Continuing Peer Commentary

Response to "Application of Single Subject Randomization Designs to Communication Disorders Research" by Susan Rvachew

Sheila V. Stager
Speech and Voice Unit, Intermural Research Program
NIDCD, NIH, Bethesda, MD

Christine Sloan
School of Human Communication Disorders
Dalhousie University, Halifax, NS

The article by Rvachew (1988) discusses the use of randomization in single-subject design. It is the intent of this response to amplify the benefits of including randomization in the design, and to include a discussion of some ethical and clinical aspects of adding randomization.

A major benefit proposed at the beginning of Rvachew's paper is that the randomization test, a specific statistical analysis, can be applied to the results. She reports that single-subject research is often criticized for incorrect use of parametric, inferential statistical analysis. The use of parametric statistics depends on having both a normal distribution and a standard variance of the data. Randomization of treatments allows the researcher to make these assumptions about the data which could not otherwise be made.

Another rationale for this discussion of statistical methods lies in her differentiation between clinical and statistical significance. Statistical significance is the probability that a particular treatment result was significantly different from a chance occurrence of that result. Clinical significance is the demonstration that treatment results were different from non-treatment results. Neither of these definitions depend on the size of the effect. It is possible to have a relatively small numerical difference between no treatment and treatment results and to still have both statistical and clinical significance. Thus, the distinction between clinical and statistical significance that the author makes is really not a major one and should not obscure the important point that randomization allows the researcher to use an appropriate statistic with single subject designs.

In the next part of the paper, Rvachew discusses the validity concerns of using randomization as part of the design and describes several designs. This is where the strength of using randomization is apparent. In designing any study measuring treatment effects, often termed experimental research, there are two major concerns. The first is to ensure that any change in behaviour results solely from the application of the experimental paradigm, not from some other factor(s). This refers to the internal validity of the study. The second is to design the study so that the treatment effects might generalize across settings, subjects, and behaviours. This refers to the external validity of the study. Single-subject research has, by its nature, poor generalizability. Thus, this discussion will focus on internal validity concerns.

Rival hypotheses for treatment effects include: (1) events occurring during the time of treatment which are unrelated to treatment (these may include intrasession events, especially in treatment which lasts multiple sessions); (2) maturation of subjects during the time treatment is applied, including fatigue and/or practice effects; (3) benefit to the subject from having the same test presented to him/her on more than one occasion; (4) changes in the measuring instrument or changes in the experimenter's criteria during treatment; (5) selection of subjects who are at one end of a performance continuum, rather than all along the continuum, whose performance across time tends towards the mean; (6) selection of subjects who share particular characteristics not included in the criteria for their selection; and (7) the selective loss of subjects (Campbell & Stanley, 1966). In order to reduce the possible influence of these seven factors, Campbell and Stanley (1966) suggest a number of designs which use a combination of a control group and randomization of subjects. The use of a control group in a design will decrease the likelihood that factors such as events occurring outside treatment, maturation, multiple testing, and instrumentation changes would explain any change. These factors may be present because treatment occurs over time. Random assignment of subjects will help ensure that subjects will represent the entire performance continuum, that subjects will not all share one specific characteristic not accounted for in selection criteria, and that subjects with specific characteristics are not selectively lost.

Single-subject designs can be considered as a category of experimental or quasi-experimental research. Thus, the same rival hypotheses that may explain the results of multi-subject designs may also explain the results of single-subject designs. The factors that are controlled for by randomly assigning

subjects do not apply to single subject designs, since the subject is usually selected for a specific purpose and, if he/she is lost, there is no study. However, the factors that are controlled for by the presence of a control group—the factors that arise because treatment takes place over time—are still possible threats to internal validity in single-subject designs. Let's look at a specific example of a design without randomization and one with randomization, in order to determine how these threats to internal validity may be controlled. We will use the example of a multiple baseline across behaviour design. In its truest from, the researcher first selects two behaviours that are independent of one another, each of which will be ameliorated by the same treatment. After a specific baseline time, the researcher applies treatment to one behaviour, while making periodic probes of the other behaviour to ensure it does not change from baseline. Then, after the first behaviour reaches criterion, the researcher applies the treatment to the second behaviour and probes the first to determine if it stays the same or if it declines when treatment is not applied to it. In effect, the researcher has verified or replicated the result of treatment, if it can be shown that treatment also improves the second behaviour.

If randomization is applied to this design, the series of events is as follows. The researcher decides if each segment of therapy (including no therapy) will last a part of a session, whole session, or longer. Then, for the total duration of therapy, the periods of treating one behaviour, the other behaviour, or no behaviour are randomly assigned. The researcher must probe for performance at the beginning and end of each segment. The difference between these pre- and post-probes are plotted across time. Hopefully, there is relatively little difference between the two probes when no treatment is given and some difference between the two probes when the treatment is applied to either one of the behaviours.

In the non-randomization design, internal validity threats are addressed by placing periods of baseline before and after treatment. It is assumed that the treatment is causing the change if there is a large increase/decrease in the behaviour when treatment is applied and no change or only a small decrease/increase when treatment is withheld. However, in the traditional design, the sequential application of no treatment, treatment, and no treatment means that factors such as maturation, instrumentation changes, historical events, and multiple testing are only controlled if we see absolutely no change during the no treatment blocks. If, for example, there is a change in the second behaviour while the researcher is working on the first, then treatment effects may or may not be responsible for this change. If the experimenter continues the study, by treating the second behaviour, he/she is now applying treatment to a behaviour which is not stable. Thus, the experimenter is not really replicating the results. This change in the second behaviour also would mean that one could no longer be certain that it was only the treatment that was causing the first behaviour to change.

The effects of these four threats to internal validity (maturation, instrumentation changes, historical events, and multiple testing) are cumulative, so are likely to be larger near the end of the total duration of therapy. In the randomized design, the treatment and no treatment segments are random in order. It is unlikely, then, that any of these four threats will specifically affect one particular segment and not another. Thus, these temporal factors are less likely to explain any change in behaviours, as Rvachew points out. If there is indeed change during the treatment phase, then a more definite statement can be made that it is really the treatment that is changing behaviours.

Although it is the control over other possible explanations that is the primary advantage to using randomization, there are others. First, randomization eliminates the need to record baseline behaviours. In non-random studies, measures during baseline are required for comparison with measures during treatment. It is the absolute level of performance which is plotted against treatment time. In a randomized study, one compares two measures taken within the same session. It is this relative difference in performance which is plotted against treatment time and not the absolute level of performance. Therefore, a series of baseline measures prior to the experimental procedures are not necessary in the randomized design. Secondly, randomization allows the researcher to make some statement about the relative effectiveness of the treatment on each behaviour. This type of conclusion can't really be made when the treatment is applied sequentially because of the possibility that other factors may have been operative. In the randomized design, these factors are equally likely to apply to each separate behaviour, leaving treatment effects to be the most likely cause of differences in relative effectiveness. Finally, because the treatments are applied randomly, it is unlikely that any order effects of two treatments will explain the results. This is always a concern in sequentially applied treatments. Therefore, the power of applying randomization to a design lies in being able to reduce other explanations for the findings and increase the specificity of how treatment may affect behaviours.

A caveat should be made at this point. Randomization can only be successfully applied to a design that answers the specific question posed by the researcher. It should not be used to bolster up a poorly designed study. This refers to Rvachew's discussion at the end of her paper on how randomization might be applied to improve an AB type design. In Campbell and Stanley (1966), the AB design is termed a one-shot case study and is the weakest design. Since treatment is occurring until the end of the study, with no following baseline condition, there is no control for variables such as maturation, instrumen-

tation changes, historical events, and multiple testing. Rvachew suggests that random assignment of the session which begins the B component will strengthen this design; however, this will not control for these validity factors because the design is still sequential. We would have to question Rvachew's statement that applying randomization to a very weak design, while allowing the use of a test statistic, significantly improves the validity of the result of such a design.

We would also like to address very briefly some ethical and clinical ramifications of doing a randomized design. Ethically, one of the largest problems clinicians face is to justify for parents or caregivers the necessary period of time to complete baseline measurements. Practically, one of the largest problems is to get them to commit the time to come back for the baseline measures following successful conclusion of the treatment phase. Using randomized treatment periods, one could eliminate the need for the prolonged baseline period at both ends of therapy. As well, it would be easier to explain that only one session or part of a session is devoted to no treatment, rather than several successive sessions.

In conclusion, Rvachew has presented a new consideration for the design of single-subject research, the randomization of treatment sessions. As discussed in this commentary and by Rvachew, there are significant advantages to adding this feature: Thus, for a small investment of time in planning, the clinical researcher reaps much benefit.

References

Campbell, D. & Stanley, J., (1966). *Experimental and Quasi-experimental designs for research.* Chicago: Rand McNally & Co.

Rvachew, S., (1988). Application of Single Subject Randomization Designs to Communicative Disorders Research. *Human Communication Canada, 12*, pp. 7-13.

Response to Stager and Sloan

I am gratified to see Stager and Sloan's support for the use of single-subject randomization designs. However, parts of their discussion suggest a need for some clarification of the basic rationale underlying these designs.

The authors state that a baseline phase is not required with the single-subject randomization design because this design relies on within session difference scores rather than a plot of absolute level of performance over time. However, either type of score can be used equally well with both the randomization design and the traditional design. Baseline measures are required when using a traditional single-subject design because interpretation of the results rests upon the assumption that extraneous variables have been eliminated or held constant for the duration of the study. Presumably, a stable baseline allows one to make this assumption.

A stable baseline is not required when using the single-subject randomization design because it is not necessary to assume that extraneous variables have been eliminated or held constant. Rather, it is assumed that the presence of such variables during any given session is independent of the treatment to which that session has been assigned. As long as the assignment of treatments to treatment times is random, this assumption has been met.

The authors also state that randomization does not improve the validity of the AB design. In one sense, this statement is correct; the use of random assignment as described for the "test for treatment intervention" does not improve the AB design; rather, it creates an entirely new design. Despite the superficial similarities, the AB design is non-experimental. Since this randomization design is experimental by virtue of the random assignment procedure, it is not subject to the same threats to internal validity which Campbell and Stanley (1966) attribute to the AB design.

The internal validity of an AB design may be threatened by a number of factors, including: (1) history, (2) maturation, and (3) testing. These factors are a problem to the researcher because they may be systematically associated either with the treatment or the control condition. This is because the B condition is always introduced on a response-contingent, non-random basis. In other words, the treatment is introduced after the subject's performance during the baseline condition has stabilized. Furthermore, introduction of the treatment phase may be confounded with an extraneous variable. For example, it is not unusual for children to show a spontaneous improvement in performance following a "plateau" in performance. Therefore, it is quite possible that the researcher might introduce the treatment just as naturally occurring maturational processes are effecting a positive change in the subject's performance.

The primary purpose of random assignment is to control for differences between subjects as well as differences within subjects over time. Between subject variation in multi-subject research is controlled for by random assignment of subjects to

treatment conditions. For example, if intelligent subjects are systematically assigned to the control condition, the internal validity of the study is seriously compromised and valid application of statistical analysis to the data is impossible. Random assignment of subjects to treatment conditions ensures that subject characteristics such as intelligence are independent of the treatment conditions and allows the researcher to apply a statistical analysis yielding a probability value which indicates the likelihood of the obtained results occurring when the null hypothesis is true. A small probability value increases the researcher's confidence in assuming that the results represent a true treatment effect.

Random assignment of treatment conditions to treatment times controls for within subject variation over time in both multi-subject repeated measures experiments and single-subject randomization experiments. For example, when using the "test for treatment intervention" design, one might assume that practice effects will be greater during session 12 than during session 6. If session 12 was systematically associated with the treatment condition and session 6 was systematically associated with the control condition the study would not be internally valid. Random assignment ensures that the degree of practice effect operating during any given session is independent of the treatment condition assigned to that session. In other words, the practice effects occurring during each session are constant across all possible assignments or data permutations. Under the null hypothesis, the subject's performance during any given session is also independent of the assignment of treatment conditions to treatment times. The randomization test determines the probability of obtaining the observed results when the null hypothesis of no treatment effect is true. Again, small *p* values lead the researcher to conclude that the observed results represent a true treatment effect.

The rationale underlying the use of random assignment and randomization tests with single-subject data is the same for all single-subject randomization designs, including the "test for treatment intervention." As I have noted previously, the particular random assignment used with this design does not compromise its internal validity; however, it does lessen its power relative to the other designs which may be used. This means that temporal effects may lead to small p values because between session differences in performance are due more to such things as practice effects rather than treatment effects. However, the probability of type I error (the probability of erroneously rejecting the null hypothesis) is not greater with this design than with any other single-subject design employing random assignment of treatments to treatment times.

I hope that I have been able to show that the validity of single-subject randomization designs rests upon the random assignment of treatment conditions to treatment times. For a more in depth discussion of the use of random assignment to control for extraneous variation, I refer the reader to Edgington (1984) and Edgington (1987).

S.R.

References

Edgington, E.S. (1984). Statistics and single case analysis. In M. Hersen, R.M. Eisler, & P. Miller (Eds.), *Progress in behavior modification* (pp. 83-119, Vol. 16). New York: Academic Press, Inc.

Edgington, E.S. (1987). *Randomization tests* (2nd edition). New York: Marcel Dekker, Inc.

Campbell, D., & Stanley, J. (1966). *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally & Co.