Why Rehabilitation Research Does Not Work (As Well as We Think It Should)

Kenneth J. Ottenbacher, PhD, OTR

ABSTRACT. Ottenbacher KJ. Why rehabilitation research does not work (as well as we think it should). Arch Phys Med Rehabil 1995;76:123-9.

• Establishing treatment effectiveness is a high priority for rehabilitation research. The use of traditional quantitative null hypotheses to achieve this priority is reviewed. Three problems are identified in the analysis and interpretation of investigations based on statistical testing of hypotheses: (1) confusion of clinical and statistical significance, (2) low statistical power in detecting clinically important results, and (3) a failure to understand the importance of replication in developing a knowledge base for rehabilitation practice. Technical aspects associated with each problem are reviewed and examples presented illustrating the impact of low statistical power and the results of misinterpreting statistical significance tests. Several specific recommendations are made to improve the clinical usefulness of quantitative research conducted in rehabilitation.

In describing the research process, Deitz[1] identified two major approaches: quantitative and qualitative.[1] She notes that the advantages of well controlled quantitative research include the ability "to obtain precise, objective information about subjects" and "to control factors that could jeopardize either the internal or external validity of a study."[1] Deitz describes several quantitative research designs, including the category of designs originally presented by Campbell and Stanley[2] as "true-experimental." The purpose of true-experimental designs is to establish a cause-and-effect relationship between the independent variable (usually treatment) and dependent variable (outcome measure). True-experimental designs are widely regarded as the most powerful research method to establish causality.[3] In describing research approaches used in medicine, Kramer and Shapiro[4] state that the true-experimental design, also referred to as a randomized clinical trial (RCT), "is generally regarded as the most potent scientific tool for evaluating medical treatments."[4] Along similar lines, Reilly and Findley[5] observed that in rehabilitation research investigations, "true random experiments are the best way to establish causation."

Poole and colleagues[6] conducted a quantitative clinical study with many of the ideal attributes associated with true-experimental designs. They examined the effectiveness of inflatable pressure splints on upper extremity function in subjects with hemiplegia. Matched pairs of subjects were randomly assigned to splint and non-splint conditions. Following 3 weeks of intervention, the upper extremity function of both groups was examined. The Fugl-Meyer Assessment (FMA)[7] was administered to all participants by raters unaware of the group to which each subject was assigned. This study is an excellent example of experimental clinical research and includes the following ideal characteristics: the independent and dependent variable were operationally defined; subjects were randomly assigned to treatment and control conditions; and, the outcome measure was blindly recorded. The attributes of this design are presented schematically in the figure. The design used by Poole et al controls the major threats to internal validity described by Campbell and Stanley.[2] These include threats of history, maturation, testing, instrumentation, statistical regression, selection bias, experimental mortality, and interaction effects.

The internal validity of an investigation is an important factor in determining the quality of an experimental study.[3] The value of conducting clinical trials with high internal validity is widely acknowledged in the rehabilitation literature.[1,4,5]

The randomized clinical trial reported by Poole and colleagues[6] includes all the attributes of a good experimental design. From a clinical viewpoint, there is only one problem with the Poole et al, investigation. The results showed a statistically nonsignificant differences between the treatment and control groups for the upper extremity outcome measures. When a randomized clinical trial produces a statistically nonsignificant result the proper interpretation is that the null hypothesis is true. The null hypothesis states that there is no difference between the groups or conditions. It is a mistake, however, to equate statistical significance with practical or clinical significance. Statistical results, from even the best designed experimental trials, can overestimate or underestimate clinical importance. Bakan[8] noted that operationalizing clinical significance by equating it with statistical significance "removes the burden of responsibility, the chance of being wrong, the necessity of making an inductive inference, from the shoulders of the investigator and places it on tests of significance." More recently, Shaver[9] has argued that "statistical significance does not speak to the probability that the null hypothesis or an alternative hypothesis

From the State University of New York at Buffalo, NY.

Submitted for publication March 30, 1994. Accepted in revised form July 18, 1994.

Supported in part by a grant from the NIDRR during the completion of this research (Grant No. H133B30041).

No commercial party having a direct financial interest in the results of the research supporting this article has or will confer a benefit upon the authors or upon any organizations with which the authors are associated.

Reprint requests to Kenneth J. Ottenbacher, PhD, SUNY at Buffalo, Department of Rehabilitation Medicine, Center for Functional Assessment Research, 232 Parker Hall, 3435 Main Street, Buffalo, NY 14214.

0003-9993/95/7602-3032$3.00/0