clarity in terms of tracking small increments in change over time. In set 4, large population surveys with random selection of participants may be the best evidence.

Once research studies are retrieved and evaluated for their informational value, the practitioner works with John in an iterative, flexible manner to synthesize the research findings, then take this synthesis and integrate it with information related to John’s preferences and the practitioner’s prior experience and expertise into a “big picture” of the intervention options for John. From here, they make tentative decisions about intervention, try them out, and revise as needed.

Future Directions

Our illustration of John is underdeveloped. Suppose that John is African-American, lives in New York City, has a prior history of substance abuse, is easily fatigued, experiences periodic bouts of depression, and has a supportive extended family. The use of research evidence becomes more challenging the more details we know about John. There will be little if any evidence about people who were exactly like John and living in John’s circumstances. Practitioners must know how to take research evidence that may not be closely applicable to the specific client or to the specific context of intervention, integrate it, communicate about it, and then decide with the client what they will do to meet the client’s needs. In this paper, we have not developed models for evaluating possibility and pattern information. Furthermore, we have not addressed how to evaluate and weigh evidence from a study as it applies to the needs of a particular client (applicability) or to the resources of the intervention setting (feasibility), though others have provided preliminary guidelines (Sackett et al., 2000; Tickle-Degnen, 2002). Nor have we addressed how to weigh information from research findings, client preferences, and practitioner expertise as decisions are made.

The hypothetical case of John illustrates the gaps we have in our current models for evaluating and using research evidence. From a heterarchical perspective, we must have multiple and linked methods for evaluating the informational value of different forms of research evidence. In their clinical reasoning, practitioners synthesize complex information in a flexible manner. Evidence-based practice must build models, based upon this natural process, that help practitioners to evaluate and synthesize research evidence from multiple studies then integrate it with information coming from other sources in a manner that is useful and practical for clinical decision making. We hope that others enter this discussion to express opinion and insight either in the venue of the Evidence-Based Practice Forum or through letters to the editor.

References


