How to Appraise Research: Elements of Sound Applied Design

The primary research priority of the profession is to generate support for occupational therapy interventions (Baum, 2006; Corcoran, 2007; Holm, 2000; Kielhofner, Hammel, Finlayson, Helfich, & Taylor, 2004). This requires the implementation of efficacy studies addressing effectiveness, cost and time efficiency, safety, and patient satisfaction. These are the primary applied research questions that should be considered when examining clinical practice. In this editorial, I review the elemental characteristics of solid applied research designs. When appraising research studies, the presence and quality of the factors described here should be considered. Separate standards exist to appraise the rigor of qualitative designs, but they are not discussed in this editorial.

Basic, Translational, or Applied Research?

A key factor to consider when appraising research involves identification of the research category. Understanding which category research falls into yields information about the study’s general purpose. Basic research is the examination of a clinical phenomenon to better understand its characteristics (e.g., examining how cortical tissue deteriorates in the progressive course of schizophrenia; Zerhouni, 2005). It does not offer direction for clinical intervention. Translational research is the process through which basic research can be used to develop treatment for disease and injury (e.g., developing a pharmaceutical intervention to slow or halt cortical deterioration in schizophrenia; Horig, Marincola, & Marincola, 2005). Much of translational research in medicine is commonly initiated using animal models (Nathan, 2002). Applied research involves the assessment of clinical intervention to determine efficacy (e.g., testing the effectiveness of a specific drug’s ability to preserve cognitive function in adults with schizophrenia; Zerhouni, 2005).

What Is the Level of Evidence?

A second factor that should be considered when appraising research involves identification of the level of evidence generated. Evidence supporting practice is commonly categorized into five levels. Levels of evidence refer to the rigor of the experimental design and the degree to which one can be certain that the results are valid and generalizable to the larger population. In accordance with the American Occupational Therapy Association’s Evidence-Based Literature Review Project (Lieberman & Scheer, 2002), Level I evidence consists of randomized controlled trials, systematic reviews, and meta-analyses. Level II evidence involves study designs of two groups that are controlled but not randomized (e.g., two-group pretest–posttest designs, cohort studies, case control studies). Level III evidence consists of designs in which one group is studied; randomization and control are not present (e.g., one-group pretest–posttest design). Level IV evidence involves descriptive studies, single-subject designs, and case series that involve analysis of outcomes. Level V evidence is not based on systematic research methodologies.
but instead reports expert opinion or case report findings. The rigor of experimental design decreases as level ratings increase.

Note that levels of evidence refer only to the rigor of experimental design and to the degree to which results can be considered credible and generalizable. Levels of evidence do not indicate the potential importance of findings. For example, case series and case reports are considered among the lowest levels of evidence. The purpose of such designs is to report novel but promising intervention techniques and to alert the professional community to possible adverse treatment reactions. A critical example of the use of such designs is the case series published by McBride (1961) regarding the absence of limbs (phocomelia) in two newborns whose mothers had taken thalidomide in early pregnancy. This report of thalidomide-induced phocomelia was the first published report alerting the medical community to this clinical problem.

What Is the Purpose of the Study?

The study’s purpose addresses the relevancy of the research question. In applied designs, can the research study directly answer questions relating to the efficacy of specific clinical practices? In other words, do the study’s findings allow practitioners to better understand whether specific interventions are effective and tolerable by patients? Do findings allow practitioners to go into the clinical setting tomorrow and better understand how to select best-practice methods or modify intervention strategies to enhance patient treatment outcomes? In the current health care environment in which all professions are competing for health care dollars and are challenged by insurer denial of services that are unsubstantiated by research, answering these clinical questions has become the most important research priority.

Who Are the Participants?

It is important for readers to understand precisely who the participants of a given study are and are not. To reduce confounding factors that may skew results, researchers frequently use inclusion and exclusion criteria to ensure the recruitment of a homogeneous sample. For example, studies commonly exclude patients with coexisting illnesses, those who are not fluent in English, and those taking specific medications. Although these exclusion criteria reduce potentially confounding factors, they also reduce the likelihood that study participants will be similar to patients actually seen in real clinical settings. Similarly, a study may have restricted inclusion to participants experiencing a specific stage in the course of an illness or injury. For example, if an intervention for patients with brain injury is found to be effective with patients functioning at a Rancho Cognitive Level (Hagen, 1998) of 4 to 6, it is unclear whether that intervention would be effective or even appropriate for patients functioning at a Rancho Level of 7 to 9. Study findings can be generalized only to patients meeting the same inclusion and exclusion criteria of a given study.

Methodological Quality

The study’s methodological quality determines the strength of the evidence generated. All studies have methodological limitations that should be clearly stated by the authors. Such methodological limitations reduce the study’s internal validity (i.e., the degree to which one can be certain that findings are a result of treatment and not chance occurrences) and external validity (i.e., the degree to which results can be generalized to the larger population). The following factors should be considered when appraising the methodological strength of specific studies.

Sample Size and Power Analysis

Sample size refers to the number of participants in a study. That number should be large enough to avoid a Type II error (i.e., failing to observe a statistically significant difference between groups when one exists). To avoid a Type II error, researchers can use power analysis—a procedure in which a statistical nomogram, or graph, is used before study implementation to calculate the sample size needed to achieve a moderate to high chance of detecting an actual difference between groups (Warner, 2008).

Sampling Method, Randomization, and Control

Sampling method refers to the way in which participants are recruited to ensure that a homogeneous group that truly represents the larger population is selected. When appraising research design, it is important to determine whether the sampling method was unbiased and best ensured representation of the larger population.

A critical determinant of the research design’s strength is the use of randomization and control. Random selection means that every person in the larger population had an equal chance of being selected for participation. In other words, a sample of participants from a larger population was randomly selected and then randomly assigned to an experimental or a control group. Random selection is difficult to achieve; researchers must be able to access members of a specific population and recruit a representative sample. Often, random selection requires financial resources and access to multiple patient bases and treatment centers.

More commonly, researchers use convenience samples to which they have ready access and then randomly assign participants to groups. Convenience samples that are subsequently randomly assigned to experimental and control groups cannot be determined to be representative of the larger population. Results cannot be generalized beyond the study.

Control refers to the ability to stabilize or eliminate variables that may interfere with and confound results. Control can be achieved through the use of control groups and participant matching. The use of control increases the likelihood that findings are a result of treatment and not extraneous variables. When control is not present, findings cannot be determined to result from treatment effects.


In double-blinded designs, both researchers and participants do not know which participants are in the experimental or the control group. In single-blinded designs, researchers know which participants have
been assigned to each group; participants do not. When researchers are not blinded and assess patient outcomes, results may be biased.

Results are also questionable when more than one rater is used and interrater reliability is found to be below moderate levels or has not been established. Similarly, one should determine whether the instruments used to measure outcomes are reliable and valid. Results based on instruments with low or untested levels of reliability and validity are subject to dispute. One should also consider whether instruments are ecolgoically valid or based on actual functional activities that patients encounter in daily life. Instruments that purport to assess functional performance through methods other than real-life activities—such as paper-and-pencil tests—may not accurately measure functional performance.

**Use of Follow-Up Points**

Determining the effectiveness of an intervention is important. Determining whether intervention effects last beyond the post-intervention period is critical to understanding whether occupational therapy benefits patients over time. Such research questions require designs using follow-up points (e.g., time series with multiple measurements). The benefit of some interventions may only become apparent over time. For example, the benefits of sensory integration therapy may only become evident when examining multiple life variables over a period of months to years.

**Effect Size**

Although it is important to determine whether a statistically significant difference exists between two or more groups (i.e., the experimental and control groups), it is equally important to assess the size of existing differences. **Effect size** is a measure that expresses how large a difference exists between groups (Warner, 2008). Effect size allows researchers to measure how well an intervention worked rather than only whether it worked. Statistical methods (e.g., Pearson r, Cohen’s d, odds ratio) that measure effect size should be part of the data analysis procedures for effectiveness studies.

**Treatment of Missing Data**

Because participants who drop out of a study are often more likely to have experienced treatment side effects, skipped treatment, or lost motivation to participate, the remaining data set of participants who completed the study may be biased. As a result, when appraising effectiveness studies, it is important to determine whether researchers accounted for participant attrition and addressed missing data points through imputation of data.

**Summary**

These factors are important to consider when appraising the strength of research evidence. Although all studies have limitations, applied designs that address many of the factors described here can provide stronger evidence of the effectiveness of occupational therapy intervention. This appraisal process is critical not only to build the profession’s body of evidence but also to ensure that our evidence is considered scientifically sound and credible by insurers, fellow health care professionals, and consumers. ▲

**References**


### ETHICS COMMISSION

**PUBLIC DISCIPLINARY ACTIONS—December 2008**

The Ethics Commission (EC) has taken the following recent disciplinary actions. According to Section 1.3 of the *Enforcement Procedures for the Occupational Therapy Code of Ethics*, with the exception of those cases involving only reprimand, the American Occupational Therapy Association (AOTA) “will report the conclusions and sanctions in its official publications and will also communicate to any appropriate persons or entities.”

<table>
<thead>
<tr>
<th>Name</th>
<th>Sanction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dana Gray, OT</td>
<td>Censure—effective December 5, 2008 Principles 4A, 5A, and 6A</td>
</tr>
<tr>
<td>Ernesto Gabriel Martinez</td>
<td>Censure—effective November 7, 2008 Principles 6C, 4B, and 5E Guidelines 1.5, 2.3, 9.1</td>
</tr>
</tbody>
</table>

Please contact Deborah Slater, AOTA liaison to the EC, atdslater@aota.org if there are questions concerning this information.