

# A Methodological Review of Randomized Clinical Trials

M. N. Hegde

*California State University—Fresno*

---

This paper is a critical examination of the claim that randomized clinical trials (RCTs) offer the “gold standard” of treatment research evidence. Often repeated in medical literature, this claim is now frequently echoed in speech-language pathology. This critical review seeks to underscore a need for speech-language pathologists to seriously consider the methodological limitations of RCTs before accepting them as the only or even a major means of establishing treatment efficacy. This review suggests that in the age of informed consent and the rights of the participants to withdraw from RCTs and to demand experimental treatment, randomization is an impossible technical goal. Consequently, all participants in RCTs are self-selected, not randomly selected or assigned to different treatment or no-treatment conditions. For various technical, financial, and practical reasons, communication disorders cannot afford RCTs even if they are technically sound and feasible. Direct and systematic replication of well controlled studies using either single-subject experimental designs or small-group experimental designs will better serve the profession than large-scale RCTs that fail to use randomization and tend to produce variable and ambiguous data that contradict individual uniqueness.

---

## A Critical Review of Randomized Clinical Trials

---

That the consumer should receive clinical services whose positive effects have been experimentally demonstrated is a basic clinical tenet. Nonetheless, in speech-language pathology, many treatment procedures whose effects are not experimentally established generally are offered to clients. This is a growing concern for both scientists and professionals. This concern has received scholarly attention in speech-language pathology and has generated articles that review or summarize treatment efficacy evidence to guide the practitioners

(e.g., Ansel, 1993; Conture, 1996; Curlee & Yairi, 1997, 1998; Gierut, 1998; Holland, Fromm, DeRuyter, & Stein, 1996; Jones, Gebiski, Onslow, & Packman, 2001; Moscicki, 1993; Ramig & Verdolini, 1998; Thomas & Howell, 2001; Yorkston, 1996). On its Web site, the American Speech-Language-Hearing Association maintains brief summaries of treatment efficacy evidence for various disorders of communication (for additional information, visit <http://www.asha.org/members/epb>).

Generally, scholars have argued that randomized control trials are the best means to establish treatment effects. Although many professional organizations and individual researchers advocate somewhat varied hierarchy of treatment research evidence, most of these hi-

erarchies place evidence generated by RCTs at the top. Some of these hierarchies do not recognize evidence generated by single-subject experimental research, even if the evidence is replicated multiple times. Those that do, downgrade single-subject experimental research evidence to second or third class; in some levels of evidence, it is often not clear where it fits. A technical report of the American Speech-Language-Hearing Association (2004), for example, describes the following four levels of evidence:

- Ia. Well-designed meta analysis of >1 randomized controlled trial
- Ib. Well-designed randomized controlled study
- Ila. Well-designed controlled study without randomization
- Ilb. Well-designed quasiexperimental study
- III. Well-designed nonexperimental studies
- IV. Expert committee report, consensus, and so forth.

One may presume that evidence generated by single-subject experiments fall under “Well-designed controlled study without randomization.” In this case, the single-subject experimental evidence is relegated to the third level (even if it is Ila). The assumption that only randomized controlled trials (RCTs) are capable of producing the best possible evidence of treatment effects may be a hurdle for an expanding treatment efficacy database. This assumption implies that the method is the end in itself. Methods used to generate data are important; however, validity, reliability, and generality of data are even more important because they are the end products of treatment research. All methods that yield valid, reliable, and generalizable data should be acceptable to researchers and clinicians because a discipline cannot afford to reject out of hand any such methods. At the same time, clinical sciences cannot afford to ignore serious limitations of historically accepted methods.

### Speech-Language Pathologists’ Dilemma

The goal of this review is to raise serious concerns about the general validity of RCTs and their applicability to speech-language pathology. An examination of the validity of RCTs requires a review of their practice in medicine where the method is most widely used. As will be clear later, limitations of RCTs are widely discussed in medical literature; it is not clear that speech-language

pathologists who advocate RCTs as the best method of establishing treatment efficacy appreciate those limitations.

In accepting RCTs as the gold standard of treatment efficacy, speech-language pathologists will face a scientific and ethical dilemma. The clinicians will have to conclude that most of their treatment procedures have not been proven effective. Indeed, those who have restricted their reviews to RCTs have concluded that there is no evidence to support treatment of most disorders of communication. Such reviews may be found in the *Cochrane Database of Systematic Reviews* published and archived by the Cochrane Collaboration on their Web site (<http://www.cochrane.org/reviews>). Most if not all reviews published by this database consider only RCTs, and make such claims as the following:

1. There may only be “some support” for treating expressive language and phonologic difficulties; but the evidence for treatment of syntactic structures is mixed; and there is no evidence to support treatment for receptive language problems (Law, Garrett, & Nye, 2006)
2. Speech and language therapy for people with aphasia following stroke has not been shown to be effective (or ineffective); there is no difference between the groups that receive formal therapy and those that receive informal support; decision to offer treatment to patients with aphasia has to be made on other grounds, not on evidence (Greener, Enderby, & Whur, 2006)
3. Nonpharmacologic swallowing treatment (of the kind speech-language pathologists offer) for dysphagia in patients with Parkinson’s disease is not shown to be effective (Deane, Whur, Playford, Ben-Shlomo, & Clarke, 2006a)
4. There is insufficient evidence to support or refute communication treatment for dysarthria in patients with Parkinson’s disease (Deane, Whur, Playford, Ben-Shlomo, & Clarke, 2006b)
5. There is no evidence to support treatment for apraxia of speech following stroke (West, Hesketh, Vail, & Bowen, 2006)
6. There is no firm evidence to support speech and language treatment for children with cerebral palsy (Pennington, Goldbart, & Marshall, 2006)
7. There is no evidence to support the speech and language treatment for dysarthria in patients who have nonprogressive brain damage (Sellars, Hughes, & Langhorne, 2006)

Given the criterion that acceptable evidence has to come from RCTs, the conclusions of the Cochrane Database Library’s systematic reviews are most likely val-

id; very few well-designed RCTs have been conducted by speech-language pathologists. If the same criterion is adopted, it is very likely that the same conclusion, that treatment is not shown to be effective (or ineffective), would hold good for articulation and phonologic disorders, fluency disorders, voice disorders, cleft palate speech, and virtually all other disorders of communication. Are speech language-pathologists ready to accept the conclusion that treatment of communication disorders is generally not shown to be effective? Does the evidence has to come from RCTs to support their professional practice? Should the clinicians now tell their clients that almost all treatments are offered on a basis other than evidence? If so, on what basis?

### The Essence of Treatment Efficacy Research

Treatment efficacy researchers have had two main goals. First, they demonstrate that a given treatment had an effect in an experimental study and thus achieve *internal validity*. To achieve this goal, researchers relate causes (treatments or independent variables) with effects (dependent variables or positive changes in client's communicative skills, health, or other measured variables). All researchers use the *experiment* as the method to relate causes to effects. An experiment can help show that when a treatment was present, certain effects also were present, and that when the treatment was absent, the effects also were absent.

Second, the researchers replicate the experiment to demonstrate that it will produce similar results in other clients and in other settings, and thus achieve *generality*. By replicating the results of an earlier experiment with new participants, the investigator shows that the earlier results were reliable. Different investigators in different settings replicate the experiment to determine the extent to which the results may be extended.

Researchers have used at least two approaches to show that a treatment works and that it has generality: the group design approach and the single-subject approach. Although the group design approach includes a control group to demonstrate that the effects were absent when the treatment was withheld, the single-subject approach arranges treatment and no-treatment conditions through which all participants pass to demonstrate the presence and absence of effects (Hegde, 2003). Insisting only on the randomized group design approach disregards this basic logic and empirical realization of cause-effect relations.

In its truest form, participants are randomly *selected* from a large population of defined subjects who are

then randomly *assigned* to different groups of the study. There is usually at least one group that receives the new treatment being evaluated. At least one other group receives a traditional treatment, receives no treatment at all, or receives a placebo. To evaluate the relative effects of two or more treatments, multiple groups may be formed randomly, each receiving a distinctly different treatment. Randomized clinical trials are hailed as the best type of treatment research mostly because of the advocated advantages of randomization (Mathews, 2000; Meinert, 1986; Pocock, 1983).

### A Brief History of Randomization

Because randomization is the cornerstone of RCTs, the present critical review is mostly concerned with its accepted promises and ignored perils. In the history of scientific experimentation, the concept of randomization had a relatively recent and humble origin in game theories with a base motivation to help the aristocratic French gamblers understand chances of winning, perhaps mostly losing. The concept evolved into a more respectable mathematical *theory of probability* which sought to predict the occurrence of an event among a possible set of such an event (Hacking, 1975; Hald, 1998). Presumed probabilities of an event were demonstrated routinely then and are being demonstrated in innumerable Statistics 101 now, with the all too familiar coin toss. A single toss of coin can only be heads or tails: in terms of probability, there is a 50% chance that the toss will be heads and an equal chance that it will be tails, or  $p = .5$  ( $p$  standing for probability). Thus evolved a formula to predict the occurrence of events in terms of their probabilities.

During the mid 17th and early 18th century, the French mathematician and astronomer Laplace further developed the probability theory into the better known *normal probability theory*. The German mathematician and astronomer Gauss postulated that when something is measured repeatedly, there is always an error (of measurement) and that these errors are distributed evenly around a mean. Quetelet, the Belgian astronomer and the father of modern statistics further advanced this concept of the mean of measurements and the symmetric errors on either side of the mean: he proposed that the mean (the average person) was the nature's aim and individual differences are the errors of nature! A blatant statement that individual differences are nature's errors is reprehensible; nonetheless, hidden in the *standard error of measurement* of group experimental research, individual differences are indeed treated the same: errors of measurement.

Soon, scientists in agriculture, psychology, and social sciences found themselves in need of a formula to predict the outcomes of their experiments. They asked research questions whose answers had to apply to a large number of entities: countless agricultural plots and millions of people; the all too important question of generality of the results of scientific experiments done on a sample of participants. The experiments they could possibly conduct could employ or recruit only a minuscule number of the possible number of entities (static agricultural plots and ever so dynamic people)—the entities to which their research answers had to apply. The theory of probability and the statistical techniques based on this theory were indeed a friend in need. None saw this more clearly or established a research tradition more firmly than the brilliant British biostatistician Ronald Fisher. Fisher's two major works, *Statistical Methods for Research Workers* (1925), and *Design of Experiments* (1942) laid the foundation for randomization as a method to achieve pre-experimental equivalence of groups, group designs of research, and statistical analysis of results. Fisher's research was greatly facilitated by the works of Galton, who believed in mental (intellectual) inheritance, regression to mediocrity (the less insolent *mean*), and the statistical method of correlation. Later, Pearson worked out the mathematical details of co-relation and called it the *coefficient of correlation*, which made it possible for Fisher to develop factorial analysis of results that randomized agricultural experiments generated.

### The Need for Randomization

Fisher was the first investigator to apply randomization to large-scale agricultural research. Faced with the problem of designing experiments to study the interaction between soil chemistry, weather conditions, moisture content, seed quality, quantity and quality of fertilizers, and a host of similar variables that affect crop growth, Fisher began to randomize plots (small sections of an acre) in which other variables (e.g., seed quality and watering schedules) were combined (Box, 1978). Without randomizing the agricultural plots, he could not equate soil chemistry across other variables (e.g., amount of water applied to the plots that contained seeds of different quality).

Although Fisher's experiments were agricultural, his methods soon spread to psychology, social sciences, and eventually, to medicine. Scientists in these and other disciplines soon saw the usefulness of randomization in their own research. An agricultural plot was like a person, a complex locus in which many variables produced their individual and interactive effects. Before forming

an experimental group and a control group to study the effects of a variable (e.g., a new form of treatment), the two groups had to be equal in all variables that may interact with treatment. The only way to have two groups of people who are equal on known and unknown variables was to randomly draw a sample from a large population so that the investigator's biases did not influence participant selection. This is the well known principle of *sampling equivalency* in group designs (Hegde, 2003). Therefore, the first function of randomization is to yield two (or more) groups of similar people.

A second function of randomization emerged because of the necessity to achieve generality while studying smaller samples drawn from the population. To achieve generality, the sample had to be *representative* of the population; the sample would be representative only if drawn from a large population and without bias. Therefore, the second function of randomization is to yield a sample that represents the population.

A sample's representativeness of its population and sampling equivalency of the two or more groups are the two functions of randomization that lead many researchers to believe that RCTs are the *gold standard* of treatment efficacy research. We will see in subsequent sections that these two functions of randomization worked well for Fisher who randomized agricultural plots but do not work at all for scientists who try to randomize human beings in treatment research.

I first review the nature and limitations of RCTs as they are implemented in medicine. I next review the relevance of RCTs to speech-language pathology and the methodological challenges they pose for speech-language pathologists.

### Randomized Clinical Trials in Medicine

Bradford Hill's 1948 report is thought to be the first prototype of an RCT, published in *The British Medical Journal* (Randal, 1998; Silverman & Altman, 1996). Hill randomly selected patients across several hospitals in England to study the effects of streptomycin on pulmonary tuberculosis. This study also was the first to blind investigators; those who read chest x-rays to evaluate the effects of treatment did not know who had the treatment and who did not. In subsequent decades, the number of such studies increased and RCTs became the established method to experimentally evaluate medical treatments. Lately, the view that RCTs set the standard for treatment efficacy is echoed in communication disorders as well (American Speech-Language-Hearing Association, 2004; Jones, Gebiski, Onslow, & Packman, 2001; Robey, 1999). As noted before, the American Speech-Language-Hearing Association (2004), consistent with vari-

ous medical organizations, puts the evidence generated by multiple RCTs at the top of an evidence hierarchy.

The critical factors to consider in evaluating the validity and applicability of RCTs is whether randomization is essential to achieve internal validity as well as generality (external validity). I begin with these two issues and then move on to methodological problems inherent to RCTs.

### Internal Validity Is a Matter of Experimental Designs, Not Randomization

Although it could help, randomization is not necessary to demonstrate the internal validity of experiments. Randomization may help rule out the influence of extraneous variables by achieving sampling equivalence; but it is not the only or the best means of showing that treatment alone, not extraneous variables, produced the effects in a treatment study. Matched groups in group designs and single-subject designs can accomplish it without randomization. Tight experimental control and ruling out extraneous variables by contrasting the behavior of a dependent variable under treatment and no-treatment conditions are all that are needed to achieve internal validity.

Historically, to demonstrate that treatment was effective, because it was better than no treatment, researchers have selected the randomized experimental group-control group design (Campbell & Stanley, 1966; Fisher, 1942, 1956). This theoretically elegant method offered a chance to show that among the two groups drawn and formed randomly, one that received a treatment changed positively and the one that did not receive it did not change positively, did not change as much, or changed negatively. For many historical reasons that cannot be explored here, ruling out extraneous variables (such as history and maturation) became a matter of having a *control group*, consisting of individuals from whom the treatment is withheld.

A control group is good, but is not the only the good thing available. To show that treatment is the cause of positive changes in people, all that an experimenter has to do is show that when the treatment was present, effects were present, and when the treatment was absent, the effects were absent (Clatterbaugh, 1999; Heise, 1975; Rothman & Greenland, 1998; Ruben, 1990; Wilson, 1985). This philosophy of science does not dictate that a *control group* is the essence of showing that in the absence of treatment, nothing worth reporting would have happened. The logical requirement of treatment research leaves it to the methodological ingenuity of researchers to create empirical conditions that reveal a cause-effect relation between such events as treatment on the one

hand and changes in symptoms or health status on the other.

The essence of treatment research is experimental *control* that helps rule out extraneous variables. Therefore, there is no logical justification for considering the *control group* as the sine qua non of *control* in an experiment (showing that the absence of a treatment means the absence of the effect). Just as a treatment variable can be introduced and withheld from *different* individuals, treatment can be introduced and withdrawn from the *same* individuals. This is what is done in single-subject experimental designs (Hegde, 2003). Showing that the same persons experience the effect when the treatment is present and lose the effect when the treatment is absent is not inferior to showing that different individuals show the presence or absence of the effect. In essence, then, control is achieved in two equally effective methods. A control group is one method. A control condition is the other method. Therefore, a control group of the RCTs is sufficient but not necessary to claim cause-effect relationships. The alternative strategies are more attractive to the researcher because of the many logical and practical problems RCTs present.

### Generality Is a Matter of Replications, Not Randomization

Campbell and Stanley (1966), among the most influential authors on group research methods, claimed that when a sample is both selected and assigned randomly, the results of a treatment study can be extended to the population. Because of the inferential statistical nature of this extension, it is also known as *inferential generality* (Barlow & Hersen, 1984; Hegde, 2003). Unfortunately, a critical examination shows that randomization does not ensure generality of results generated by RCTs.

Although either the group (including RCTs) or single-subject design approach is adequate for establishing internal validity (causality), neither is sufficient to establish generality (external validity) when a single treatment study is completed. This is because generality is not a matter of randomization; it is not even a matter of experimental designs per se; it is a matter of replications (Hegde, 2003). Mook (1998) has asserted that Campbell and Stanley (1966) were simply wrong when they claimed that designs that are strong in both internal and external validity should be selected for treatment evaluation. Mook has stated that “External validity is not an automatic desideratum; it *asks a question*” (italics in the original) (1998, p. 146). It asks whether data from a given study *can* be extended to people not included in that study. The answer is not an automatic affirmative solely because an investigator used randomly

selected groups in the study. An affirmative answer is unthinkable when nonrandomly selected samples are only assigned randomly to experimental and control groups. The answer is unknown when data stem from a single study. In all cases, an answer has to be found, and only by replication.

An effectively used group experimental design (used in RCTs) or a single-subject experimental design helps ensure internal validity, but not external validity. Until other clinicians in other settings replicate the study by using different clients, external validity of the treatment procedure remains doubtful. The central notion of external validity is that a treatment's effects under similar (and perhaps more importantly) different conditions will be comparable. Regardless of the sample size, asserting external validity of findings based on a single study is perilous. Therefore, prudent researchers of group design approach to treatment evaluation would replicate before asserting clinical generality. Researchers of single-subject approach, who typically do not claim generality for unreplicated studies, accept the burden of replication from the beginning. Essentially then, the trouble of otherwise and multiply flawed randomization is not worth the efforts when the trouble of replication is inevitable.

Medicine is more successful than speech-language pathology in reporting randomized group treatment evaluations with large numbers of patients. Nonetheless, subsequent replications frequently question the conclusions of earlier studies that did randomize participants into large groups of patients. Reviews of methodological quality of several hundred randomized control trials have shown that there are significant deficiencies in most of them (Dreyfuss, 2004; Feig, 2005; Fletcher, 2002; Gøtzsche & Olsen, 2000; Horwitz, 1987; Koes, Bouter, & van der Heijden, 1995; Schulz, Chalmers, Grimes, & Altman, 1994). In some instances, only a quarter of all trials may be methodologically adequate (Bizzini, Childs, Piva, & Delitto, 2003). Screening mammography for breast cancer is a recent example of controversy regarding contradictory findings produced by a large number of RCTs. Over the decades, RCTs on screening mammography had randomized as many as half a million women in multiple countries leading to the conclusion that screening mammography saves lives. Nonetheless, authors of a reanalysis of those studies claimed that the conclusions are flawed because of inadequate randomization and other methodologic deficiencies (Gøtzsche & Olsen, 2000); the controversy still continues (Feig, 2005; Fletcher, 2002). Another example involves a randomized clinical trial in surgery that was once considered an ideal (EC/IC Bypass Study Group, 1985), but severely questioned just two years later (Relman, 1987; Sundt, 1987). The study concluded that intracranial-extracranial arterial bypass surgery is ineffective in reducing the risk of

ischemic strokes in patients with symptomatic atherosclerotic disease of the internal carotid artery. Only a small number of eligible patients were randomized into the study and a much larger number of patients received surgical treatment outside the scope of the study; hence, the patients who consented to participate in the study may not have represented those who did not, but received the same treatment outside the scope of the study with presumably better outcome. Therefore, the conclusions of the study were judged invalid. A recent editorial concluded that several RCTs in hypertension treatment succeeded only in giving confusing messages (Moser, 2006). These are but a few of the many examples (see Dreyfus, 2004) widely discussed and heavily criticized in medicine that show that large numbers of individuals randomized into experimental groups and control groups do not necessarily ensure internal validity; and without it, there is no external validity.

Moye and Deshwal (2003) point out that even though medical RCTs recruit hundreds and thousands of participants in multiple countries with enormous effort and expense, the population of patients to which the results are extended is so big that the sample, no matter how large, is miniscule. For instance, they state that a study that recruits several hundred patients with Type II diabetes may randomly assign patients to a new form of treatment or control therapy, but the population of that disease in the United States is 15 million. A sample of 300 patients (a difficult sample to achieve) is only 0.002% of the population. Even large samples of thousands of individuals may still be biased and nonrepresentative of the relevant population. For instance, RCTs on cancer have recruited thousands of participants. Nonetheless, researchers have estimated that only 1% of cancer patients may be enrolled in RCTs (Taylor, et al., 1994). Another study on multiple sclerosis has found that only 3% of patients in a particular setting were recruited for a randomized trial (Schwartz & Cox, 1995). In fact, many medical researchers believe that participants in randomized clinical trials do not represent their respective populations (Begg & Engstrom, 1986; Feinstein, 1970; Taylor, et al., 1994). As the concept of RCT has spread to nonmedical human service professions, the difficulty of implementing them in such diverse fields as learning disabilities in children to adolescents and adults with intellectual deficiencies has been a major concern for researchers (Gates & Atherton, 2001; Oliver et al., 2002). Before jumping on the RCT bandwagon, speech-language pathologists should consider questions like these: What should be a representative sample size in a treatment experiment on language disorders in children with a prevalence rate of 12 to 13% of school-age children in the United States? What financial resources are needed to conduct such RCTs in multiple settings? How do we

train multiple speech-language pathologists to administer an experimental treatment in a uniform format?

No matter how large the number of participants in a study, one needs replication to claim generality. No matter how small a study's sample, multiple replications will eventually establish generality. The tacit assumption that a group design treatment study of some 50 persons with aphasia or stuttering (of whom only 25 received treatment) can automatically establish the clinical validity of that treatment whereas a single-subject design study of some 6 or 8 subjects (all of whom received treatment) establishes little is highly questionable. This questionable assumption is based on the theoretically elegant assertion that samples represent populations and that the measured behaviors of the samples reflect the unmeasured behavior of all in the population. How often such theoretical elegance translates into empirical reality in treatment research should be a concern for all researchers, not just for the critics of RCTs.

It is much easier to replicate smaller studies of both group and single-subject designs. It is much harder to randomly (often inadequately) draw another 50 clients for a replication. Consequently, an unreplicable result of a group design study is more likely to mislead practitioners than the unreplicable result of a single-subject study because generality is not promoted by that single study and a failure to replicate the result is likely to be reported soon.

### Randomization in the Post-Informed Consent Era

The limitations of randomization (and the RCTs that use it) described so far have been there from its very inception. The problems have exacerbated in the recent decades, however. The dawn of the informed consent era has turned the concept of randomization into a thin shadow of what Fisher eminently and effectively practiced in agricultural research.

Random *selection* and random *assignment* are the two stages of randomization. In the first stage, participants should be randomly selected from a defined population to which the results of a study are expected to be extended. In a truly random selection, (1) all persons in the population with the defined characteristics are potential participants, (2) all should be accessible, and (3) all or most who have been randomly selected should participate in the study until its completion. None of these elements characterize RCTs.

Those basic requirements of random selection run counter to the legal assurances given to participants in all research studies. Before 1974, when there were no

federal laws in the United States to protect human participants in research studies, people were selected for experiments (though rarely randomly) to various kinds of research without their consent and without giving them a clear understanding of the risks and benefits. In many abusive experiments that were conducted in medicine and many other disciplines, the participants did not know they were experimental subjects (Hegde, 2003, for a review of some of the abusive experiments). The passage of the National Research Act in 1974 changed this to a great extent. The law required that the potential participants (a) be fully told about the purpose and methods of the study; (b) should give free and voluntary consent to participate only after fully understanding the risks and benefits of their participation; (c) be free to withdraw from a study in spite of their initial voluntary consent; (d) retain their right to demand alternative treatments, demand or reject the experimental treatment, refuse to be a part of the control group, or drop out without giving reasons. All researchers guarantee these conditions and participant rights; many patients freely exercise their rights. In essence, investigators do not randomly select participants; they can only offer a chance to participate. Patients decide whether to participate or not. A significant number of people who initially decide on participation eventually drop out or ask for other treatments. These realities essentially negate *random selection* of participants.

### Random Selection Does Not Happen in RCTs

Most treatment studies with group designs describe their random *assignment* of participants, not random selection that ensures sample representativeness. Those who assert that randomized controlled clinical trials are "the 'gold standard' against which other trials are judged" (Moscicki, 1993, p. 186) do not mention the need for random selection of participants. Several sampled books on medical statistics and randomized controlled clinical trials in medicine fail to address random selection of participants (Bland, 1995; Brody, 1998; Campbell & Machin, 1999; Finn, 1999; Meinert, 1986; Mike & Stanley, 1982; Mould, 1998; Pocock, 1983; Tygstrup, Lachin, & Juhl, 1982). Most medical reports on randomized control trials of new drugs or surgical procedures neglect to include random selection of subjects from a defined population, as pointed out in the next section on medical treatment studies.

Fisher (1925, 1942, 1956) could easily randomize agricultural plots in a large field to experimentally manipulate such independent variables as seed quality and water quantity to produce differential effects on crops. The divided plots of land in Fisher's experiments had an

equal chance of being selected for a particular treatment and could not reject or change a group membership. Marbles in a bucket placed on a classroom desk to demonstrate random selection may have an equal chance of being picked for nothing, but in a study on treatment of aphasia, few, if any, persons with aphasia in the population have the same chance of being selected for the study. Most of those patients are not even identified. Dice repeatedly thrown may help illustrate the principle of random outcomes, but children recruited with great difficulty for a treatment study are not randomly selected. Except in national survey research (a *nonexperimental* type of research), random selection has not worked in human *experimental* research. It has not worked in any human treatment research. In human clinical treatment research, random sampling of patients is not only impossible, but it also is unsustainable in light of legal research mandates. As Feinstein put it in 1970, “one of the most pernicious scientific delusions now prevalent in the world of medical research is the idea that concepts of ‘random sampling’ can be readily applied to clinical populations” (p. 288). Many medical researchers now believe that RCTs are unsustainable (Bjerklie, 2002; Blichert-Toft, Mouridsen, & Andersen, 1996; Dreyfuss, 2004; Feinstein, 1970; Gorkin, et al., 1996; MacIntyre, 1991; Silverman & Altman, 1996). One would hope that speech-language pathologists critically examine the methods, requirements, promises, and failures of RCTs before embracing them as the “gold standards” of treatment efficacy research.

The idea that randomization prevents bias in participant selection is a widely accepted statistical dogma, not an empirical reality of RCTs. Feinstein (1995), has recommended that researchers “stop believing the often stated dogma that ‘randomization prevents bias’” (p. 78). Pringle and Churchill (1995) have stated that “It would be wrong to stick blindly to a gold standard which is likely to produce the wrong findings—methodologically pure but clinically meaningless” (p. 1382).

### **Random Assignment Is Not Sustained in RCTs**

Random assignment of subjects is the second stage of randomization. Presumably randomly selected or simply available and initially willing participants may be randomly assigned to either the experimental treatment group or a variety of control group (e.g., a placebo-controlled group, no-treatment group, or a standard treatment group). In RCTs, investigators may initially assign recruited (not randomly selected) potential participants to the experimental and control groups, but they have no control on sustaining this random assignment until

the end of the study. In practice, these initially randomly assigned groups rapidly regroup themselves into self-selected individuals.

Group design researchers are often dismayed that single-subject researchers will do treatment research with small samples that offer no possibility of generality (external validity). Unfortunately, RCTs that cannot randomly draw participants and then sustain random assignment into groups do not offer generality either. In fact, some medical research experts who recommend only the randomized control group method for establishing treatment effectiveness have nonetheless suggested that “any generalization that goes beyond the study population must be made with caution and is judgmental rather than statistical” (Meinert, 1986, p. 206). Edgington had emphatically stated that “in the absence of random samples . . . generalization to other individuals [is] logical, nonstatistical considerations” (1967, p. 195).

Random procedure does not offer any special advantages for group designs because the method is rarely used according to the theory of probability. However, it may be fine to argue that random selection is too unrealistic an expectation and that random assignment is an acceptable compromise. In practice, both single-subject and group designs force acceptable compromises. If so, one would have to conclude that group design research does not offer superior statistical generality. And as argued in this paper, both single-subject and the group design researchers will have to settle for time-consuming replications as the only means of achieving generality.

Although educational, behavioral, and social researchers recognize the importance of generality (Barlow & Hersen, 1984; Campbell & Stanley, 1966; Hegde, 2003; Kerlinger, 1986; Sidman, 1960), medical researchers and medical statisticians pay only a meager attention to it (Bland, 1995; Brody, 1998; Campbell & Machin, 1999; Finn, 1999; Meinert, 1986; Mike & Stanley, 1982; Mould, 1998; Pocock, 1983; Tygstrup, Lachin, & Juhl, 1982). If generality is not an issue, mere random assignment is sufficient to claim internal validity, provided that random assignment was held until the end of the experiment. But serious objections can be raised when it is insisted that randomized control trials are the standard for establishing internal validity as well as generality.

### **All Participants in Randomized Clinical Trials Are Self-Selected**

Self-selection of participants is a known data-distorting bias in all human experimental research (Campbell & Stanley, 1966). In fact, randomization is used in human research only to avoid self-selection or investigator’s biased selection of participants. Self-selection pos-

es more serious problems for generality than for internal validity. Self-selection has been the de facto subject selection method since the dawn of the era of informed consent in research. Random assignment of subjects to groups—even if it was sustained until the end of the experiment—does not counter this problem. This is true of randomized control trials in medicine that have included several thousand patients in multiple centers. All treatment researchers recruit available and willing clients and complete studies only with those who *choose* to stay until the end.

As noted before, in many countries where RCTs are conducted, government laws and ethical guidelines require researchers to use only those participants who consent to participate and willingly continue until the trial is completed. Many eligible patients do not consent to participate, and those who do, retain their right to refuse randomization, reject the randomly offered treatment, demand the new experimental treatment or an established treatment when assigned to a control group, drop out of the study at any point, request their principal investigator to change the treatment protocol in ways that will invalidate the results, and so forth. These factors produce two effects. First, those who fully participate in clinical trials are indeed self-selected, not randomly selected. Second, many participants assign themselves to treatment, no-treatment, or placebo control groups. Therefore, there is very little justification for claiming random assignment of subjects to different groups of an experiment. Self-selected participants are fine for claiming internal validity, but the results of studies with such subjects cannot have statistical generality, believed to be the greatest strength of randomized clinical trials (Paci & Alexander, 1997).

The trouble with randomization begins when investigators first identify eligible patients and offer information to obtain informed consent. To recruit participants, some trials use either an “opt-in” or an “opt-out” approach (Junghans et al., 2005). In the opt-in approach, the patients are requested to actively signal that they are interested in trial participation. Patients who do not signal to be included may not be contacted further. In the opt-out approach, patients are repeatedly contacted until they refuse participation, raising potential question of subtle coercion. In either strategy, many eligible patients refuse participation (Fairhurst & Dowrick, 1996; Feine, Awad, & Lund, 1998; Pullman & Wang, 2001; Silverman & Altman, 1996; Snowden, Elbourne, & Garcia, 1999; Taylor, 1985; Torgerson & Roland, 1998). In one analysis, the opt-out strategy helped recruit up to 50% of eligible patients whereas the opt-in strategy helped recruit only 38% (Junghans, 2005). Blichert-Toft, Mouridsen, and Andersen (1996) estimate that generally, the number of eligible patients who refuse participation in randomized

studies is 50% or higher. In a Cooperative VA clinical trial on total parenteral nutrition in malnourished surgical patients, 395 consented, and 233 (58%) refused participation (Williford, Krol, & Buzby, 1993). Another study that analyzed selection and participation bias found that only 13% of the 325 eligible and invited patients eventually participated in a randomized study on a psychosocial intervention for multiple sclerosis (MS). Even that number of eligible patients was only a small portion of 1,500 MS patients seen at the research facility (Schwartz & Cox, 1995). Certain diseases seem to generate very high refusal rates. Many patients with prostate cancer, for example, refuse to be randomized (O’Reily, Martin, & Collins, 1999).

Even the dismal 50% or so recruitment rate of eligible patients may not hold when the patients are offered their first appointment for initial assessment. Up to 20% of patients who initially accept to participate may not show up for appointments (Junghans, et al., 2005), resulting in a further reduction in the number of eligible patients recruited for the study.

The very small number of eligible patients (even in a large sample) who select themselves into the study begin to have second thoughts about their participation when the investigators begin to explain random assignment to different groups of the study. Such second thoughts tend to create further drop-outs. The critical point here is that who gets what treatment or no treatment is greatly influenced by patients, not by random assignment, prompting the comment that patients get what they want (even in RCTs) “by voting with their feet” (Silverman & Altman, 1996, p. 171). In many randomized clinical trials, self-selected patients who have given informed consent may refuse to be randomized to a particular treatment or a placebo. For instance, in a coronary artery surgery study, nearly two-thirds of the 2,099 eligible patients refused random assignment to surgical treatment when the alternative was medical treatment (CASS Principal Investigators and Their Associates, 1984). Patients’ refusal to be randomized to unwanted treatments or placebo is so strong that medical researchers are concerned about the future of randomized clinical trials, or if conducted, the validity of the results (Blichert-Toft, Mouridsen, & Andersen, 1996; MacIntyre, 1991; Silverman & Altman, 1996). Some have suggested that as medical insurance becomes more universal, recruitment to randomized trials will become even more difficult because many who enroll themselves in RCTs do so because they lack that protection (Gorkin, et al., 1996). It is a disturbing thought that RCTs are partly supported by unfavorable socioeconomic conditions. Feinstein (1970) has stated that “as we enter the 21<sup>st</sup> century, the glaring handwriting on the wall is that randomized trials will be impossible—logical-

ly, ethically, and fiscally . . ." (p. 78). According to Dreyfus (2004, p. 574), "we may see in the future the twilight of randomized controlled trials in critically ill patients because scientific, ethical, and sociological substrata will be progressively lacking as will be funding." According to one survey, 80% of clinical trials have difficulty recruiting participants (Bjerklie, 2002), and most people who refuse participation in treatment research do so because of their aversion to randomization (Llewellyn-Thomas, et al., 1991).

Over the years, there has been an increase in the number of medical researchers who recognize that self-selection is a fact of all medical treatment research, including "randomized" clinical trials, and that this fact invalidates the claim of statistical generality (Barnett, Sackett, & Taylor, 1987; Begg & Engstrom, 1986; Feinstein, 1970; Taylor, et al., 1994). In medical literature, self-selection bias is also described as *nonconsent bias* (Marcus, 1997). Feinstein (1970) warned more than 30 years ago that participants in RCTs are volunteers, and this very fact makes them unrepresentative of the population of patients. Feinstein stated that the idea of a clinical sample representing the population of patients is "completely vitiated by the use of patients as the 'material' of clinical investigation, because the patient—unlike an agricultural field, chemical vat, or the material of any other type of experimentation—chooses the investigator, rather than vice versa" (1970, p. 288). Indeed, it is the patients who choose, not the investigators.

### **Investigators Induce Their Own Recruitment Biases**

In addition to patients, physicians, too, introduce bias into the participant pool. Physicians decide who among their patients initially volunteers for a study and who does not. In making that decision, physicians often use their own criteria, not necessarily the study criteria (Schwartz & Cox, 1995; Taylor, 1985). Their personal judgments about the treatments being evaluated will influence their recruitment effort; they may handpick patients for RCTs—hardly anything resembling randomization. In addition, many clinicians experience ethical dilemmas in RCTs in which their patients may be offered a less effective treatment, a new and risky treatment, or a placebo (Verdu-Pascal & Castello-Ponce, 2001). Ninety percent of responding pediatricians in a survey reported their ethical dilemmas when asked to refer critically ill children to randomized clinical trials; more than one-third of the principal investigators of RCTs may not recommend a significant portion of their eligible patients to their own study (Taylor, 1985). A parallel example would

constitute a speech-language pathologist who is serving or has access to 50 patients with aphasia, but asks only a few selected patients to volunteer to be randomized. The farcical nature of already dysfunctional randomization applied to patients who have been handpicked to be volunteers for a study is obvious.

After volunteering their selected patients to an RCT, physicians may disregard the study protocols, often with good justification and sound clinical judgment. For instance, in one survey, 65% of pediatricians who referred children to RCTs tended to alter the experimental treatment protocols or offer the experimental treatment to control group children whose condition deteriorated, resulting in questionable validity of results (Morris, Zaritsky, & LeFever, 2000). Another survey revealed that among physicians who were a part of a large oncology treatment research cooperative established by the National Cancer Institute, only 9% believed that their primary responsibility was to the future generation of patients who might benefit from RCTs (Taylor et al., 1994). Most believed that their primary responsibility was to their current patients. Randomization nullifies a physician's decision-making power in offering appropriate treatment to his or her patients. Consequently, many physicians who are a part of RCTs nonetheless select a small number of patients based on their own criteria (not necessarily the study criteria) and treat most of their patients outside the scope of the study. As stated previously, surgeons may perform surgery on large numbers of patients outside the scope of RCTs to which they are committed (Relman, 1987; Sundt, 1987). It is very likely that speech-language pathologists also will be reluctant to refer the clients they serve, especially children they serve, to RCTs that include no-treatment or sham-treatment control groups. The clinicians also may selectively refer their clients to RCTs, thus removing any semblance of randomization.

Physicians who agree to serve as investigators in RCTs may nonetheless be so reluctant to carry out randomized treatments on their patients that some of them may make systematic efforts to subvert randomization even after randomizing patients in a study. Schulz (1995) has summarized a variety of anonymous efforts on the part of the medical research personnel to subvert randomization. Such efforts range from simple to more complex, including transillumination of envelopes that contain random assignment of treatments to secretly searching for codes of randomization in the office of the principal investigators. Schulz has stated that such efforts are more common than realized, underscoring physicians' dilemma about the widespread acceptance of an ethically questionable method of documenting treatment efficacy.

## Consenters and Nonconsenters Are Different People

The representativeness of randomized samples, on thin ice to begin with, is further eroded by systematic differences that are evident in the three groups of people: those who consent to participate, those who refuse to participate, and those who initially consent but do not persist with the study. It has been shown that those who select themselves to participate in RCTs and those who refuse participation are heterogeneous (Fairhurst & Dowrick, 1996; MIAMI Trial Research Group, 1985; Smith & Arnesen, 1990; Williford, et al., 1993). Generally, consenters are more sick than nonconsenters. In many cases, those who volunteer for a study are desperate patients who do not have good alternative treatments for their disease, or if they are available, the patients cannot afford them because of lack of medical insurance (Lemonick & Goldstein, 2002; Minogue, Palmer-Fernandez, Udel, & Waller, 1995). Studies have shown that those who give consent to participate may not, and often do not, represent the population to which they belong in critical prognostic factors. Consenters and nonconsenters differ in their demographic and geographic characteristics. Most consenters live near medical research centers and those who refuse participation tend to live farther, often in remote rural areas (Schwartz & Fox, 1995). Furthermore, consenters and nonconsenters may differ in their mortality rate during or after a treatment study. Smith and Arnesen (1990) have shown that those who refused to participate in a trial of oral anticoagulation therapy after acute myocardial infarction had a higher mortality rate than those who were assigned to the placebo group.

Different kinds of RCTs have differential types of consenters and nonconsenters. There is evidence that those who consent to participate in randomized control trials on *treatment for diseases* tend to be poorer, less educated, uninsured for health services, and more disease prone, than those who refuse participation (Gorkin et al., 1996). Many patients who accept the riskier and newer treatment in RCTs are more seriously ill than those who do not participate or do not persist (Mitchell, 1981). To the contrary, those who consent to participate in a disease *prevention* study tend to be more affluent, better educated, and generally healthier than those who refuse participation (McKee et al., 1999).

Treatment knowledge and preferences for treatment will partly determine whether people will consent or not. Some patients who enter a randomized control trial tend to have a strong preference for a given treatment and regardless of the treatment to which they are randomized, may demand and get the preferred treatment. Others have no such preference and accept the random-

ly assigned treatment. Consequently, people who receive the experimental treatment because of lack of preference may be different from those who demand and receive their preferred treatment (Feine, Awad, & Lund, 1998). Some patients who reluctantly accept a randomly offered but nonpreferred treatment may not comply well with the treatment protocol, thus introducing additional factors that limit the validity of the results (Torgerson & Roland, 1998). Edlund, Craig, and Richardson (1985) have claimed that informed consent is a form of volunteer bias that may produce both false-positive and false-negative results. Schooler (1980) has maintained that refusal to consent may affect the final outcome of studies to an extent that significant results may turn out to be nonsignificant and vice versa. In some cases, the direction of effect may be reversed because of the informed consent bias (Schooler, et al., 1980).

There is evidence that people who quit a study and those who persist with it are different kinds of people (Fairhurst & Dowrick, 1996; MIAMI Trial Research Group, 1985; Smith & Arnesen, 1990; Williford, et al., 1993). In fact, those who drop out from the study may not be homogeneous with anyone, and many do drop out in large randomized clinical trials. Generally, though, patients who persist with the riskier and newer treatment in RCTs are more severely ill than those who do not persist (Mitchell, 1981). Such patients may also be less well informed about treatment options, and consequently, do not have a preferred treatment (Feine, Awad, & Lund, 1998).

There is evidence that being informed about the negative side effects of drugs may unravel another highly regarded but ethically questionable procedure of typically tattered randomized clinical trials: blinding (Brownell & Stunkard, 1982; Moscucci, Byrne, Weintraub, & Cox, 1987). Levine has written that "the subjects have been able to 'guess' correctly their treatment assignment. Such unblinding is, of course, detrimental to the purpose of using double-blind design in the first place" (1987, p. 247). That it is not possible to blind either the judges or patients to behavioral intervention procedures used in speech-language pathology is beside the point.

Informed consent, though an ethical and legal necessity of any kind of research with human subjects, tends to induce other kinds of data distortions. Levine (1987) has argued that informed consent and randomized placebo-controlled clinical trials may be incompatible with each other. Based on data from several clinical trials, Levine (1987) has concluded that when patients are informed of adverse effects of drugs, patients' "behaviors are likely to be influenced by having been so informed. In some cases, these behavioral changes may introduce serious biases in the results of the study" (p. 247). For instance, when patients are told about potential nega-

tive side effects of a new drug being evaluated, drop-out rate may increase among all patients—even among those assigned to the placebo-controlled group because, they, too, may begin to experience the negative side effects. A disturbing fact associated with this trend is that many patients who consent to participate in RCTs nonetheless do not fully understand the concept of randomization and the risks involved in subjecting themselves to new treatments (Cuttini, 2000; Lidz et al., 2004). It is likely that the number of nonconsenters may soar if patients gain a better understanding of randomization and risks associated with RCTs. Thus, universal health insurance and a clearer understanding of risks randomization poses to patients may make RCTs socially untenable.

In essence, both the patients and clinicians have serious problems with informed consent and randomization. The problem is so serious that some investigators believe both physicians and patients are obstacles to RCTs (Lacher, 1981). If so, one wonders for whose benefit and for what purpose randomization is used in clinical treatment experiments. Obviously, there are only two parties that remain interested in RCTs in medicine: the drug manufacturers and the RCT investigators. That informed consent is a hurdle to RCTs is widely and intensely debated in medical research literature (Doyal, 1997, 1998; Ellenberg, 1997; Tobias & Souhami, 1993 for reviews). There are suggestions that informed consent is not necessary in all cases, and in some cases it is acceptable to offer the participants only partial information on the study (Tobias & Houghton, 1994; Tobias & Souhami, 1993, among others). Although no one wants a return to the Nazi era of human experimentation, the devastating consequences of truncated or suspended informed consent are historically demonstrated (Hegde, 2003) and currently unacceptable.

### Other Critical Problems with RCTs

The main criticism against RCTs is that there is no randomization in them. In addition, there are many other problems associated with treatment efficacy studies that recruit large number of participants. It is instructive to consider two problems that produce negative effects on data interpretation and generality.

#### Large and Heterogeneous Samples Produce Ambiguous Data

It is the legacy of inferential statistics that requires researchers to select large samples for experiments. The greater the sample size, the higher the chances that in-

ferential statistical analyses will detect a treatment effect. Therefore, statisticians typically recommend clinical researchers to increase the sample size if they want to show that a given treatment is effective. Statisticians also offer a bonus to clinical researchers who sample large numbers of individuals: statistical analyses might still detect small or weak effects of treatment if the samples are large enough. Generally, this has led to the belief that bigger samples are better than smaller samples.

Bigger samples are indeed better in nonexperimental survey research where the goal is to predict the behavior of large sections of a society (e.g., all voting-age individuals who might elect a president, all consumers who might buy a car in a given year). To the contrary, bigger samples bring only bigger problems in treatment research in all disciplines, including medicine, clinical psychology, and speech-language pathology. Theoretically, larger samples have a better chance of representing the population than smaller samples. Unfortunately, the larger the sample, the greater the heterogeneity; the greater the heterogeneity, the greater the magnitude of variability of results found in the sample. In a large sample study, some participants may have experienced a great deal of improvement, others much less, some negligible amount, some no improvement at all, and some may even have deteriorated. But if a certain number of participants experienced a great deal of improvement, the post-test mean may reflect an overall positive effect, wiping out the negative effects in some and masking the limited effects in the others. The mean calculated for such a sample may be ambiguous at best and meaningless at worst. Assumed statistical generality of such a study only means that similar results would be found if the same treatment is applied to the successive samples drawn from the same population. This is true even in the rare case of a truly randomly *selected* and *assigned* samples that do ensure statistical generality. Even though an inferential statistical analysis may reveal a significant effect, the meaning of this effect remains ambiguous to individual practitioners. It is not possible to predict who improves and who does not under the same treatment when practitioners apply it to their individual patients.

In some RCTs, efforts may be made to reduce the sample heterogeneity by using an extremely narrow set of inclusion criteria. They do it for the obvious reason that more heterogeneous the sample, the harder it is to obtain impressive treatment effects (Fletcher, 2002). This strategy of creating more homogeneous samples, though desirable in some ways, negates the most widely touted advantage of RCTs: generality to the diverse population from which such samples are drawn. Homogeneous samples, no matter how large, are neither randomly selected nor supportive of generality.

## Inferential Generality Is a One-Way Street

Even in the best of RCTs, statistical (inferential) generality is a one-way street: Conclusions based on a representative sample may be extended to the population as a whole, but not to an individual or small numbers of individuals the practitioners treat (Barlow & Hersen, 1984; Feinstein, 1970; Hegde, 2003; Johnston & Pennyacker, 1993; Kazdin, 1982). Generality to individual patients gets progressively shrunk with successively larger samples. This is because the practitioners cannot judge whether the researched treatment will produce a significant favorable effect, an insignificant favorable or unfavorable effect, no effect, or a significant unfavorable effect in their individual clients. The pretest and post-test means calculated for the experimental and the control groups will mask this variability, but they cannot mask the difficulty in applying the results in professional practice. Consequently, RCTs are of no help when the help is urgently needed in clinical practice: selecting treatments that are good for individual clients. The belief that clinicians in health care professions can extend the results of large sample studies (especially those of RCTs) to their individual clients they serve is as fallacious as it is popular.

Many medical scientists now recognize this limitation of large-sample randomized studies. Dissatisfaction with the large N randomized trials is widespread (Blichert-Toft, Mouridsen, & Andersen, 1996; Cook, 1996; Dekkers & Boer, 2001; Demitrack, Faries, Herrera, DeBrotta, & Potter, 1998; Dreyfus, 2004; Feinstein, 1970, 1985, 1995; Kaptchuk, 2001; Marcus, 1997; Pringle & Churchill, 1995; Schooler, 1980, among others). Feinstein (1995) has stated that results of treatment studies that report average effects may hold good for the group as a whole but not to individual patients. He and others have taken exception to another prestigious statistical tool—meta-analysis of randomized trials. Feinstein (1995) dubbed the meta-analysis as the statistical alchemy for the 21<sup>st</sup> century because the method averages the mean changes in patients reported in individual studies, only to violate the scientific principles of precision and homogeneity. After reviewing selected meta-analyses in medicine, Bailar (1995) concluded that “problems [with meta-analyses] were so frequent and so serious, . . . that it was difficult to trust the overall ‘best-estimates’ that the method often produces” (p. 153).

Feinstein (1995) has stated that a problem with randomized trials “is the frequent claim that a large heterogeneous population gives the results ‘wider applicability’. . . The scientific effect of a diverse poorly identified mixture is to produce imprecision, confusion, and perhaps delusion, not generalizability” (p. 76). Feinstein (1970), consistent with the claim of single-subject re-

searchers, has pointed out that the results of RCTs “will be clinically meaningless because a clinician will not know how to apply them in the future; he cannot determine whether ‘good risk’ and ‘poor risk’ patients responded the same way to each therapeutic agent” (p. 289). In a similar vein, Fletcher (2002) stated that “Good clinicians think of their patients as individuals . . . In contrast, randomized controlled trials are about average results in groups of patients. It is a bad match” (p. 1188). Feinstein believed that in establishing cause-effect relationships in treatment research, “the challenging basic elements are raw data and small human groups” (1995, p. 78). Another proven alternative is single-subject experimental approach in which small numbers of participants are intensively studied.

## Control Groups of RCTs Are an Ethical Quagmire

The function of control groups is to help rule out extraneous variables in controlled experiments. In other words, control groups help demonstrate that treatment is better than no treatment. Control groups will work well in nonclinical research in which certain variables are manipulated with no negative consequences to participants. It will also work well in animal research and human basic laboratory research designed to produce temporary effects. However, in clinical treatment research, control groups generate negative consequences to participants that may match or exceed its benefits.

Denial of treatment to those who need it as badly as those who get it is the basic ethical dilemma of RCTs. Patients not only reject, but are fearful of randomization. For instance, Finn (1997, p. 17) stated that “fear of randomization is one of the most significant factors preventing cancer patients from considering clinical trials.” An increasing number of potential patients reject randomization mostly to avoid being randomly assigned to one of the control groups. Therefore, societal rejection of control groups, along with their rejection by many physicians, makes RCTs untenable to those who wish to respect the wishes of patients as well as professionals. Very few large-scale RCTs are attempted in speech-language pathology, but we already know that in one study, parents of control children who stuttered have demanded treatment before the experiment was over (Onslow, Andrews, & Lincoln, 1994). When people in the control group demand treatment, there is no choice but to terminate the experiment—a costly but inevitable, and the only available—ethical course of action.

There are other ethical concerns with control groups, some of them serious. For example, Brody (1998) and Dekkers and Boer (2001) raise such troubling questions

as the following: What exactly is the ethical justification for giving sham treatments in the form of placebo to patients who badly need a real treatment? What is the ethical justification for randomization when sick patients do not want to be randomized to no-treatment or deceptive sham-treatment control groups? What is the ethical justification of sham surgical treatment (e.g., burring a whole in the skull of patients with Parkinson's disease to simulate intracerebral transplantation of fetal mesencephalic tissue) for the sake of fooling some control group patients into thinking that they have gone through a beneficial surgical treatment? Many others have raised these and other ethical issues with randomized control trials in which patients are denied treatment or were offered sham treatment; most medical journals are replete with responses from physicians who question the concept of randomized control groups that involve placebo controls (Dekkers & Boer, 2001; Morris, Zaritsky, & LaFeever, 2000; Verdu-Pascual & Costello-Ponce, 2001). An examination of these and other serious issues relative to control groups is necessary before RCTs are accepted as the standard for establishing treatment effects in speech-language pathology. If the number of randomized control trails were to increase in speech-language pathology, the urgency with which we examine the need for a control group will only become progressively more acute, with only distressing consequences for both the clients and researchers.

## Randomized Control Trials in Medicine

### Versus Speech-Language Pathology

There are important methodological differences in treatment efficacy research in medicine and speech-language pathology and similar disciplines (e.g., education, clinical psychology, health sciences, social work, physical therapy). These differences should be considered before emulating RCTs of medicine.

### Medicine and Speech-Language Pathology Have Vastly Different Resources

Large N studies conducted simultaneously in multiple centers require enormous resources and trained personnel that are usually not available to support such treatment research in speech-language pathology. Government, private individuals, charitable organizations, and pharmaceutical industry offer vast financial support to medical research that speech-language pathology lacks. To help recruit patients to RCTs, physicians may be paid a per capita fee, often a substantial amount.

Many randomized drug trials in multiple centers and countries are entirely supported by global pharmaceutical companies that need their new drugs evaluated. Such companies spend up to a billion dollars to have a new drug evaluated in RCTs. When multiple centers are recruited internationally, literally hundreds of medical personnel in several centers might participate in surgical or pharmaceutical treatment evaluation studies (e.g., see the International Stroke Trial Collaborative Group, 1997; Post Coronary Artery Bypass Graft Trial Investigators, 1997; the Scandinavian Simvastatin Survival Study Group, 1994).

Although pharmaceutical companies generously support RCTs, one cannot ignore the ethical issues associated with them. Companies that sponsor RCTs own the data the medical researchers produce. The researchers are not free to publish their findings, and may not even have access to all the findings collected in a study. To further intensify ethical concerns, some RCT investigators may be financial partners with the sponsoring entities (Perlis et al., 2005; Tuech et al., 2005). There may be a tendency to report only positive findings while suppressing negative findings about new drugs, medical devices, or other treatment procedures. A recent example is the alleged partial suppression of negative evidence to obtain Federal Drug Administration's approval for Guidant Corporation's implantable cardiac defibrillator, which later failed and caused deaths (Eggerston, 2005). The tendency to report only positive findings is generally much higher when the investigators also are financial partners with the study sponsors (Perlis, et al., 2005). Reviewing 102 randomized clinical trials, Hrobjartsson et al. (2005) found that nonsignificant treatment outcomes tended to be reported only partially whereas significant outcomes were reported more completely. Hrobjartsson et al. concluded that "outcomes in randomized trials are often reported selectively" (p. 3189). Analyzing 45 industry-sponsored RCTs in arthritis drug development, Fries and Krishnan (2004) found that each and every study reported positive findings. A knowledge of study sponsorship was all that was needed to predict a study's results with 100% certainty.

Undisclosed payment to physicians to recruit patients for RCTs raise additional ethical issues. Recruited patients are totally unaware that their physicians have steered them into a study because of financial gain. Believing that their physicians have acted in the interest of science, patients may submit themselves to dangerous treatments, lack of treatment (control group) when one is urgently needed, or sham treatment (placebo group) that will worsen their disease. Such financial incentives offered to physicians may induce participant selection biases. In general, such incentives are unethical (Puma, Stocking, Rhoades, & Darling, 1995; Rao & Cassia, 2002),

and fortunately, unlikely to be available to most speech-language pathologists.

Thousands of self-selected patients in hospitals and clinics are available for medical treatment research, but almost nonexistent in behavioral treatment research done in speech-language pathology. This is an important distinction with compelling consequences for an unqualified recommendation of RCTs for treatment research in speech-language pathology and many other human service professions.

### Medical and Behavioral Treatment Procedures Are Vastly Different

Another, equally important difference between treatment research in medicine and speech-language pathology lies in the independent variable (nature of treatment) and the dependent variables (treatment outcomes or effects) evaluated in the two disciplines. This difference has an effect on treatment research design selection.

*Treatment in medicine* is relatively easily administered by a large number of well-trained professionals in multiple centers when they are given strict and detailed protocols on how, when, and how much of a new drug should be administered to thousands of study patients. With minimal or no additional training cost, medical personnel in multiple centers in several countries can administer a drug according to a prescribed protocol. For example, the International Stroke Trial (International Stroke Trial Collaborative Group, 1997) consisting 19,435 stroke patients in 467 hospitals in 36 countries evaluated the antithrombotic effects of aspirin, subcutaneous heparin, both, or neither administered for 14 days to patients in randomly assigned groups. The dependent variables were the number of patient deaths in 14 days and the number of deaths or dependency in 6 months. One can only contrast the logistics of administering a pill with the administration of a phonologic treatment to a large number of children selected across several schools in multiple countries.

*Treatment in speech-language pathology*, compared to that in medicine, is much more difficult to administer to large numbers of individuals. Several reasons are obvious. First, treatment in speech-language pathology is not standardized as it is in medicine. For instance, treatments of aphasia, dysarthria, stuttering, articulation, and language disorders in children do not have standardized protocols on which a large number of clinicians agree or are similarly trained. Drugs come with standard protocol of administration that medical personnel in single or multiple settings can use without controversy or variation across clinicians. In a large drug evaluation study, the administration protocols on frequency and dosage

can be written briefly and clearly to limit variations. Treatment in speech-language pathology is much more varied, involved, and complex. Even the same treatment is often administered with great variations. For instance, the Lidcombe Program for early stuttering intervention is currently used in different countries, but the method is locally modified in multiple ways, some described and some perhaps never described (Onslow, Packman, & Harrison, in press). Speech-language treatment protocols can be written for multicenter randomized or single-subject design studies on treatment effects. However, the use of such protocols will require a large training effort, costing a great deal of money. This issue has received little or no attention in the discipline, however.

Second, treatment in speech-language pathology is labor intensive, much more so than medical treatment. A typical treatment in a university clinic for a child or an adult who stutters might take 9 months of individual therapy, offered twice weekly, in sessions lasting 45 to 60 minutes. A randomized clinical trial in the treatment of language disorders in children with a large enough number to be randomized will involve an incredible number of clinician-child contact hours. However, the administration of a drug (e.g., aspirin to stroke patients), may require a few minutes of daily patient contact (if that) and is likely to be over in a few days or weeks. Medical researchers do publish studies in which patients were treated for years or followed up for many more. Significant structural and financial differences in speech-language-hearing clinics and medical facilities explain why such studies are more practical in medicine than in speech-language pathology.

Third, training of a substantial number of speech-language pathologists required to administer treatment uniformly in a large randomized control trial will pose significant practical and economic hurdles. The task of training medical personnel in following a drug administration protocol for a randomized study would involve much less effort, training time, and research funds than training a large number of speech-language pathologists to uniformly administer a treatment for stuttering, phonologic disorder, aphasia, or dysarthria. This situation is not likely to improve soon just because speech-language pathologists want to emulate the medical model of randomized clinical trials.

Fourth, the differences in the measurement of dependent variables in medicine and communicative disorders also pose problems for speech-language pathologists conducting randomized control trials. Documenting changes in communication skills due to treatment (dependent variables) is more labor intensive, riddled with controversies, and takes more time than documenting changes due to medical treatment. A large randomized clinical trial may compound this problem to an extent

that achieving reliable measures will be extremely difficult. Although unreliable dependent variable measurement is a problem in medicine as well (e.g., questionable results of a single laboratory test for blood cholesterol), less time-consuming multiple measures help establish reliability with manageable time, effort, and cost. To the contrary, achieving reliability through multiple measures of communicative behaviors is much more difficult because of the time, effort, and expense involved (e.g., recording, transcribing, and analyzing multiple language samples for each of several hundred children in a randomized language treatment study). In addition, some dependent variable measures as mortality rates are much more easily (and unquestionably) documented than subtle changes in speech and language skills.

Although the differences between medical and speech-language pathology treatment procedures contrast the most, even different branches of medicine and medicine versus clinical psychology provide interesting contrasts. Although drug and medical screening (e.g., screening mammography) studies report on exceptionally large number of initially randomized but eventually self-selected participants, psychiatric and clinical psychological studies, which offer more varied and labor intensive psychotherapy and counseling, report on much smaller number of participants. For instance, Shapiro and Shapiro (1983) have documented that in psychotherapy research, the average number of participants in a majority of RCTs did not exceed 12 to 13 participants. When the validity of randomization is questionable with large numbers of participants, it is meaningless with such small numbers.

### Multiple Strategies for Speech-Language Pathology

In his critical examination of randomized and masked (blinded) clinical trials in medicine, Kaptchuk (2001) makes an observation that will also characterize the motivation to write this critical review. Kaptchuk wrote that

While a gold standard is valuable, any worshipping at an altar of a golden calf, like the Biblical Exodus story, may obscure "reality" . . . Unless one is aware of a research methodology's weaknesses, scientific activity may become a mechanical ritual (p. 548).

Treatment researchers select experimental designs based on their training, experience, expertise, and research philosophy. And it will continue to be that way. Those who typically use a particular strategy will retain

a healthy critical disposition toward the one they do not use. This is good for the science and the profession because the skeptics of any approach will help keep the enthusiasts a notch below extremists. However, a serious impediment to varied treatment research is created when the professional establishment upholds one approach to treatment efficacy research as the gold standard. In any case, a *gold standard* is a hyperbole. Such overstatements have prompted some investigators to wonder whether RCTs are a gold standard or fool's gold (Pringle & Churchill, 1995).

### Retooling the Group Treatment Research Designs

The criticism offered in this paper applies most forcefully to theoretical and methodological aspects of *randomization* in group treatment research. Group research designs, though traditionally and supposedly are based on randomization, can serve the profession better if they are reconceptualized and retooled to (a) maximize experimental control in treatment research, (b) de-emphasize the importance of randomization, (c) show restraint in making claims of generality with single or just a few studies, or (d) offer data on each individual participant with more detailed information on subject characteristics. Group experimental designs that recruit small numbers of subjects and then produce large treatment effects may generate data that ensure internal validity even better than large N randomized trials. Groups formed for treatment research may be either homogeneous or heterogeneous, although the traditional concepts of homogeneity or heterogeneity of participants is of limited validity. For example, all white persons in the United States (or African Americans or Hispanic Americans in the country) who are of the same age and gender, belong to the same socioeconomic class, enjoy the same health status, are educated similarly, and have the same living conditions may be considered homogeneous; but they may all not react the same to the same treatment. Conversely, individuals who belong to different ethnocultural, educational, age, and socioeconomic groups may be considered heterogeneous; nonetheless, they may all react the same to the same treatment. Therefore, what is more important than the statistical means for the groups (no matter whether the groups were homogeneous or heterogeneous) is individual profiles of participants, some of whom may react well, others poorly, and some not at all to the experimental or control treatments.

Individual profile of participants, when provided, will greatly attenuate the limitations of most group design treatment studies. Obviously, the need for such profiles will preclude large samples, randomized or not; but

such a preclusion that leads to smaller samples, it is argued here, is beneficial to practitioners because of the details the study can offer on individual variability in response to treatment. Individual profiles will help the practitioner determine who in the study (with what characteristics) improved, who did not, and who deteriorated. The clinicians can then match their clients with the characteristics of individuals who experienced different outcomes in a treatment research studies. Currently as practiced, even the small-group treatment studies fail to give adequate individual profiles of participants and individual data points; such profiles and data points are impractical in large-scale studies (presumably randomized or not).

All treatment studies, no matter what designs they use, should report the details on drop-outs. The report should include not just the number of individuals who dropped out from a study, but also their personal profiles. This kind of information will help restrain practitioners from overgeneralizing the results to the kinds of individuals who dropped out.

Group design studies may have to limit the use of the no-treatment control group or sham-treatment control group. For the most part, group research designs may be limited to studies in which all groups receive treatment, one of which may be the experimental treatment. This approach may not be used to show that treatment is better than no treatment, but it may be used to document the *relative effectiveness* of different treatments. In this strategy, a new and untested treatment may be compared only against other proven treatments or against standard treatments that are typically offered with little or no experimental evidence. This should not be difficult at all in speech-language pathology; there are both proven techniques and unproven but widely practiced techniques in speech-language pathology that can serve as controls against new treatments to be evaluated. Group treatment researchers may be reluctant to accept this strategy, but progressively better-informed consumers may force this option on them.

A potential problem in using a widely practiced but unproven technique as the control for a new technique in a multigroup, multitreatment research design is that the control group may show no improvement at all. In this case, the study will have used a de facto control group (no change, no improvement). However, a generally applied treatment that eventually proves ineffective may be more acceptable to consumers than no treatment at all.

Because no attempt will be made to randomize the participants, the approach of conducting small-group studies will emphasize direct and systematic replications (Hegde, 2003). Replication of small-group experimental treatment studies is a more practical research ef-

fort than either conducting the initial large-scale studies or to replicate them. Although direct replications will establish the reliability of treatment effects, systematic replications will establish generality (or lack thereof).

To make the data generated from small-group experimental studies more useful to clinicians, the current practice of reporting group design research needs to be changed. All investigators who use group experimental designs also should report data for individual participants, not just the means for the different experimental and control conditions to which the groups were exposed. If journals do not have space for such detailed data reporting, a system of freely and publicly available archival of database into which investigators can deposit individual data should be created. Limiting journal space made available for speculative discussions and tentative interpretation of results may free some space for more valuable individual data.

The group experimental designs are neither the most effective nor the most efficient in establishing the effects of a single treatment; single-subject designs do better. Small-group designs will avoid some of the problems of RCTs if they only used multiple groups to evaluate the relative effectiveness of two or more treatments without a control group that does not receive treatment. But the *factorial* group experimental designs (Hegde, 2003) are better than single-subject designs in analyzing interactions between two or more treatments and interactions between a treatment and subject characteristics (e.g., the severity of the disorder; and age, gender, education, and socioeconomic status of the participants). Thus, the group designs may play a primary role in studying relative effects of different treatments and various forms of interactions. Single-subject designs may play a primary role in establishing the effects of various treatments.

### Recognizing the Importance of Single-Subject Designs

One of the ironies of the recent emphasis on RCTs in speech-language pathology treatment research is that much of the evidence to support its professional practice comes from single-subject designs. Most speech-language pathologists do not subscribe to the conclusions of the *Cochrane Database of Systematic Reviews* that many treatment procedures in communication disorders have not been shown to be effective. Although this is not the place to review evidence for treatment effects generated by single-subject designs, behavioral treatment procedures used in modifying articulation and phonologic disorders, language disorders in children, communication disorders associated with autism, language dis-

orders associated with intellectual disabilities, stuttering in adults and children, and aphasia in adults have considerable experimental support, derived mostly from single-subject experimental analysis, not RCTs.

Single-subject experimental designs produce convincing evidence of treatment effects in small numbers of participants (not necessarily single participants). Not all single-subject experimental designs are quasiexperimental; both single-subject and group research approaches have their versions of relatively weak designs. Unlike group designs, single-subject designs contrast the same participants' behaviors or skills under experimental (treatment) and control (no treatment) conditions. Showing the appearance and disappearance of treatment effects in the same individuals may be even more powerful than comparing the presence and absence of effects in different individuals. By offering treatment to all participants, most single-subject designs avoid control groups and the problems associated with them.

Treatment effects studied by behavioral scientists and speech-language pathologists have been widely replicated, in spite of "small samples" used in their research designs. Such studies are much easier to replicate than large-scale RCTs, and thus encourage replication. The results of single-subject designs are immediately useful to practitioners because they can match the profiles of participants to those of their own clients in selecting treatment procedures for them—something that cannot be done with the results of RCTs.

### Creating a Dual-Strategy of Treatment Research

The single-subject experimental designs and the retooled group experimental designs provide an effective means for establishing treatment effects and demonstrating generality of those effects. A dual strategy in which the results of both kinds of design options are valued equally will be more helpful in expanding the treatment efficacy database than the currently popular "gold standard" approach.

Within this dual strategy, speech-language pathologists can exploit the strengths of each approach while minimizing the weaknesses of both. Retooled small group experimental designs may be preferred over single-subject designs to evaluate the relative effects of multiple treatments; single-subject designs may be preferred over group designs to establish the effects of single treatments. Interactions of two or more treatments and interaction between a treatment and participant characteristics may be better studied with factorial group designs than single-subject designs; clients' preference of one treatment over the other may be better studied in single-

subject design studies than group design studies. There are other ways in which the strengths of the two approaches may be exploited (Hegde, 2003).

When the group design researchers begin to emphasize small groups that are studied more intensively, generate clinically obvious and personally meaningful effects, provide more specific and useful individual data, and use replication instead of randomization as the means to establish generality of treatment effects, much of the criticism offered against RCTs will be neutralized. More importantly, the profession will have expanded options to enhance treatment efficacy database.

**Acknowledgment** I am grateful to the editorial feedback on an earlier version of this paper; constructive criticisms and suggestions from the Editor and the two anonymous reviewers have been especially useful in revising this paper.

**Address correspondence to** M. N. Hegde, Department of Communicative Disorders, 5310 N. Campus Drive M/S PH80, California State University-Fresno, Fresno, CA 93740-8019; e-mail: girih@csufresno.edu; Telephone: (559) 278-3938.

### References

- American Speech-Language-Hearing Association. (2004). *Evidence-based practice in communication disorders: An introduction* [Technical report]. Available at: <http://www.asha.org/members/deskref-journals/deskref/default>
- Ansel, B. M. (1993). Treatment efficacy research in stuttering. *Journal of Fluency Disorders, 18*, 121-123.
- Bailar, J. C. (1995). The practice of meta-analysis. *Journal of Clinical Epidemiology, 48*, 149-157.
- Barlow, D. H., Hayes, S. C., & Nelson, R. O. (1984). *The scientist practitioner: Research and accountability in clinical and educational settings*. New York: Pergamon.
- Barlow, D. H., & Hersen, M. (1984). *Single-case experimental designs* (2<sup>nd</sup> ed.). New York: Pergamon.
- Barnett, H. J. M., Sackett, D., & Taylor, D. W. (1987). Are the results of the intracranial-extracranial bypass trial generalizable? *New England Journal of Medicine, 316*, 820-824.
- Begg, C. B., & Engstrom, P. F. (1986). Eligibility and extrapolation in cancer clinical trials. *Journal of Clinical Oncology, 4*, 972-974.
- Bizzini, M., Childs, J. D., Piva, R. S., & Delitto, A. (2003). Systematic review of the quality of randomized trials for patellofemoral pain syndrome. *Journal of Orthopedic Sports and Physical Therapy, 33* (1), 4-20.
- Bjerklie, D. (2002). They are dying to get in. *Time*, April 22 [retrieved from [timeinc.com](http://timeinc.com) on April 20, 2002].
- Bland, M. (1995). *Introduction to medical statistics*. New York: Oxford University Press.

- Blichert-Toft, M., Mouridsen, H., & Andersen, K. W. (1996). Clinical trials. *Seminar in Surgical Oncology*, 12(1), 32–38.
- Box, J. F. (1978). *R. A. Fisher: The life of a scientist*. New York: Wiley.
- Brody, B. A. (1998). *The ethics of biomedical research: An international perspective*. New York: Oxford University Press.
- Brownell, K. D., & Stunkard, A. J. (1982). The double-blind in danger: Untoward consequences of informed consent. *American Journal of Psychiatry*, 139, 1487–1489.
- Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.
- Campbell, M. J., & Machin, D. (1999). *Medical statistics: A common sense approach* (3<sup>rd</sup> ed.). New York: Wiley.
- CASS Principal Investigators and Their Associates. (1984). Coronary Artery Surgery Study (CASS): A randomized trial of coronary artery bypass surgery. *Journal of the American College of Cardiology*, 3, 114–128.
- Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology*, 66, 7–18.
- Clatterbaugh, K. (1999). *The causation debate in modern philosophy 1637–1739*. New York: Routledge.
- Couture, E. (1996). Treatment efficacy: Stuttering. *Journal of Speech and Hearing Research*, 39, S18–S26.
- Cook, D. J. (1996). Randomized trials in single subjects: The N of 1 study. *Psychopharmacological Bulletin*, 32(3), 363–367.
- Curlee, R., & Yairi, E. (1997). Early intervention with early childhood stuttering: A critical examination of the data. *American Journal of Speech-Language Pathology*, 6(2), 8–18.
- Curlee, R., & Yairi, E. (1998). Treatment of early childhood stuttering: Advances and research needs. *American Journal of Speech-Language Pathology*, 7(3), 20–26.
- Cuttini, M. (2000). Proxy informed consent in pediatric research: A review. *Early Human Development*, 60, 89–100.
- Deane, K. H. O., Whur, R., Playford, E. D., Ben-Shlomo, Y., & Clarke, C. E. (2006a). Speech and language therapy versus placebo or no intervention for dysarthria in Parkinson's disease. *Cochrane Database of Systematic Reviews*, 3. Available in the Cochrane Library (ISSN 1464-780X).
- Deane, K. H. O., Whur, R., Playford, E. D., Ben-Shlomo, Y., & Clarke, C. E. (2006b). Nonpharmacological therapies for dysphagia in Parkinson's disease. *Cochrane Database of Systematic Reviews*, 3. Available in the Cochrane Library (ISSN 1464-780X).
- Demitrack, M. A., Faries, D., Herrera, J. M., DeBrot, D., & Potter, W. Z. (1998). The problem of measurement error in multi-site clinical trials. *Psychopharmacology Bulletin*, 34(1), 19–24.
- Dekkers, W., & Boer, G. (2001). Sham neurosurgery in patients with Parkinson's disease: Is it morally acceptable? *Journal of Medical Ethics*, 27(3), 151–156.
- Doyal, L. (1997). Journals should not publish research to which patients have not given fully informed consent. *British Medical Journal*, 414, 1107–1111.
- Doyal, (1998). Informed consent in medical research. *British Medical Journal*, 316, 1000.
- Dreyfuss, D. (2004). Beyond randomized, controlled trials. *Current Opinion in Critical Care*, 10, 574–578.
- EC/IC Bypass Study Group (1985). Failure of extracranial-intracranial arterial bypass to reduce the risk of ischemic stroke: Results of an international randomized trial. *New England Journal of Medicine*, 313, 1191–1200.
- Edlund, M. J., Craig, T. J., & Richardson, M. A. (1985). Informed consent as a form of volunteer bias. *American Journal of Psychiatry*, 142, 624–627.
- Eggerston, L. (2005). Physicians want transparency as Guidant lawsuits grow. *Canadian Medical Association Journal*, 173(8), 855–856.
- Ellenberg, S. S. (1997). Informed consent: Protection or obstacle? Some emerging issues. *Controlled Clinical Trials*, 18, 628–636.
- Fairhurst, K., & Dowrick, C. (1996). Problems with recruitment in a randomized controlled trial of counseling in general practice: Causes and implications. *Journal of Health Service Research Policy*, 1, 77–80.
- Feig, S. A. (2005). Screening mammography controversies: resolved, partly resolved, and unresolved. *Breast Journal*, 11(Suppl. 1), S3–S6.
- Feine, J. S., Awad, M. A., & Lund, J. P. (1998). The impact of patient preference on the design and interpretation of clinical trials. *Community Dental and Oral Epidemiology*, 26(1), 70–74.
- Feinstein, A. R. (1970). Statistics versus science in the design of experiments. *Clinical Pharmacology and Therapeutics*, 11, 282–292.
- Feinstein, A. R. (1985). An additional basic science for clinical medicine: II. The limitations of randomized trials. *Annals of Internal Medicine*, 99(4), 544–550.
- Feinstein, A. R. (1995). Meta-analysis: Statistical alchemy for the 21<sup>st</sup> century. *Journal of Clinical Epidemiology*, 48, 71–79.
- Finn, R. (1999). *Cancer clinical trials*. Sebastopol, CA: O'Reilly.
- Fisher, R. A. (1925). *Statistical methods for research workers*. London: Oliver & Boyd.
- Fisher, R. A. (1942). *Design of experiments*. London: Oliver & Boyd.
- Fisher, R. A. (1956). *Statistical methods and scientific inference*. London: Oliver & Boyd.
- Fletcher, R. H. (2002). Evaluation of interventions. *Journal of Clinical Epidemiology*, 55, 1183–1190.
- Fries, J. F., & Krishnan, E. (2004). Equipoise, design bias, and randomized controlled trials: The elusive ethics of new drug development. *Arthritis Research and Therapy*, 6(3), R250–255.
- Frisell, J., Lidbrink, E., Hellstrom, L., Rutqvist, L. E. (1997). Follow-up after 11 years: Update of mortality results in the Stockholm mammographic screening trial. *Breast Cancer Research and Treatment*, 45, 263–270.
- Gates, W., & Atherton, H. (2001). The challenge of evidence-based practice for learning disabilities. *British Journal of Nursing*, 10(8), 517–522.
- Gierut, J. A. (1998). Treatment efficacy: Functional phonological disorders in children. *Journal of Speech, Language, and Hearing Research*, 41, S85–S100.
- Gorkin, L., Schron, E. B., Handshaw, K., Shea, S., Kinney, M. R., Branyon, M., Campion, J., Bigger, J. T., Jr., Sylvia, S. C., Dug-

- gan, J., Stylianou, M., Lancaster, S., Ahern, D. K., & Follick, M. J. (1996). Clinical trial enrollers vs. nonenrollers: The Cardiac Arrhythmia Suppression Trial (CAST) Recruitment and Enrollment Assessment in Clinical Trials (REACT) project. *Controlled Clinical Trials, 17*(1), 46-59.
- Gøtzsche, P. C., & Olsen, O. (2000). Is screening for breast cancer with mammography justifiable? *Lancet, 355*, 129-134.
- Greener, J., Enderby, P., & Whur, R. (2006). Speech and language therapy for aphasia following stroke. *Cochrane Database of Systematic Reviews, 3*. Available in the Cochrane Library (ISSN 1464-780X).
- Hegde, M. N. (1998, April). *Treatment research: Market-driven or science-driven?* Paper presentation at the 1998 Treatment Efficacy Research Conference, Vanderbilt University.
- Hegde, M. N. (2001, November). *Treatment research: Designs and evaluative criteria*. A symposium presented at the Annual Convention of the American Speech-Language-Hearing Association, New Orleans.
- Hegde, M. N. (2003). *Clinical research in communicative disorders: Principles and procedures* (3rd ed.). Austin, TX: Pro-Ed.
- Heise, D. R. (1975). *Causal analysis*. New York: Wiley.
- Holland, A. L., Fromm, D. S., DeRuyter, F., & Stein, M. (1996). Treatment efficacy: Aphasia. *Journal of Speech and Hearing Research, 5*, S27-S36.
- Horwitz, R. I. (1987). Complexity and contradiction in clinical trial research. *The American Journal of Medicine, 82*(3), 498-510.
- Hrobjartsson, A., Chan, A. W., Haahr, M. T., Gotzsche, P. C., & Altman, D. G. (2005). [Selective reporting of positive outcomes in randomized trials: A comparison of protocols with published reports]. *Ugeskr Laeger, 167*(34), 3189-3191. Article in Danish; PubMed abstract (PMID: 16117920).
- International Stroke Trial Collaborative Group. (1997). The International Stroke Trial (IST): A randomized trial of aspirin, subcutaneous heparin, both, or neither among 19435 patients with acute ischaemic stroke. *Lancet, 349*, 1569-1581.
- Jones, M., Gebiski, V., Onslow, M., & Packman, A. (2001). Design of randomized controlled trials: Principles and methods applied to a treatment of early stuttering. *Journal of Fluency Disorders, 26*, 247-267.
- Junghans, C., Feder, G., Hemingway, H., Timmis, A., & Jones, M. (2005). Recruiting patients to medical research: Double blind randomized trial of "opt-in" versus "opt-out" strategies. *British Medical Journal, 331*(7572), 940.
- Kaptchuk, T. (2001). The double-blind, randomized placebo-controlled trial: Gold standard or golden calf? *Journal of Clinical Epidemiology, 54*, 541-549.
- Kazdin, A. E. (1982). *Single-case research designs: Methods for clinical and applied settings*. New York: Oxford University Press.
- Kerlinger, F. (1986). *Foundations of behavioral research* (3rd ed.). New York: Holt, Rinehart, & Winston.
- Koes, B. W., Bouter, L. M., & van der Heijden, G. J. (1995). Methodological quality of clinical trials on treatment efficacy in low back pain. *Spine, 20*(2), 228-235.
- Lacher, M. J. (1981). Patients and physicians as obstacles to a randomized clinical trial. *Seminars in Oncology, 8*, 424-429.
- Law, J., Garrett, Z., & Nye, C. (2006). Speech and language therapy interventions for children with primary speech and language delay or disorder. *Cochrane Database of Systematic Reviews, 3*. Available in the Cochrane Library (ISSN 1464-780X).
- Llewellyn-Thomas, H. A., McGreal, M. J., Thiel, E. C., Fine, S., & Erlichman, C. (1991). Patients' willingness to enter clinical trials: Measuring the association with perceived benefit and preference for decision participation. *Social Science and Medicine, 1*, 35-42.
- Lemonick, M. D., & Goldstein, A. (2002). Human guinea pigs. *Time*, April 22. Retrieved April 20, 2002 from timeinc.com
- Levine, R. J. (1987). The apparent incompatibility between informed consent and placebo-controlled clinical trials. *Clinical Pharmacology and Therapeutics, 42*, 247-249.
- Lidz, C. W., Appelbaum, P. S., Grisso, T., & Renaud, M. (2004). Therapeutic misconception and the appreciation of risks in clinical trials. *Social Science and Medicine, 58*, 1689-1697.
- MacIntyre, I. M. C. (1991). Tribulations of clinical trials: Poor recruitments hampering research. *British Medical Journal, 302*, 1099-1100.
- Marcus, S. M. (1997). Assessing non-consent bias with parallel randomized and nonrandomized clinical trials. *Journal of Clinical Epidemiology, 50*, 823-828.
- Mathews, J. N. S. (2000). *An introduction to randomized controlled clinical trials*. London, England: Arnold.
- McKee, M., Gritton, A., Black, N., McPherson, K., Sanderson, C., & Bain, C. (1999). Interpreting the evidence: Choosing between randomized and non-randomized studies. *British Medical Journal, 319*, 312-315.
- Meinert, C. L. (1986). *Clinical trials: Design, conduct, and analysis*. New York: Oxford University Press.
- MIAMI Trial Research Group (1985). Patient population. *American Journal of Cardiology, 56*, 10G-14G.
- Mike, V., & Stanley, K. (1982). *Statistics in medical research*. New York: Wiley.
- Minogue, B. P., Palmer-Fernandez, G., Udel, L., & Waller, B. N. (1995). Individual autonomy and the double-blind controlled experiment: The case of desperate volunteers. *Journal of Medical Philosophy, 20*(1), 43-55.
- Mitchell, J. R. (1981). Trinolol after myocardial infarction: An answer or a new set of questions? *British Medical Journal, 282*, 1565-1570.
- Mook, D. G. (1998). In defense of external validity. In A. E. Kazdin, (Ed.), *Methodological issues and strategies in clinical research* (pp. 145-161). Washington, DC: American Psychological Association.
- Morris, A. D., Zaritsky, A. L., & LeFever, G. (2000). Evaluation of ethical conflicts associated with randomized, controlled trials in critically ill children. *Critical Care Medicine, 28*(4), 1152-1156.
- Moscicki, E. (1993). Fundamental methodological considerations in controlled clinical trials. *Journal of Fluency Disorders, 18*, 183-196.
- Moscucci, M., Byrne, L., Weintraub, M., & Cox, C. (1987). Blinding, unblinding, and the placebo effect: An analysis of patients' guesses of treatment assignment in a double-blind clinical trial. *Clinical Pharmacology and Therapeutics, 41*, 259-265.

- Moser, M. (2006). More confusing message from the hypertension treatment trials. *Journal of Clinical Hypertension*, 8(1), 8-11.
- Mould, R. F. (1998). *Introductory medical statistics* (3rd ed.). Philadelphia: Institute of Physics Publishing.
- Oliver, P. C., Piachaud, J., Done, J., Regan, A., Cooray, S., & Tyrer, P. (2002). Difficulties in conducting randomized controlled trial of health service interventions in intellectual disability: Implications for evidence-based practice. *Journal of Intellectual Disabilities Research*, 46(Pt. 4), 340-345.
- Onslow, M., Andrews, C., & Lincoln, M. (1994). A control/experimental trial of an operant treatment for early stuttering. *Journal of Speech and Hearing Research*, 37, 1244-1259.
- Onslow, M., Packman, A., & Harrison, M. (in press). *Lidcombe program for early stuttering intervention: Clinician's guide*. Austin, TX: Pro-Ed.
- O'Reilly, P., Martin, L., & Collins, G. (1999). Few patients with prostate cancer are willing to be randomized to treatment. *British Medical Journal*, 318(7176), 126
- Paci, E., & Alexander, F. E. (1997). Study designs in randomized controlled trials of breast cancer screening. *Journal of National Cancer Institute*, 22, 21-25.
- Pennington, L., Goldbart, J., & Marshall, J. (2006). Speech and language therapy to improve the communication skills of children cerebral palsy. *Cochrane Database of Systematic Reviews*, 3. Available in the Cochrane Library (ISSN 1464-780X).
- Perlis, R. H., Perlis, C. S., Wu, Y., Hwang, C., Joseph, M., & Nierenberg, A. A. (2005). Industry sponsorship and financial conflict of interest in the reporting of clinical trials in psychiatry. *American Journal of Psychiatry*, 162(10), 1957-1960.
- Place, U. T. (1996). Linguistic behaviorism as a philosophy of empirical science. In W. O'Donohue & R. Kitchener (Eds.), *The philosophy of psychology* (pp. 126-140). Thousand Oaks, CA: Sage Publications.
- Pocock, S. J. (1983). *Clinical trials: A practical approach*. New York: Wiley.
- Post Coronary Artery Bypass Graft Trial Investigators (1997). The effect of aggressive lowering of low-density lipoprotein cholesterol levels and low-dosage anticoagulation on obstructive changes in saphenous-vein coronary-artery bypass grafts. *The New England Journal of Medicine*, 336, 153-163.
- Pringle, M., & Churchill, R. (1995). Randomized controlled trials in general practice: Gold standard or fool's gold? *British Medical Journal*, 311, 1382-1383.
- Pullman, D., & Wang, X. (2001). Adaptive designs, informed consent, and the ethics of research. *Controlled Clinical Trials*, 22(3), 203-210.
- Puma, J. L., Stocking, C. B., Rhoades, W. D., & Darling, C. M. (1995). An ethical debate: Financial ties as part of informed consent to postmarketing research: Attitudes of American doctors and patients. *British Medical Journal*, 310, 1660-1661.
- Ramig, L. O., & Verdolini, K. (1998). Treatment efficacy: Voice disorders. *Journal of Speech, Language, and Hearing Research*, 41, S101-S109.
- Randal, J. (1998). Randomized clinical trials come into their own. *Journal of National Cancer Institute*, 90, 1257-1258.
- Rao, J. N., & Cassia, L. J. S. (2002). Ethics of undisclosed payment to doctors recruiting patients in clinical trials. *British Medical Journal*, 310, 36-37.
- Relman, A. S. (1987). The extracranial-intracranial arterial bypass study: What have we learned? *The New England Journal of Medicine*, 316(13), 809-810.
- Robey, R. R. (1999). Speaking out: Single-subject versus randomized group design. *Asba* (November/December), 14-15
- Rothman, K. J., & Greenland, S. (1998). Causation and causal inference. In K. J. Rothman & S. Greenland (Eds.), *Modern epidemiology* (2nd ed., pp. 7-28). Philadelphia: Lippincott Williams & Wilkins.
- Ruben, D. (1990). *Explaining explanation*. New York: Routledge.
- Scandinavian Simvastatin Survival Study Group. (1994). Randomized trial of cholesterol lowering in 4444 patients with coronary heart disease: The Scandinavian Simvastatin Survival Study. *Lancet*, 344, 1383-1389.
- Schooler, N. (1980). How generalizable are the results of clinical trials? *Psychopharmacology Bulletin*, 16(3), 29-31.
- Schooler, N., Levine, J., & Severe, J. B., Brauzer, B., DiMascio, A., Klerman, G. L., & Tuason, V. B. (1980). Prevention of relapse in schizophrenia: An evaluation of fluphenazine decanoate. *Archives of General Psychiatry*, 37, 16-24.
- Schulz, K. F. (1995). Subverting randomization in controlled trials. *Journal of the American Medical Association*, 274(18), 1456-1458.
- Schulz, K. F., Chalmers, J., Grimes, D. A., & Altman, D. G. (1994). Assessing the quality of randomization from reports of controlled trials published in obstetrics and gynecology journals. *Journal of the American Medical Association*, 272(2), 125-128.
- Schwartz, C. E., & Cox, B. H. (1995). Who says yes? Identifying selection biases in a psychosocial intervention study of multiple sclerosis. *Social Sciences and Medicine*, 40(3), 359-370.
- Sellers, C., Hughes, T., & Langhorne, P. (2006). Speech and language therapy for dysarthria due to nonprogressive brain damage. *Cochrane Database of Systematic Reviews*, 3. Available in the Cochrane Library (ISSN 1464-780X).
- Shapiro, D. A., & Shapiro, D. (1983). Comparative therapy outcome research: Methodological implications of meta-analysis. *Journal of Consulting and Clinical Psychology*, 51, 42-53.
- Shapiro, S., Venet, W., Strax, P., & Venet, L. (Eds.) (1998). *Periodic screening for breast cancer*. Baltimore: Johns Hopkins University Press.
- Shepherd, J., Cobbe, S., Ford, I., Isles, C., Lorimer, R., McFarlane, P., McKillop, J., & Packard, J. (1995). Prevention of coronary heart disease with pravastatin in men with hypercholesterolemia. *New England Journal of Medicine*, 333, 1301-1308.
- Sidman, M. (1960). *Tactics of scientific research*. New York: Basic Books.
- Silverman, W. A., & Altman, D. G. (1996). Patients' preferences and randomized trials. *Lancet*, 347, 171-174.
- Smith, P., & Arnesen, H. (1990). Mortality in nonconsenters in a post-myocardial infarction trial. *Journal of Internal Medicine*, 228, 253-256.

- Snowdon, C., Elbourne, D., & Garcia, J. (1999). Zelen randomization: Attitudes of parents participating in neonatal clinical trial. *Controlled Clinical Trials*, 20(2), 149-171.
- Sundt, T. M., Jr. (1987). Was the international randomized trial of extracranial-intracranial arterial bypass representative of the population? *New England Journal of Medicine*, 316(13), 814-816.
- Taylor, K. M. (1985). The doctor's dilemma: Physician participation in randomized clinical trials. Washington, DC: U.S. Department of Health and Human Services, *Cancer Treatment Report*, 69, 1095-1100.
- Taylor, K. M., Feldstein, M. L., Skeel, R. T., Pandya, K. J., Ng, P., & Carbone, P. P. (1994). Fundamental dilemmas of randomized clinical trials process: Results of a survey of the 1737 Eastern Cooperative Oncology Group Investigators. *Journal of Clinical Oncology*, 12(9), 1796-1805.
- Thomas, C., & Howell, P. (2001). Assessing efficacy of stuttering treatments. *Journal of Fluency Disorders*, 26, 311-333.
- Tobias, J. S., & Houghton, (1994). Is informed consent essential for all chemotherapy studies? *European Journal of Cancer*, 30A, 907-910.
- Tobias, J. S., & Souhami, R. L. (1993). Fully informed consent can be needlessly cruel. *British Medical Journal*, 307, 1199-1201.
- Torgerson, D. J., & Roland, M. (1998). What is Zelen's design? *British Medical Journal*, 316, 606.
- Tuech, J. J., Moutel, G., Pessaux, P., Thoma, V., Schraub, S., & Herve, C. (2005). Disclosure of financial interests and role of sponsors in phase III cancer trials. *European Journal of Cancer*, 41(15), 2237-2240.
- Tygstrup, N., Lachin, J., & Juhl, E. (Eds.) (1982). *The randomized clinical trial and therapeutic decisions*. New York: Marcel Dekker, Inc.
- Verdu-Pascal, F., & Castello-Ponce, A. (2001). Randomized clinical trials: A source of ethical dilemmas. *Medical Ethics*, 27(3), 177-178.
- West, C., Hesketh, A., Vail, A., & Bowen, A. (2006). Interventions for apraxia of speech following stroke. *Cochrane Database of Systematic Reviews*, 3. Available in the Cochrane Library (ISSN 1464-780X).
- Williford, W. O., Kroll, W. F., & Buzby, G. P. (1993). *Journal of Clinical Epidemiology*, 46(9), 1025-1034.
- Wilson, F. (1985). *Explanation, causation, and deduction*. Boston: D. Reidel Publishing Company.
- Yorkston, K. M. (1996). Treatment efficacy: Dysarthria. *Journal of Speech and Hearing Research*, 5, S46-S57.