IDENTIFYING PREDICTORS OF TREATMENT RESPONSE

Paul Yoder[*1] and Don Compton

Vanderbilt University, Nashville, Tennessee

This article provides a rationale for considering predictors of growth in a treatment group as inadequate to identifying predictors of treatment response. When we interpret predictors of growth in a treatment group as synonymous with predictors of treatment response, we implicitly attribute all of the treated children's growth to the treatment, an untenable assumption under most conditions. We also contend that the use of standard scores in predictors of growth studies does not allow us to differentiate growth from treatment, from growth from other factors. We present two research methodologies that are appropriate methods of identifying predictors of treatment response: (a) single-subject experimental logic utilized to identify the specific participants in which treatment responses (not just growth) were found, combined with follow-up group comparison logic to identify the characteristics on which responders and nonresponders differ, and (b) statistical interactions among child/family/context characteristics and randomly assigned group membership. Principles for selecting potential predictors of treatment response are provided. © 2004 Wiley-Liss, Inc.
MRDD Research Reviews 2004;10:162–168.

Key Words: research methodology; predictors of treatment response

Parents and practitioners must select among a large menu of treatment choices for the people with disabilities for whom they are responsible. The efficacy of these treatments often varies as a function of the individual, the family, or the context of treatment. The origin of special education is motivated, in part, by the hypothesis that some individuals need different instructional methods than others. However, the exclusive focus on main effects in most treatment studies may lead to our functionally ignoring those who do not respond to our treatments. Studies that identify predictors of treatment response seek to address these issues.

Unfortunately, few studies appropriately identify predictors of treatment response. In fact, there are several instances of studies that inappropriately interpret predictors of growth as equivalent to or informative of predictors of treatment response. Therefore, there is a need for a review that addresses the appropriate and inappropriate research designs for predicting treatment response. The extant literature on predicting response to treatment [e.g., Cronbach and Snow, 1977] does not deal with the problems inherent in using predictors of growth as informative of predictors of treatment response.

This chapter seeks to address eight points. First, we define *predictors of treatment response* and indicate the two related questions one is trying to address when inquiring about such predictors. Second, we describe how a response to treatment is only a small portion of the growth one observes, even when that growth occurs while being treated and when measured by standard scores. Third, we describe two appropriate approaches to identifying predictors of treatment response. Fourth, we review the rationale for identifying predictors of treatment response. Fifth, we describe the classes of predictors of treatment response. Sixth, we make an appeal for the importance of conducting confirmatory and replication studies on the prediction of treatment response. Seventh, we provide principles for selecting predictors of a target treatment. Eighth, we illustrate that predicting growth addresses different research questions from predicting treatment response. Ultimately, we hope this discussion will encourage researchers and consumers to differentiate predictors of growth from predictors of treatment response, because the former is not necessarily informative of the latter.

When considering which variables are probable predictors of treatment response, it is useful to specify a particular treatment. It is hypothesized here that predictors of treatment response are frequently, perhaps always, specific to the treatment(s) being tested. If this hypothesis is supported over time, it will be clear that it is not realistic to think in terms of "general" predictors of response to any and all treatments. In this chapter, we use our experience with predictors of a particular prelinguistic communication treatment to provide examples of the *types* of variables future researchers may consider as predictors of response to their treatments. This treatment, which we have used with children with either mental retardation or autism, is prelinguistic milieu teaching [PMT; Yoder and Warren, 1998] or its successor, responsive education and prelinguistic milieu teaching [RPMT; Yoder and Warren, 2002]. These child-centered treatments target prelinguistic child communication through responsive interaction, a hierarchy of prompts for target behaviors that are used when children show communicative intent, and informative consequences to child communication. It should be noted that the intensity of treatment implementation may be highly relevant when evaluating the knowledge generated by these studies. In the empirical studies testing PMT

or RPMT, the treatments were implemented in 20-min sessions 3 times a week for 6 months. It is possible that longer or more frequent sessions would result in different predictors of response to PMT or RPMT. It should be noted that this article is not about RPMT, per se. Rather, findings from the PMT and RPMT are used to illustrate the methodological points.

DEFINITION OF PREDICTORS OF TREATMENT RESPONSE

To define the concept of *predictors of treatment response*, it is useful to consider the components of the phrase. *Predictors* are correlating variables. Detecting a *treatment response* requires differentiating change due to the treatment from change due to other factors. The change due solely to the treatment is called the "treatment effect" or "treatment response." Therefore, the concept of predictors of treatment response refers to *correlates of change due exclusively to the treatment*. When scientists attempt to identify predictors of treatment response, they are asking one of two questions: (1) "For which child/family/context is Treatment A effective?" or (2) "For which child/family/context is Treatment A more effective than Treatment B?"

A TREATMENT RESPONSE IS USUALLY A SMALL PORTION OF THE TOTAL GROWTH OBSERVED

Many treatment reviews for special populations have shown that treatments tend to account for less than 50% of the total variance in growth [e.g., Shonkoff and Hauser-Cram, 1987; Yoder and McDuffie, 2002]. To account for over 50% of the variance in total growth the Cohen's *d* for the treatment's effect size (standardized difference between experimental and control group means) would have to be over 3.0. Thus, most of the time, most of the variance in growth is due to factors other than the treatment we are testing. Real growth can occur for factors extraneous to our target treatment for a number of reasons. One important set of reasons is treatments the children experience outside of our treatment studies. These include educational, medical, dietary/supplemental, and prosthetic treatments. Other influences on real change outside of our researched treatments include such factors as mother–child interaction style [Yoder et al., 1998]. Such influences have been called "history effects" [Shadish et al., 2002]. History and other alternative explanations to change have been referred to as "threats to internal validity" [Shadish et al., 2002].

These threats to internal validity affect all types of dependent variables. It is widely recognized that raw scores, age equivalency scores, and criterion-referenced scores change because of factors outside of our target treatments. However, it is important to note that using standard scores, percentile rankings, or z scores (i.e., indices of degree of delay) do not allow us to differentiate change due to our treatment of interest versus other factors that affect real change. We cannot assume that such indices of degree of delay will remain the same over time unless treated by a particular treatment, particularly in children with disabilities. We know this is so because it has been shown that indices of delay can and do change in children in our control groups. For example, Stevenson, Bax, and Stevenson [1982] used a randomized experiment in which children with language impairments were randomly assigned to either a treatment or a control group. Fifty percent of their control group gained at least 1 SD in standard scores on a standardized language test from pretreatment to posttreatment. Therefore, when increases in standard scores occur, we cannot assume that all of that change is because of a particular treatment.

APPROPRIATE METHODS OF IDENTIFYING PREDICTORS OF TREATMENT RESPONSE

If growth in a treatment group is not necessarily a treatment response, then predicting growth in a treatment group is not the same as predicting a treatment response. Methods that accurately identify predictors of treatment response must differentiate change because of a treatment alone versus change at least partially because of other factors. Two research approaches control for threats to internal validity while allowing identification of predictors of treatment response. The first is the use of a series of single-subject experiments to identify treatment responders and nonresponders on an individual basis followed by a group comparison analysis that tests the mean differences on pretreatment variables between responders and nonresponders. The second is a statistical interaction between a pretreatment variable and treatment group assignment in a randomized group design.

Single-Subject Experimental Logic Combined with Between-Group Comparison Logic

Simply put, the first research approach that can legitimately identify predictors of treatment response is a group comparison design that tests the mean difference between treatment responders and treatment nonresponders. It is critical to note that this class of research design uses a single subject experiment to identify "treatment responders" and "treatment nonresponders" on an individual, not a group, basis. Like any set of research designs, there are circumstances under which single subject experiments (particularly AB type designs) do not control for threats to internal validity. However, it should be noted that under proper circumstances, there is a long tradition of appropriate use of single-subject experiments that do control for threats to internal validity [Kazdin, 1982]. Readers are referred to Kazdin [1982] to read about how single-subject experiments control for threats to internal validity, as well as the limitations of such designs.

Treatment responders are those that meet the requirements of the design to differentiate change because of a host of factors from change because of the treatment. Treatment nonresponders may very well make changes in the expected direction but the immediacy and magnitude of the changes are not sufficiently great to attribute them to the targeted treatment. Additionally, it is quite possible for some treatment responders to change less than some nonresponders because only the former's data meet the criteria for inferring a treatment effect by single-subject experimental standards. For example, the responder's change may be immediately after the onset of the treatment and flat baseline, but the nonresponder's change may be quite delayed and after an extremely variable baseline. Unfortunately, there are no published examples of this approach. However, this approach has much promise as a research design that accurately identifies characteristics of children, families, or contexts that respond to a particular treatment as tested by single-subject design. It is also important to note that this design is most appropriate for identifying predictors of response to single treatment rather than predictors of differential response to one treatment over another.

Although the single-subject experimental design is used to differentiate a treatment effect from growth because of other factors, the part of the design that identifies predictors of these treatment responses is a group design. In general,

Table 1. Hypothetical Example of Child Characteristics That May or May Not Predict Treatment Response Tested by a Single-Subject Design

	Participant A	Participant B	Participant C	Participant D	Participant E
Does visual analysis indicate a treatment effect?	Yes	Yes	No	No	Yes
MLU stage	I	II	IV	III	III
IQ	80	76	67	72	71
Type-token ratio	.37	.34	.46	.40	.43

group designs are necessary to identify predictors of treatment response for two related reasons. The effect sizes for conditional treatment effects are likely to be moderate [Cohen, 1988]. For example, the average effect size for the conditional treatment effects in our own past studies is 13% of the variance in the outcome [Yoder and Warren, 1998, 1999, 2001b, 2002]. Power analysis indicates that 56 participants are necessary to detect this size effect with over 80% power. Second, large sample sizes are needed to detect moderate effect sizes because much information is needed to detect a relatively weak association or difference. The unit of analysis for predicting treatment response is the individual case (i.e., participant/family/dyad). No matter how many repeated measures of the dependent or independent variables are present, the "case" contributes only one score in the analysis that identifies predictors of treatment response.

When one has the typical sample size for single subject studies (e.g., three to five people), only one or two violations of the apparent pattern of results will affect our confidence that we have identified a predictor of treatment response. Table 1 shows a hypothetical situation in which the first four participants (A through D) show a pattern of results that may lead readers to think that mean length of utterance (MLU), intelligence quotient (IQ), or type-token ratio (i.e., number of different words/total instances of word use) describes who responded to the treatment. When just one participant breaks the pattern (participant E), our confidence that we have found predictors of treatment response becomes very low. This hypothetical example illustrates why it usually takes many subjects to confidently identify predictors of treatment response.

Randomized Group Experiment with Pretreatment Measures

The second class of research approaches that legitimately identify predic-

tors of treatment response is the randomized group experiment with pretreatment measures of potential predictors of treatment effects. Readers are directed to Shadish et al. [2002] to read in detail how randomized group designs control for threats to internal validity. Briefly, changes due to treatment effects alone are judged by between-group differences in gain scores, growth curves, or endpoint outcomes when groups are initially equivalent. As the simplest case, two groups are compared. These two groups are either a control group and a treatment group or two treatment groups. When two treatment groups are used, it is important that there be no reason to suspect that either treatment could slow development. Otherwise, group differences could be the result of slowed growth in the inferior group and neutral effect in the superior group.

The treatment effect size in a randomized group design is the magnitude of the difference between groups. When the size of this difference varies as a function of a predictor (i.e., child, parent, community variable), a statistical interaction between the predictor and treatment group assignment predicting the endpoint outcome, gain score, or growth curve of the outcome has occurred. We use the term *moderated treatment effect* to describe such statistical interactions. These interactions can be identified in (a) factorial ANOVA designs (e.g., 2 × 2 ANOVA); (b) multiple regressions with a dummy-coded treatment variable, a continuous predictor, and their product term predicting the dependent variable; or (c) the analogous analyses in mixed-level model framework such as hierarchical linear model. Aptitude by treatment interaction (ATI) designs are a small class of moderated treatment effect studies in which at least two treatments are tested and the predictor is a child variable that is essentially unalterable by treatment. However, we do not use this term because it unnecessarily limits our thinking about what constitutes appropriate predictors of treatment response. For

example, alterable child variables or environmental variables can also predict treatment response.

An example of a moderated treatment effect includes our finding of a statistical interaction between presence/absence of Down syndrome and treatment group assignment (RPMT versus control) predicting growth curves of generalized requesting [Yoder and Warren, 2002]. RMPT affected growth in requesting only in children with mental retardation for reasons other than Down syndrome. Although this predictor was dichotomous, many moderators of treatment responses are continuous. For example, pretreatment frequency of commenting, a continuous variable, has been found to moderate RPMT effects on later generalized and maintained comments [Yoder and Warren, 2002]. By "commenting," we mean the use of gestures, vocalizations, and facial expressions to communicate "I like that." Whenever possible, it is desirable to maintain the continuous form of predictors of treatment response to avoid loss of statistical power and to avoid arbitrary or sample-specific decision making regarding where to divide the continuous variable into "low" and "high" groups [Cohen, 1983]. In fact, under some conditions, dichotomizing a continuous variable can also result in type I errors [Maxwell and Delaney, 1993].

RATIONALE FOR CONDUCTING STUDIES TO IDENTIFY PREDICTORS OF TREATMENT RESPONSE

Studies identifying predictors of treatment response define for whom and under what conditions a treatment response is seen. This information can increase (a) the efficiency and effectiveness in our decision making and (b) the efficacy of our treatments.

Two types of decisions about treatment choices are likely to be affected by identifying predictors of treatment response. First, research that identifies pre-

dictors of treatment response could affect personnel training and resource allocation decisions. For example, practitioners could use such research to decide which treatment approaches to learn for use with specific target populations. Second, research that identifies predictors of treatment response can provide provisional guidance regarding the selection of a treatment approach for a particular individual. Although group research designs are particularly well-suited to identifying predictors of treatment response, they provide only probabilistic information for decision making about a particular individual. Any group design finding should be a starting point and only one consideration for individual decision making.

To understand how research on the predictors of treatment response can lead to treatment adaptations, it is important to note that, in some cases, our predictors of treatment response are easily altered by treatment itself, at least for some individuals. In fact, the primary way we can identify whether predictors of treatment response "cause" respondability to a treatment is by comparing the treatment response to (a) a treatment that targets the alterable predictor in addition to the more important outcome versus (b) a treatment that targets only the important outcome.

When we add treatment elements to existing treatments to include alterable predictors of treatment response in the treatment goals, we are "treating the constraint." For example, imagine an oral language treatment that only works for children with relatively good oral motor functioning. An adaptation using the "treat the constraint" approach might be to add a component of the treatment that attempts to address the children's oral motor functioning (e.g., certain feeding programs attempt to address oral motor skills used in speech) before or concurrent with addressing their oral language skills.

An alternative approach to responding to the identification of predictors of treatment response is to create new treatments that "utilize the relative strength" of the children. It is tempting to suggest that such approaches are particularly appropriate when we identify predictors of treatment response that are not easily altered by treatment itself. For example, assume research shows that we are generally not able to address oral motor functioning via educational treatments. Further assume that a hypothetical study finds that oral language treatment works only for children with good oral

motor functioning. A productive response to such a study may be to use nonspeech communication systems for children with poor oral motor functioning. The hypothesis would be that such a treatment would result in better symbolic communication outcomes than an oral language therapy method for the children with the worst oral motor control.

However, the identification of an alterable predictor of treatment response does not necessarily indicate it *should* be treated. Future research is necessary to determine whether it is more efficient to "treat the constraint" versus "utilize the relative strength" and for which subgroup of children. Additionally, there may be a practical limit to the degree to which we can keep adding to the complexity of a particular treatment. For some individuals, the number of constraints may be so great that even if the predictors of treatment response are alterable by treatments, it may be more efficient or practical to utilize the relative strength.

In summary, we have noted that studies identifying predictors of treatment response may guide decision making and motivate treatment adaptation or development. Such studies can guide decision making regarding treatment selection for target populations and can provide provisional guidance on selecting an appropriate treatment for a particular individual. The identification of modifiable predictors can motivate altered treatments that "treat the constraint" by including the predictors in the treatment goals along with the primary outcomes of interest. Finally, identification of less alterable predictors of treatment response or a relatively large number of predictors of treatment response may motivate development of new treatments for those groups of people who do not respond to existing treatments. These new treatments seek to utilize the relative strengths of these people, instead of attempting to treat their constraints. Ideally we would determine relative efficacy of "treat the constraint" versus "utilize the strength" treatments for subgroups that do not respond to previously tested treatments and seek to determine whether the difference was conditional on certain environmental or child factors.

CLASSES OF PREDICTORS OF TREATMENT RESPONSE

Although we usually think of predictors of treatment response as characteristics of children, child variables are only one of three broad classes of predictors of treatment response. These three

classes are child, family, and community variables.

Child Predictors

As mentioned in the previous section, frequency of pretreatment commenting and presence/absence of Down syndrome have been found to predict children's response to RPMT [Yoder and Warren, 2002]. Specifically, children with low-frequency pretreatment commenting learned more generalized and maintained commenting than children in the control group. RPMT did not facilitate commenting in children with high-frequency commenting at the pretreatment period. This finding may simply indicate that children who had the lowest frequency of commenting prior to treatment were most likely to need and thus benefit from what RPMT had to teach. This finding also indicates that commenting, a predictor of response to RPMT, is alterable by educational treatments [Yoder and Warren, 1999, 2002]. Therefore, if comments are the goal and the target children are low in comments before treatment begins, RPMT is a reasonable treatment and requires no alteration. However, children who were high in comments initially may be better off in a different treatment such as linguistic milieu teaching [Yoder and Warren, 2002]. At a certain point in development, most commenting is linguistic and addressing the primary form through which such commenting will take place makes sense.

Down syndrome is clearly not alterable by educational treatment. We found that children with mental retardation without Down syndrome learned generalized and maintained requesting from RPMT, but those children with Down syndrome did not. However, it is important to determine whether there is a variable that covaries with Down syndrome that is responsible for this finding. If future research identifies that a covarying variable of Down syndrome accounts for the result that identified Down syndrome as a predictor of RPMT response, we may find the covarying variable is alterable by treatment itself. If this covarying variable is not alterable by treatment, then we will need to find alternative treatments for children with Down syndrome when targeting requests. Hypothetically, we may find that a passive interaction style and low motivation for difficult tasks covary with Down syndrome and affect learning to request through RPMT. If so, the Picture Exchange Communication System [PECS; Bondy and Frost, 1995], which

uses more explicit reinforcement sampling and a different approach to teaching self-initiation, may be effective in facilitating generalized requests for children with Down syndrome. This would be an example of "utilize the relative strength" philosophy to treatment selection. This hypothesis has not yet been tested.

Parental Predictors

Parent variables can also predict treatment response. For example, the proportion of communication acts to which parents responded prior to treatment predicted children's response to PMT on later requesting, commenting, and language [Yoder and Warren, 1999; Yoder and Warren, 1998, 2001b]. This was true even though staff, not parents, implemented the treatment. Additionally, parents were not even allowed to view the treatment in this research. Theory suggests that parental responsivity could reinforce and thus support generalization and maintenance of changes induced by and first occurring during PMT treatment sessions. Additionally; subsequent findings support the notion that the direct effects of PMT (e.g., increased intentional communication) elicited parental responding, which in turn may have facilitated later language [Yoder and Warren, 2001a]. Because these findings and theory suggested that parental responsivity could support children's response to PMT, the authors of PMT added a component of the treatment to support parental responsivity, thus creating responsivity and prelinguistic milieu teaching [RPMT; Yoder and Warren, 2002]. This is an example of "treat the weakness" philosophy of treatment adaptation. However, not all parental predictors of treatment response are reasonably accommodated by treatment adaptations. For example, maternal education level also predicted response to PMT [Yoder and Warren, 1999]. It is probable that maternal education level is merely a covarying variable for some more proximal predictor of response to PMT (e.g., maternal interaction style). This covarying variable has not yet been identified.

Community Predictors

Community variables are the least studied class of potential predictors of treatment response. One question regarding community variables that has particular salience for children with disabilities is how nonproject treatments affect children's response to within-project treatments. Nonproject treatments may have either synergistic or inhibiting effects on children's response to a within-

project treatment. For example, Yoder and Stone [2000] are currently testing whether number of hours of discrete trial training experienced outside of our project affects children's response to RPMT, a within-project child-led treatment, with children with autism. This is particularly important for children with autism because some receive up to 30 or 40 hours of discrete trial training weekly and many have specific deficits in the type of social communication for which child-led treatments are particularly designed (i.e., initiating joint attention). One can imagine that many hours of being taught to be compliant to adult prompts would lead children to be passive in interactions with adults. This passivity could generalize to interactions with practitioners using RPMT, which depends on child initiation of communication to teach. This hypothesis has not yet been tested.

In summary, child, familial, and community variables are potential predictors of treatment response in any one sample. Even when these predictors are identified in the context of exploratory studies (i.e., many significance tests without appropriate protection against experiment-wise error), the results are valid for that sample. However, only replicable results have societal value.

REPLICATION IS KEY

For the value of identifying predictors of treatment response to be realized, it is critical to replicate the findings of well-conducted and appropriately interpreted studies that identify predictors of treatment response. We need replication of existing moderated treatment effects and replication of the difference between responders and nonresponders, as detected in single-subject experiments. Unfortunately, there has been a general lack of replication of statistical interactions [Cronbach and Snow, 1977]. This lack of replication probably occurs, in part, because many presently identified statistical interactions may have been found through multiple significance testing without adjustment of the alpha for multiple testing. This practice results in inflated experiment-wise error [i.e.; finding a significant result when one is not present in the population; Westfall and Young, 1992]. Confirmatory studies that test a few theoretically predictable correlates of treatment response are more likely to identify replicable results than are exploratory studies. Therefore, it is useful to review principles by which potential predictors of treatment response

can be selected before beginning the study.

PRINCIPLES FOR SELECTING POTENTIAL PREDICTORS OF TREATMENT RESPONSE

Even though predictors of treatment response are not necessarily causal explanations for treatment response, thinking in causal terms helps identify such predictors a priori. It is much more difficult to select predictors of differential response to two or more treatments (e.g., Treatment A versus Treatment B) than it is to select predictors of response to a particular treatment (i.e., Treatment A versus Control group). Many predictors of response to a particular treatment are predictors of response to other treatments as well. For example, theory suggests that children with autism with extremely low object exploration will respond less well to child-centered treatments (e.g., RPMT) than will children who readily play with a variety of objects. However, theory suggests that such children will also have difficulty acquiring generalized communication skills from other, more adult-directed communication treatments that do not directly and effectively address object exploration and play skills (e.g., PECS). If future research supports these predictions, frequency of object exploration would *not* be expected to be a predictor of *differential* response to one of these communication treatments over the other.

In general, child skills, familial contexts, or community treatment conditions that one treatment requires or benefits from but that the other does not tend to predict differential response to one treatment over another. In the case of prerequisite child skills, predicting differential response to two or more treatments can occur because the superior treatment successfully addresses the prerequisite child skills before or while treating other skills (i.e., treat the constraint) or circumvents the need for otherwise prerequisite child skill (i.e., utilize the relative strength); whereas the other requires but does not directly treat the child skill. In the case of familial context, predicting differential response to two or more treatments can occur because the superior treatment provides the stimulation that the familial context lacks while the inferior treatment does not provide, but depends on, the needed stimulation. In the case of community treatments, predicting differential response to two or more treatments can occur because the superior within-project treatment is compatible with the dominant commu-

nity treatment, but the inferior within-project treatment is not. Again, the principles for selecting potential predictors of treatment response do not make reference to variables that predict growth of the outcome in general.

The following principles can be used to guide selection of predictors of response to one particular treatment (as opposed to differential response to one treatment over another). Predictors of response to a particular treatment tend to be variables in one of four classes as follows: (a) the child's need for what the treatment teaches (e.g., low commenting for commenting treatment), (b) the child's skills that enable complete and consistent implementation of the treatment (e.g., compliance), (c) the child's skills that are prerequisite for the child to benefit from the treatment (e.g., good speech processing for a linguistic treatment), and (d) the family context or community-based treatments that support development of (b) and (c) (e.g., consistency or responsivity). Children who score low on variables in Class A are predicted to benefit from the treatment. Children who score high on variables in Classes B–D are predicted to benefit from the treatment.

It should be noted that utilizing these principles to select potential predictors of response to a particular treatment requires a firm grasp of the treatment and theory for why the treatment should be effective. None of the principles makes reference to variables that predict growth on the outcome. However, because relatively high scorers on some of the types of predictors of treatment response (i.e., those in Classes B–D) and almost all types of predictors of growth are predicted to have "positive outcomes," it is tempting to think that predictors of growth are informative in selecting predictors of treatment response. After all, the rich do get richer. The problem with this truism is that it is too general to be of much scientific value. The *particular* variables on which one is rich and on which variables one gets richer are critically important. This is not to say that predictors of growth *never* predict treatment response. Predictors of treatment response can be predictors of growth, by chance. But using the results of predictors of growth studies to select predictors of treatment response is generally not a useful strategy. Next, we discuss why this is so.

INEFFECTIVE METHODS OF IDENTIFYING PREDICTORS OF TREATMENT RESPONSE

It is important to contrast the two appropriate methods for identifying predictors of treatment response with commonly seen, but misguided, approaches to doing so. These misguided methods have in common the prediction of presence of or amount of change, not treatment response. Here, we include change scores as measures of growth. It should be emphasized that predicting growth is clearly an important scientific goal. It is the misinterpretation of such results as identifying predictors of treatment response to which we are objecting.

Predictors of Growth within a Treatment Group

The most common misguided approach to identifying predictors of treatment response is to identify predictors of gain on the outcome within the experimental group. A variant of this approach replaces gain scores with parameters of growth curves or endpoint outcome scores controlling for initial outcome scores. A third variant is to dichotomize the outcome within the treatment group into those with favorable outcomes versus those with less favorable outcomes. A fourth variant is to dichotomize the outcome within the treatment group into those that change versus those that do not change. This latter approach is usually conducted with standard scores. The futility of using standard scores to differentiate change due to other factors from change due to a particular treatment has already been addressed, as have the problems with dichotomizing a normally continuous outcome.

When it comes to predicting treatment response, the problem with studies that predict growth (or outcome level) of the experimental group is that individual differences in the gain (or outcome) is caused by many sources, not just the target treatment(s). This problem exists even when prior analyses have demonstrated experimental versus control group differences on the dependent variable, including on standard scores. Significant between-group mean differences do not indicate that all of the change in the treatment group is due to the treatment, only some of it. Some readers may wonder why it is informative to use single subject experiments, but not group experiments, prior to a subsequent analysis thought to identify predictors of treatment response. The difference between the approaches is that group design logic detects a treatment effect for a group, not individuals. Therefore, the subsequent prediction within the experimental group can no longer differentiate change due to the treatment versus that due to many factors. Such within-treatment-group analyses implicitly treat all growth as "treatment effect," an untenable assumption in most cases. In contrast, single subject experiments can potentially allow identification of a treatment effect at the individual level. Every member of the treatment responders group has been found to show change that is greater than expected due to other influences on change. The subsequent analysis in the "single-subject experimental logic combined with between-group comparison logic" approach seeks to identify the characteristics on which treatment responders are different from treatment nonresponders. No slip in logic has occurred in the latter approach.

An example from our own data will demonstrate that predicting gain scores in the treatment group answers a different research question than do analyses identifying moderated treatment effects. Using unpublished results from the Yoder and Warren [2003] sample, the R^2 for the association of rate of Time 1 intentional communication and gain in the rate of commenting from Time 1 to Time 2 in the RPMT group was -0.004; $P = 0.99$. The R^2 for the statistical interaction between rate of Time 1 intentional communication and treatment group predicting rate of Time 2 comments above and beyond main effects was 0.10; $p = 0.04$. One explanation for the difference in the findings is that the treatment changed the association of intentional communication and comments.

More commonly, we find variables that predict growth within the treatment group that do not interact with the treatment group. For example, frequency of requests at Time 1 predicted number of different words 6 months after treatment ended within the treatment group ($r = 0.60$; $P = 0.008$). However, frequency of requests did not statistically interact with treatment group assignment to predict follow-up number of different words [unpublished data from the Yoder and Warren, 2002].

Predictors of Growth Regardless of Treatment Experience

A second misguided approach is to view correlates of outcome or gain regardless of treatment experience as informative of what predicts treatment response. The difference between this approach and the first of the misguided approaches is that the analysis pools participants that vary on experimental versus control group membership on the type and intensity of treatment they experience. As with the above design, growth is implicitly equated with treatment re-

sponse when one interprets such results as identifiers of response to a treatment.

Data from our lab illustrates that this last misguided approach does not necessarily inform us about the types of children that benefit from a treatment. A correlational analysis on the pooled sample (i.e., across control and experimental groups) indicated that children with relatively *high* frequency vocal communication with consonants prior to treatment had better language developmental rates than children with relatively low frequencies of such vocal communication [Yoder and Warren, 2003]. A moderated treatment effect analysis showed that RPMT facilitated maintained language development in children with *low* frequency vocal communication with consonants prior to treatment, but not in children with high frequency vocal communication with consonants [Yoder and Warren, 2002]. If we had used the information regarding predicting rate of gain to inform us regarding which children should receive RPMT, the wrong subgroup, but right variable, would have been identified. However, there are even more examples of predictors of growth in a pooled sample analysis that do not interact with treatment group assignment. For example, Yoder and Warren [2003] found that number of parental facilitating responses predicted later language in children with mental retardation, but this variable did not interact with the treatment group assignment to predict later language.

CONCLUSION

By its nature, predicting treatment response requires conceptually testing a conditional treatment effect. That is, the treatment works for some people or contexts but not others. The only way to know whether a treatment works is to differentiate total change from change due only to the treatment (i.e., a treatment effect). The single subject experimental logic combined with between-group comparison logic approach to identifying predictors of treatment re-

sponse examines differences between people or contexts in which treatment works versus those for which it does not work. Moderated treatment effects test this conditional treatment effect through a statistical interaction between characteristics of the participant or his context and treatment group assignment. We have suggested that replicating conditional treatment effects is necessary to know whether they generalize to an appropriate population. Such replication studies are very expensive. To identify enough participants in a particular clinical population, we may very well need to conduct collaborative multisite research. However, the importance of identifying predictors of treatment response justifies the expense.

At the very least, we hope that the present discussion will lead researchers and consumers of research to show appropriate restraint in interpreting study results. Specifically, we hope it is clear that predictors of growth are not equivalent to, and are not necessarily informative of, identifying predictors of treatment response. This is true even when that growth occurs within a treatment group, when growth is measured with indices of degree of delay such as standard scores, and when prior group analyses find evidence of a main effect for the treatment on the dependent variable. ∎

REFERENCES

Bondy AS, Frost LA. 1995. Educational approaches in preschool: behavior techniques in a public school setting. In: Schopler E, Mesibov GB, editors. Learning and cognition in autism. New York: Plenum. p 311–333.

Cohen J. 1983. The cost of dichotomization. Applied Psychological Measurement, 7(3):249–253.

Cohen J. 1988. Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

Cronbach L, Snow R. 1977. Aptitudes and instructional methods. New York: Irvington.

Kazdin A. 1982. Single-case research designs: methods for clinical and applied settings. New York: Oxford University.

Maxwell S, Delaney H. 1993. Bivariate median splits and spurious statistical significance. Psychological Bulletin 113:181–190.

Shadish W, Cook T, Campbell D. 2002. Experimental and quasi-experimental designs for generalized causal inference. Boston: Houghton-Mifflin.

Shonkoff JP, Hauser-Cram P. 1987. Early intervention for disabled infants and their families: a quantitative analysis. Pediatrics 80:650–658.

Stevenson P, Bax M, Stevenson J. 1982. The evaluation of home based speech therapy for language delayed pre-school children in an inner city area. British Journal of Disorders of Communication 1:141–148.

Westfall PH, Young SS. 1992. Resampling-based multiple testing: examples and methods for P-value adjustment. New York: Wiley.

Yoder PJ, McDuffie A. 2002. Treatment of primary language disorders in early childhood: evidence of efficacy. In Accardo P, editors, Disorders of language development. Baltimore: York. p 151–177.

Yoder P, Stone W. 2000. Early communication intervention for children with autism. NIDCD grant application, Nashville, TN.

Yoder P, Warren SF. 1998. Maternal responsivity predicts the prelinguistic communication intervention that facilitates generalized intentional communication. Journal of Speech, Language, and Hearing Research 41(5):1207–1219.

Yoder P, Warren SF. 1999. Self-initiated proto-declaratives and proto-imperatives can be facilitated in prelinguistic children with developmental disabilities. Journal of Early Intervention 22:337–354.

Yoder PJ, Warren SF. 2001a. Intentional communication elicits language-facilitating maternal responses in dyads with children who have developmental disabilities. American Journal on Ment Retard, 106(4):327–335.

Yoder PJ, Warren SF. 2001b. Relative treatment effects of two prelinguistic communication interventions on language development in toddlers with developmental delays vary by maternal characteristics. Journal of Speech, Language, and Hearing Research, 44(1):224–237.

Yoder P, Warren SF. 2002. Effects of prelinguistic milieu teaching and parent responsivity education in dyads with children with intellectual disabilities. Journal of Speech, Language, and Hearing Research 45:1158–1174.

Yoder P, Warren SF. 2003. Early predictors of language in children with and without Down syndrome. American Journal of Mental Deficiency. In press.

Yoder PJ, Warren SF, McCathren R, Leew SV. 1998. Does adult responsivity to child behavior facilitate communication development? In Wetherby AM, Warren SF, Reichle J, editions. Transitions in prelinguistic communication. Baltimore: Brookes. Vol. 7, p 39–58.