Experimental Design and Primary Data Analysis Methods for Comparing Adaptive Interventions

Inbal Nahum-Shani
University of Michigan

Min Qian
Columbia University

Daniel Almirall
University of Michigan

William E. Pelham and Beth Gnagy
Florida International University

Gregory A. Fabiano
University at Buffalo, State University of New York

James G. Waxmonsky
Florida International University

Jihnhee Yu
University at Buffalo, State University of New York

Susan A. Murphy
University of Michigan

In recent years, research in the area of intervention development has been shifting from the traditional fixed-intervention approach to *adaptive interventions*, which allow greater individualization and adaptation of intervention options (i.e., intervention type and/or dosage) over time. Adaptive interventions are operationalized via a sequence of decision rules that specify how intervention options should be adapted to an individual's characteristics and changing needs, with the general aim to optimize the long-term effectiveness of the intervention. Here, we review adaptive interventions, discussing the potential contribution of this concept to research in the behavioral and social sciences. We then propose the sequential multiple assignment randomized trial (SMART), an experimental design useful for addressing research questions that inform the construction of high-quality adaptive interventions. To clarify the SMART approach and its advantages, we compare SMART with other experimental approaches. We also provide methods for analyzing data from SMART to address primary research questions that inform the construction of a high-quality adaptive intervention.

Keywords: adaptive interventions, experimental design, sequential multiple assignment randomized trial (SMART)

In adaptive interventions the type or the dosage of the intervention offered to participants is *individualized* based on participants' characteristics or clinical presentation and then *repeatedly adjusted* over time in response to their ongoing performance (see, e.g., Bierman, Nix, Maples, & Murphy, 2006; Marlowe et al., 2008; McKay, 2005). This approach is based on the notion that individuals differ in their responses to interventions: In order for an intervention to be most effective, it should be individualized and, over time, repeatedly adapted to individ-

ual progress. An *adaptive intervention* is a multistage process that adapts to the dynamics of the "system" of interest (e.g., individuals, couples, families, or organizations) via a sequence of decision rules that recommend when and how the intervention should be modified in order to maximize long-term primary outcomes. These recommendations are based not only on the participant's characteristics but also on intermediate outcomes collected during the intervention, such as the participant's response and adherence. Adaptive interventions are also known as

dynamic treatment regimes (Murphy, van der Laan, Robins, & the Conduct Problems Prevention Research Group, 2001; Robins, 1986), *adaptive treatment strategies* (Lavori & Dawson, 2000; Murphy, 2005), *multistage treatment strategies* (Thall, Sung, & Estey, 2002; Thall & Wathen, 2005), and *treatment policies* (Lunceford, Davidian, & Tsiatis, 2002; Wahed & Tsiatis, 2004, 2006).

The conceptual advantages of adaptive interventions over a fixed-intervention approach (in which all participants are offered the same type or dosage of the intervention) have long been recognized by behavioral and social scientists. For example, in the area of organizational behavior, Martocchio and Webster (1992) studied the effect of feedback on employee performance in software training, stressing the need to develop training programs in which training design characteristics are adapted to employee level of cognitive playfulness (cognitive spontaneity in human–computer interactions). Recently, in their study of career goal setting, Hirschi and Vondracek (2009) discussed the development and adaptation of goals as a dynamic process, in which individuals have to select goals according to personal preferences and environmental opportunities and limitations, optimize their behavior to achieve those goals, and compensate and adjust if goals become unattainable or unattractive (also see Abele & Wiese, 2008; Baltes & Baltes, 1990). In the area of psychotherapy, Laurenceau, Hayes, and Feldman (2007) conceptualized individuals' growth and change during therapy as a dynamic (i.e., nonlinear and discontinuous) process and highlighted the need to use experimental designs that improve the ability of practitioners to effectively adjust therapy over time in response to the dynamics of change (Laurenceau et al., 2007).

Here, we discuss an experimental design and data analysis methods useful for developing (i.e., for constructing or revising) adaptive interventions in the social and behavioral sciences. First we review how adaptive interventions use decision rules to operationalize the individualization of intervention options (i.e., intervention type and dosage) based on participants' characteristics and the repeated adaptation of intervention options over time in response to the ongoing performance of participants. Next we discuss how the sequential multiple assignment randomized trial (SMART; Murphy, 2005; Murphy, Collins, & Rush, 2007; see also Lavori & Dawson, 2000, 2004, for related ideas), an experimental design approach used across the medical fields (e.g., Fava et al., 2003; Lavori et al., 2001; Schneider et al., 2001; Stone et al., 1995), can be used to develop adaptive interventions in the social and behavioral sciences. More specifically, we discuss how data from SMART designs are useful in addressing key research questions that inform the construction of a high-quality adaptive intervention. Building on ideas from the clinical literature, we compare the SMART design with other common experimental approaches. We also provide data analysis methods that can be used with the resulting SMART data to compare the effectiveness of intervention options at different stages of the adaptive intervention and to compare the effectiveness of adaptive interventions that are embedded in the SMART design. To clarify ideas and illustrate the data analysis methods, we employ data from the Adaptive Interventions for Children with ADHD SMART study (Center for Children and Families, SUNY at Buffalo, William E. Pelham, principal investigator [PI]). Finally, we discuss

directions for future research for social and behavioral scientists aiming to develop adaptive interventions.

Adaptive Interventions

Adaptive interventions are conceptualized as sequential treatment processes in which the varying needs of participants are taken into consideration (Collins, Murphy, & Bierman, 2004). This conceptualization has two components: (a) The intervention is *individualized* based on the characteristics and specific needs of participants; and (b) the intervention is *time varying*; that is, it repeatedly adapts over time in response to participants' ongoing performance and changing needs.

Sackett, Rosenberg, Gray, Haynes, and Richardson (1996, p. 71) defined evidence-based practice as "the conscientious, explicit, and judicious use of current best practice evidence in making decisions about the care of individual patients." The practical appeal of the adaptive approach is outlined in Weisz, Chu, and Polo's (2004, p. 302) discussion of dissemination and evidence-based practice in clinical psychology, which suggests that evidence-based practice should ideally consist of much more than simply obtaining an initial diagnosis and choosing a matching treatment. Evidence-based practice

> is not a specific treatment or a set of treatments, but rather an orientation or a value system that relies on evidence to guide the entire treatment process. Thus, a critical element of evidence-based care is periodic assessment to gauge whether the treatment selected initially is in fact proving helpful. If it is not, adjustments in procedures will be necessary, perhaps several times over the course of the treatment. (Weisz et al., 2004, p. 302)

In fact, many behavioral and cognitive therapies can be seen as adaptive interventions (Bierman et al., 2006). For example, group therapy processes are adjusted over time based on group dynamics and the developmental stage of the group (M. B. Cole, 2005; Yalom, 1995). Cognitive therapy is tailored to address the unique cognitive conceptualization of the patient and to respond to his/her progress or regression during the process, with therapists basing the format, content, duration, and intensity of the upcoming sessions on the results of the prior sessions (Beck, Liese, & Najavits, 2005). While in practice the idea of adaptive interventions is not new, limited attention was given to developing research designs and methods that help scientists and practitioners decide how to individualize and repeatedly adapt intervention options (i.e., intervention type and/or dosage) in order to optimize long-term outcome. In recent years, intervention scientists are becoming increasingly interested in experimental designs that explicitly incorporate the adaptive aspects of the intervention. The general aim is to obtain data useful for addressing key questions concerning the individualization and adaptation of intervention options so as to inform the construction of high-quality (i.e., highly efficacious) adaptive interventions.

Decision Rules and Tailoring Variables

One approach to operationalizing the conceptual idea of an adaptive intervention is to use decision rules (Bierman et al., 2006) that link individuals' characteristics and ongoing performance with specific intervention options. The aim of these decision rules is to

guide practitioners in deciding which intervention options to use at each stage of the adaptive intervention, using available information relating to the characteristics and/or ongoing performance of the participant. This approach is conceptually appealing because it mimics decision processes in real life, in which decision makers select their actions based on information obtained from the environment and modify their actions based on this information with the general aim to achieve good long-term outcomes.

In an adaptive intervention, the adaptation of the intervention options is based on the participant's values on *tailoring variables*. These variables strongly moderate the effect of intervention options, such that the type or dosage of the intervention should be tailored according to values of these variables. Not all moderators are tailoring variables (Gunter, Zhu, & Murphy, 2011). Consider an example in which both men and women respond better to high-intensity than to low-intensity behavioral intervention, but high-intensity behavioral intervention is more beneficial for women. In this case, although gender is a moderator, both men and women should be offered high-intensity behavioral intervention because both groups are likely to benefit. Thus, gender is not a tailoring variable (otherwise different genders should be offered different levels of intervention intensity).

Although the list of candidate tailoring variables depends on the problem at hand, potential tailoring variables include individual, group, or contextual characteristics representing risk or protective factors that influence responsiveness to (or need for) intervention type or dosage. For example, assume an investigator is interested in evaluating an intervention (i.e., assessing the effectiveness of the intervention compared to control) that includes 13 weekly group sessions for improving perceived social support among community residents. Suppose the investigator considers two types of intervention: one in which group sessions emphasize cognitive restructuring (e.g., positive assertions to self and others, conflict resolution strategies, active listening) and one in which group sessions emphasize developing social skills (e.g., identifying and correcting dysfunctional attitudes in relationships, positive self-statements, and self-acceptance). Assume that prior scientific evidence suggests that community residents who are characterized by particular risk factors, for example low self-esteem, are more likely to benefit from an intervention that emphasizes cognitive restructuring than on social skills training. Those with high self-esteem, on the other hand, might find other interventions more beneficial. Accordingly, the investigator might choose to offer a supportive intervention in which the type of the intervention is individualized based on baseline levels of self-esteem (i.e., to offer different types of intervention to community residents with different baseline levels of self-esteem).

Participants' responsiveness might also be an important tailoring variable. For example, assume the investigator conjectures that community residents who do not adequately respond (report low levels of support) to the support intervention within a certain period of time (e.g., 6 weeks) might need a more intensive subsequent intervention (e.g., during the last 7 weeks of the intervention), rather than more of the same. In this case, the investigator might choose to offer a supportive intervention that adapts to community residents' response to the first 6 weeks of the intervention.

The choice of intervention options can also be tailored by the intervention options the individual has already received. For ex-

ample, assume the investigator conjectures that intensifying the intervention for nonresponders might be more effective for those residents who received cognitive restructuring during the first 6 weeks of the intervention. However, it is better to add a cognitive component to those nonresponders who received social skills training during the first 6 weeks of the intervention. In this case, the investigator might choose to offer a supportive intervention in which the decision of whether to intensify the support intervention for nonresponders to the first 6 weeks of the intervention varies depending on the type of support intervention given during the first 6 weeks.

To demonstrate the way decision rules link the tailoring variables to intervention options, assume the investigator decided to offer a supportive intervention that begins with six weekly sessions of social skills training (denoted as the first stage of the intervention) and then the sessions offered during the subsequent 7 weeks (denoted as the second stage of the intervention) are tailored according to the participant's response to the first stage of the intervention. In this case, the decision rule can be expressed as

First-stage intervention = {social skill development}
  IF evaluation = {nonresponse}
    THEN second-stage intervention = {intensify first-stage intervention}
  ELSE IF evaluation = {response}
    THEN at second stage = {continue on present intervention}

Notice that the IF and ELSE IF parts of the rule contain the tailoring variables; intervention options are expressed in the THEN parts of this rule. To develop high-quality adaptive interventions, we focus on constructing good decision rules.

Researchers have used decision rules to operationalize adaptive interventions in several existing programs. In the Fast Track adaptive intervention for preventing conduct problems among high-risk children (Conduct Problems Prevention Research Group, 1992), the frequency of home visits was adapted, and at half-year intervals, re-adapted. The tailoring variables were two types of clinical judgment: ratings of parental functioning and global assessments of family need for services. Program guidelines were used to link these two tailoring variables to three prescribed levels of home visiting (e.g., weekly home visits for low-functioning families, biweekly home visits for moderate-functioning families, and monthly home visits for high-functioning families (see Bierman et al., 2006). In the misdemeanor drug court adaptive intervention developed by Marlowe and colleagues (Marlowe et al., 2008), the frequency of court hearings and type of counseling session were adapted using pretreatment tailoring variables (history of formal drug abuse treatment and evidence of antisocial personality disorder) as well as time-varying tailoring variables (participants' compliance with the drug court program and results of drug tests). McKay (2005) evaluated the effectiveness of an adaptive intervention for patients who are alcohol dependent. This intervention was built around brief telephone contacts, consisting of risk-for-relapse assessment and problem-focused counseling. When risk levels increased above specified criteria, participants received stepped-up care, such as more frequent telephone sessions and sessions of motivational interviewing.

In these programs, the sequence of decision rules that operationalized the adaptive intervention was prespecified based on clinical

experience and scientific evidence. However, in many cases, there is not enough evidence a priori to inform the construction of high-quality sequences of decision rules. There might be open questions (or even debates in the literature) concerning the best intervention option at specific intervention stages or other important questions relating to which tailoring variables to use. For example, consider the previous example concerning the group-based intervention for improving the social support of community residents. Suppose the investigator is interested in developing a two-stage adaptive intervention but there is not enough scientific evidence to determine (a) whether the group sessions at the first stage of the intervention should focus on cognitive restructuring or social skills development and (b) what is the most efficacious intervention option for participants who do not respond to the first six sessions. In this case, the investigator might be interested in obtaining data that inform the development of an adaptive intervention prior to its evaluation (i.e., assessing the effectiveness of the intervention compared to control). More specifically, the investigator might be interested in collecting data that enable him/her to compare the effectiveness of various intervention options at different stages of the intervention, as well as to compare different sequences of decision rules (i.e., different adaptive interventions). To better understand this, consider the following example, based on the Adaptive Interventions for Children With ADHD study (Center for Children and Families, SUNY at Buffalo, William E. Pelham, PI).

Adaptive Interventions for Children With ADHD Example

Suppose an investigator is interested in planning an adaptive two-stage intervention to improve the school performance of children with attention-deficit/hyperactivity disorder (ADHD). There are two critical decisions confronting the development of this adaptive intervention: Which intervention option should be provided initially? and Which intervention option should be provided to children whose ADHD symptoms did not sufficiently decrease

during the first-stage intervention (i.e., nonresponders)? First, the researcher considers two possible first-stage intervention options: low-intensity behavioral intervention or low-dose medication. Second, the researcher considers two possible second-stage intervention options for the nonresponding children: to increase the dose of the initial intervention or to add the alternative type of intervention. More specifically, if the child is not responding to low dose of medication, the dose of medication can be increased or medication can be augmented with behavioral intervention. If the child is not responding to behavioral intervention, the intensity of the behavioral intervention can be enhanced or the behavioral intervention can be augmented with medication. Table 1 presents four simple adaptive interventions and the four associated decision rules drawn from this example.

Assuming that an investigator considers the four adaptive interventions presented in Table 1, the question is how to obtain high-quality data that can be used to contrast these four adaptive interventions and that, beyond this simple comparison, more deeply inform the construction of an adaptive intervention for children with ADHD.

Sequential Multiple Assignment Randomized Trial

A SMART is a multistage randomized trial in which each stage corresponds to a critical decision. Each participant progresses through the stages and is randomly assigned to one of several intervention options at each stage (Murphy, 2005). The SMART was designed specifically to aid in the development of adaptive interventions. Data from a SMART design can be used to address primary research questions concerning (a) the comparison of different intervention options at different stages of the intervention (e.g., the difference between the first-stage intervention options or the difference between the second-stage intervention options for nonresponding participants); (b) the comparison of adaptive interventions (i.e., sequences of decision rules) that are embedded in

Table 1
Four Adaptive Interventions and Decision Rules Based on the ADHD Example

Adaptive intervention	Decision rule
(1, −1) First, offer low-intensity behavioral intervention; then add medication for nonresponders and continue low-intensity behavioral intervention for responders.	First-stage intervention option = {BMOD} IF evaluation = {nonresponse} THEN second-stage intervention option = {AUGMENT} ELSE continue on first-stage intervention option
(−1, −1) First, offer low-dose medication; then add behavioral intervention for nonresponders and continue low-dose medication for responders.	First-stage intervention option = {MED} IF evaluation = {nonresponse} THEN second-stage intervention option = {AUGMENT} ELSE continue on first-stage intervention option
(1, 1) First, offer low-intensity behavioral intervention; then increase the intensity of behavioral intervention for nonresponders and continue low-intensity behavioral intervention for responders.	First-stage intervention option = {BMOD} IF evaluation = {nonresponse} THEN second-stage intervention option = {INTENSIFY} ELSE continue on first-stage intervention option
(−1, 1) First, offer low-dose medication; then increase the dose of medication for nonresponders and continue low-dose of medication for responders.	First-stage intervention option = {MED} IF evaluation = {nonresponse} THEN second-stage intervention option = {INTENSIFY} ELSE continue on first-stage intervention option

Note. The notation (a, b) in column 1 is used here and throughout the article as shorthand for the different adaptive interventions embedded within the sequential multiple assignment randomized trial (SMART) in Figure 1. They should not, however, be confused as denoting a nonadaptive sequence of treatments that begins with a and is followed by b for everyone. Rather, treatment b in the notation (a, b) is the treatment recommended to nonresponders to initial treatment a. In the simple adaptive interventions considered in this article, all participants who responded to initial treatment a continued to receive treatment a subsequently; this part of second-stage treatment is not encoded in this notation. ADHD = attention-deficit/hyperactivity disorder.

the SMART design (e.g., the difference between the four adaptive interventions specified in Table 1; Oetting, Levy, Weiss, & Murphy, 2007); and (c) as described in the companion article (Nahum-Shani et al., 2012) the construction of more deeply tailored adaptive interventions that go beyond those explicitly embedded as part of the SMART design.

Trials in which each participant is randomized multiple times have been widely used across the medical fields. Precursors of SMART include the CALGB Study 8923 for treating elderly patients with primary acute myelogenous leukemia (Stone et al., 1995) and the STAR*D for treatment of depression (Fava et al., 2003; Lavori et al., 2001). In recent years, a number of SMART trials have been conducted. These include trials for treating cancer at the University of Texas MD Anderson Cancer Center (for more details see Thall et al., 2007; Thall, Millikan, & Sung, 2000); the National Institute of Mental Health Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) study concerning antipsychotic drug effectiveness for persons with schizophrenia (for more details see Stroup et al., 2003); and the Adaptive Interventions for Children With ADHD conducted by one of the authors (W. Pelham) at SUNY-Buffalo. The latter SMART study is discussed above and in the following.

The ADHD SMART Study

In the first stage of the ADHD SMART study (the beginning of the school year), children were randomly assigned (with probability .5) to a low dose of medication or a low dose of behavioral intervention. Beginning at 8 weeks, children's response to the first-stage intervention was evaluated monthly until the end of the school year. Monthly ratings from the Impairment Rating Scale (IRS; Fabiano et al., 2006) and an individualized list of target behaviors (ITB; see e.g., Pelham, Evans, Gnagy, & Greenslade, 1992; Pelham et al., 2002) were used to evaluate response. At each monthly assessment, children whose average performance on the ITB was less than 75% and who were rated by teachers as impaired on the IRS in at least one domain were designated as nonresponders to the first-stage intervention and entered into the second stage of the adaptive intervention. These children were then rerandomized (with probability .5) to one of two second-stage intervention options, either to increasing the dose of the first-stage intervention option or to augmenting the first-stage intervention option

with the other type of intervention (i.e., adding behavioral intervention for those who started with medication or adding medication for those who started with behavioral intervention). Children categorized as responders continued with the assigned first-stage intervention option. Consistent with each randomization, this trial was designed to investigate two critical decisions: the decision regarding the first-stage intervention and then the decision regarding the second-stage intervention for those not responding satisfactorily to the first-stage intervention.

The structure of this SMART study is illustrated in Figure 1. The four adaptive interventions listed in Table 1 are embedded in this SMART study. Notice that each adaptive intervention embedded in this design is operationalized by a sequence of decision rules that specifies the intervention options at each stage for both responders and nonresponders. Accordingly, there are responding and nonresponding children who are *consistent* with each adaptive intervention (i.e., who were assigned to experimental conditions that form a specific sequence of decision rules). More specifically, each responding child receives only the initially assigned intervention and thus the child would have received the same treatment under two different adaptive interventions; these two adaptive interventions differ only in how nonresponding children are treated. As illustrated in Figure 1, children in Subgroups 1 and 2 are consistent with an adaptive intervention that begins with a low dose of medication and then nonresponders receive an increased dose of medication (Subgroup 2), whereas responders continue with the same low dose of medication (Subgroup 1); children in Subgroups 1 and 3 are consistent with an adaptive intervention that begins with a low dose of medication and then behavioral intervention is added to nonresponders (Subgroup 3), whereas responders continue with the same low dose of medication (Subgroup 1). Notice that responders to a low dose of medication (Subgroup 1) are consistent with the two adaptive interventions specified above (i.e., the two adaptive interventions that begin with a low dose of medication). Figure 1 also illustrates that children in Subgroups 4 and 5 are consistent with an adaptive intervention that begins with low-intensity behavioral intervention and then the intensity of the behavioral intervention is increased for nonresponders (Subgroup 5) and responders continue with the same low-intensity behavioral intervention (Subgroup 4); children in Subgroups 4 and 6 are
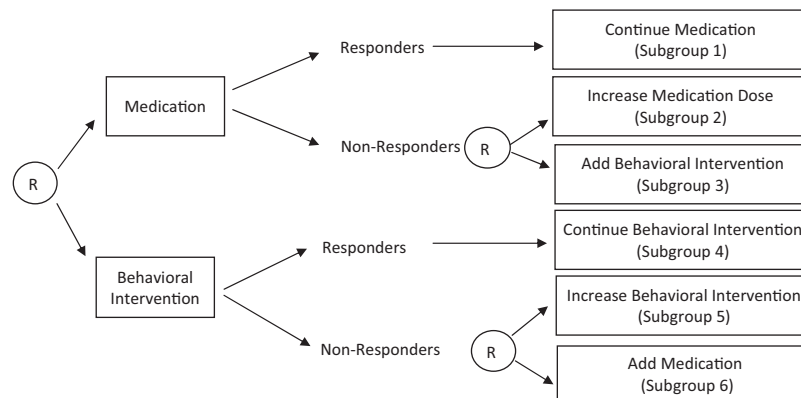
Figure 1. Sequential multiple assignment randomized trial for attention-deficit/hyperactivity disorder (ADHD) study.

consistent with an adaptive intervention that begins with low-intensity behavioral intervention and then medication is added to nonresponders (Subgroup 6) and responders (Subgroup 5) continue with the same low-intensity behavioral intervention. Notice that responders to low-intensity behavioral intervention (Subgroup 4) are consistent with two adaptive interventions (the two adaptive interventions that begin with low-intensity behavioral intervention).

Other Types of SMART Designs

The ADHD study described above is a SMART study in which rerandomization to the second-stage intervention options depends on an intermediate outcome (i.e., response/nonresponse to first-stage intervention). Other types of SMART designs might vary in the extent and form of the tailoring that is incorporated in the design. Such aspects of the design are determined by the investigator based on scientific evidence, as well as ethical considerations. In the following we describe three other common types of SMART studies: SMART designs with no embedded tailoring variables, SMART designs in which participants are rerandomized to different second-stage intervention options depending on an intermediate outcome, and SMART designs in which the decision whether to rerandomize or not depends on an intermediate outcome and prior treatment.

SMARTs with no embedded tailoring variables. In these SMART designs participants are first randomized to two different first-stage intervention options (e.g., B or C), and then (e.g., after 12 weeks) all participants are rerandomized to two second-stage intervention options (e.g., D or E) regardless of any information observed prior to the second-stage randomization (i.e., regardless of any intermediate outcome or prior treatment). This design is illustrated in Figure 2. Because there are no embedded tailoring variables in this SMART design, the embedded interventions are nonadaptive. More specifically, there are four nonadaptive interventions embedded in this SMART design: (a) Begin with B and then offer D, (b) begin with B and then offer E, (c) begin with C and then offer D, and (d) begin with C and then offer E.

SMARTs in which rerandomization to different second-stage intervention options depends on an intermediate outcome. Consider a SMART design in which participants are first randomized to two different first-stage intervention options (e.g., B or C). Then (e.g., after 12 weeks), all participants are rerandomized
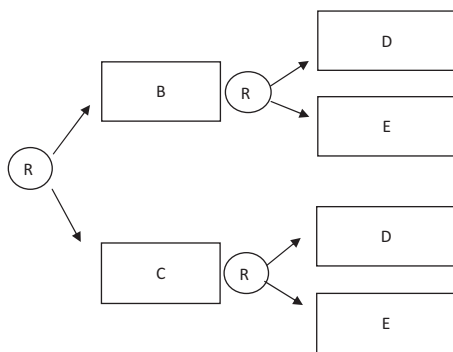
to second-stage intervention options, but second-stage intervention options vary depending on an intermediate outcome (e.g., response/nonresponse to first-stage intervention options). More specifically, assume responders are rerandomized to one of two maintenance interventions, M or M+, whereas nonresponders are rerandomized to either switch to a third intervention E or to receive the combined intervention B + C. This design is illustrated in Figure 3.

There are eight adaptive interventions embedded in this design: (a) Begin with B, and then offer E for nonresponders and M for responders; (b) begin with B, and then offer E for nonresponders and M+ for responders; (c) begin with B, and then offer B + C for nonresponders and M for responders; (d) begin with B, and then offer B + C for nonresponders and M+ for responders; (e) begin with C, and then offer E for nonresponders and M for responders; (f) begin with C, and then offer E for nonresponders and M+ for responders; (g) begin with C, and then offer B + C for nonresponders and M for responders; (h) begin with C, and then offer B + C for nonresponders and M+ for responders.

SMARTs in which the decision whether to rerandomize or not depends on an intermediate outcome and prior treatment. Assume for example that participants are first randomized to two different first-stage intervention options (e.g., B or C), and then (e.g., after 12 weeks) only nonresponders to B are rerandomized to either intensify B (B+) or add the alternative intervention option (B + C). Responders to B are not rerandomized and remain on the same first-stage intervention option (B); responders to C are not rerandomized and remain on the same first-stage intervention option (C); and nonresponders to C are not rerandomized and are offered an intensified version of C (C+). This design is illustrated in Figure 4. Notice that in this design, whether or not a participant is rerandomized to second-stage intervention options depends on both the observed intermediate outcome and the prior intervention option offered to the participant. There are three adaptive interventions embedded in this design: (a) Begin with B and then offer B to responders and B+ to nonresponders; (b) begin with B and then offer B to responders and B + C to nonresponders; and (c) begin with C and then offer C to responders and C+ to nonresponders.

To better understand the SMART design and its advantages, in the following we compare the SMART with other common experimental designs.

SMART Relative to Other Experimental Approaches

Factorial design. A classic 2 × 2 factorial is an experimental design with two factors, each with two levels that are crossed, resulting in four experimental conditions (Collins, Dziak, & Li, 2009). The SMART design in Figure 2 might seem like a 2 × 2 factorial design in which the first factor corresponds to the first-stage intervention options (with two levels: B and C) and the second factor corresponds to the second-stage intervention option (with two levels: D and E). However, in the SMART design, the second factor is provided after the provision of the first factor (e.g., after 12 weeks). Furthermore, in many SMART designs, intermediate outcomes are used to determine whether an experimental unit will be rerandomized (e.g., the design in Figure 1) or the levels of the second factor to which an experimental unit will be rerandomized (e.g., the design in Figure 3). Hence, data analysis methods for
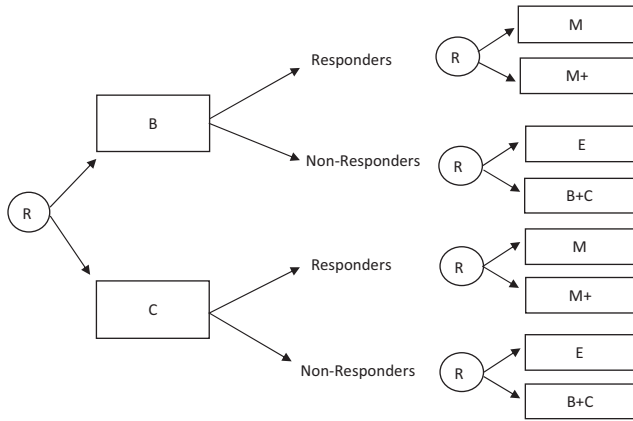

Figure 2. Sequential multiple assignment randomized trial (SMART) with no embedded tailoring variables.

Figure 3. Sequential multiple assignment randomized trial (SMART) design in which participants are rerandomized to different second-stage intervention options depending on an intermediate outcome (e.g., response/ nonresponse).

addressing primary research questions might differ from those typically used in a classic 2 × 2 factorial design (as we later discuss in the Analysis Methods section).

Just as some factorial designs are unbalanced (e.g., some of the factors are not completely crossed; see Collins et al., 2009; Dziak, Nahum-Shani, & Collins, 2012), SMART designs can be unbalanced as well. In particular the SMART design will be unbalanced when the prior intervention option determines whether or not an experimental unit will be rerandomized (e.g., the designs in Figure 4) and/or the second-stage intervention options depend on the prior intervention option (e.g., the design in Figure 1). To clarify this, consider the ADHD SMART (in Figure 1), in which the intervention option offered to the child at the first stage (i.e., the first factor) was used to decide the set of second-stage intervention options if the child did not respond in the first stage. That is, the second-stage interventions for children who do not respond to low-intensity behavioral modification are different from the second-stage options for children who do not respond to low-dose medication.

Randomized trial with multiple groups. Sometimes the SMART is equivalent to a randomized trial with multiple groups, each corresponding to a different adaptive intervention. For example, in the ADHD example (in Figure 1), instead of randomizing to two first-stage intervention options and then rerandomizing nonresponders at the time of nonresponse, an investigator might conduct both randomizations prior to the first-stage intervention. In this case the second randomization will only be used if the participant does not adequately respond to the first-stage intervention. This approach is equivalent to randomizing children upfront to the four adaptive interventions in Table 1 with randomization probability .25 each. However, performing all randomizations prior to the first-stage intervention (e.g., randomizing to four adaptive interventions) is disadvantageous if there are important first-stage intermediate outcomes (e.g., adherence to the first-stage intervention or side effects experienced during the first-stage intervention) that might be strongly predictive of second-stage primary outcomes. Such first-stage intermediate outcomes are prognostic factors

for the second-stage primary outcome. In this case, the second-stage intervention groups might, by chance, differ in the distribution of the prognostic factor, leading to spurious differences between the second-stage intervention groups in terms of the primary outcome.

To clarify this, suppose that in planning the ADHD (in Figure 1) study we believed that some families/children are less adherent to any intervention option and thus will generally exhibit poorer school performance. In this case a natural prognostic factor for the second-stage primary outcome is adherence to the first-stage intervention. In general, stratification or blocking is used to achieve distributional balance of a prognostic factor between intervention groups within each level of a stratum (Efron, 1971; Simon, 1979). In stratification we partition participants into mutually exclusive subsets defined by the prognostic factor. Within these strata, we then randomize between intervention groups. The purpose of stratified randomization is to provide increased confidence that the compared groups are similar with respect to known prognostic factors (e.g., adherence to first-stage intervention) and thus reduce the suspicion that differences in the primary outcome are simply due to chance differences in the composition of the prognostic factors between intervention groups (Hedden, Woolson, & Malcolm, 2006; Kernan, Viscoli, Makuch, Brass, & Horwitz, 1999).

Accordingly, if we believe adherence to the first-stage intervention is a prognostic factor for the second-stage primary outcome, we might stratify randomization to the second-stage intervention options by the level of adherence to the first-stage intervention. Because information concerning adherence to the first-stage intervention is not available prior to the first-stage intervention, we would not be able to randomize a priori to the four adaptive interventions in Table 1. Instead, we would delay the rerandomization for nonresponders until the time of nonresponse (at which time we will also know the child's level of adherence to the first-stage intervention). This is one advantage of multistage randomizations over conducting a trial that randomizes participants upfront to the different adaptive interventions.

A standard randomized controlled trial (RCT). The traditional approach to intervention development involves constructing an intervention a priori and then evaluating it in an RCT. For
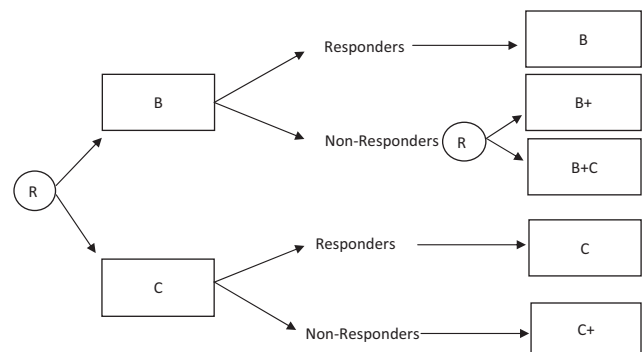
Figure 4. Sequential multiple assignment randomized trial (SMART) design in which the decision whether to rerandomize depends on an intermediate outcome and prior treatment.

example, Connell and colleagues (Connell, 2009; Connell & Dishion, 2008; Connell, Dishion, Yasui, & Kavanagh, 2007) conducted an RCT to evaluate an adaptive family-centered intervention (described in details in Dishion & Kavanagh, 2003) for reducing adolescent problem behaviors in public schools. High-risk adolescents were randomly assigned at the individual level to either control (school as usual) classrooms (498 youths) or adaptive intervention classrooms (500 youths) in the seventh grade. Typically, the adaptive intervention is constructed using theory, real-world experience, surveys of the literature and expert opinion.

In RCTs, the goal is to assess whether the intervention results in better primary outcomes than the control condition (e.g., placebo control, attention control, or treatment as usual). Further development of the intervention is most frequently informed by nonrandomized comparisons. For example, a common secondary analysis uses lack of fidelity to intervention assignment (e.g., variation in intervention adherence by staff and participants) to inform which intervention options are best for which participants and to inform whether the intervention should vary in response to how the participant is progressing. These analyses, due to their dependence on nonrandomized comparisons, provide lower quality evidence than do analyses based on randomized comparisons (Cook & Campbell, 1979; Shadish, 2002; Shadish, Cook, & Campbell, 2002). Thus, although RCTs are well suited for evaluating an adaptive intervention (i.e., assessing its effectiveness compared to control), they are not as well suited for developing adaptive interventions (i.e., for revising or constructing adaptive interventions). On the other hand, as we demonstrate below and in the companion article (Nahum-Shani et al., 2012), SMART designs enable the data analyst to conduct randomized comparisons in order to investigate which intervention options are best for which participants and to investigate whether the intervention should vary in response to how the participant is progressing.

Single-stage-at-a-time experimental approach.   This is a natural alternative to SMART, particularly when randomized trials involving initial treatments have already been conducted. In this approach, a separate randomized trial is conducted for each stage, and each randomized trial involves a new group of participants. Then, the results from each separate randomized trial are pieced together to infer what the best adaptive intervention ought to be. Essentially, the single-stage-at-a-time approach focuses separately on each individual intervention stage in the sequence, as opposed to considering the entire sequence as a whole. To ascertain the best adaptive intervention, this approach compares available intervention options at each stage based on a randomized trial for that stage alone and/or based on historical trials and the available literature. For example, Moore and Blackburn (1997) recruited outpatients with recurrent major depression who failed to respond to acute treatment with antidepressant medication, aiming to determine whether cognitive therapy is a better follow-up treatment than medication. Walsh et al. (2000) recruited patients with eating disorders who had not responded to, or had relapsed following, a course of cognitive behavior therapy or interpersonal psychotherapy, aiming to determine the utility of a pharmacological intervention for patients whose response to psychological treatment was not satisfactory. In these two examples, investigators conducted randomized trials in order to compare second-stage intervention options among nonresponders, while explicitly or implicitly

making conclusions concerning the best initial intervention option based on the available scientific literature.

To further clarify this point, assume that instead of the ADHD SMART study discussed earlier (in Figure 1), an investigator decides to first review the literature on childhood ADHD in order to choose whether to begin with low-dose medication or low-intensity behavioral intervention. Then, the researcher conducts a randomized trial for nonresponders in which nonresponding children are recruited and then randomized to one of the two second-stage intervention options (i.e., increase the dose of the first-stage intervention option, or augment the first-stage intervention option with the alternative type of intervention). While conceptually simpler than a SMART design, this approach has at least three disadvantages.

First, this approach might fail to detect possible delayed effects in which an early stage intervention option has an effect that is less likely to occur unless it is followed by a particular subsequent intervention option. This might occur when subsequent intervention options intensify, facilitate, or weaken the impact of initial intervention options. To clarify this concept, consider research in the area of experimental psychology that suggests that there might be a delayed effect when attention-control training is provided for improving individuals' ability to cope with tasks that require dividing attention (e.g., driving a car). Gopher and colleagues (see Gopher, 1993; Shani, 2007) have shown that the knowledge and abilities acquired through attention-control training influence and benefit performance beyond the specific situation in which these skills were acquired (e.g., when a new concurrent task was introduced or when no feedback was provided). In fact, such training appears to contribute to the development of attention management skills that transfer beyond the specific situation in which the individual was originally trained, such that the individual continues to improve as he/she experiences other types of tasks that require the allocation of attention. In other words, subsequent training/experience that involves different types of tasks than those used during the initial training provides an opportunity for the individual to internalize the specialized knowledge acquired during the initial training and hence facilitates the development of attention management skills. Accordingly, the advantages of the initial training cannot be fully detected without exposing the participant to subsequent training or experiences.

Overall, when using the single-stage experimental approach, conclusions concerning the best initial intervention options are based on studies in which participants were not necessarily exposed to subsequent intervention options. This might reduce the ability of the researcher to detect these types of delayed effects and hence might lead to the wrong conclusion as to the most effective sequence of intervention options.

Second, the SMART approach enables the researcher to detect prescriptive information that the initial intervention elicits. The investigator can use this information to better match the subsequent intervention to each participant and thus improve the primary outcome. This is not possible by piecing together the results of single-stage specific randomized trials. For example, in the context of the ADHD study, results of secondary data analysis (reported in the companion article, Nahum-Shani et al., 2012) indicate that the child's adherence to the first-stage intervention might be a promising tailoring variable for the

second-stage intervention. Nonresponders (to either medication or behavioral intervention) with low adherence to the first-stage intervention performed better when the first-stage intervention was augmented with the other type of intervention, relative to when the dose of the first-stage intervention was increased. Assume the investigator had chosen to take the single-stage-at-a-time approach and conducted a nonresponder trial in which children who did not respond to medication are randomized to two intervention options (increase the dose of medication or augment with behavioral intervention). In this more limited study the investigator would not obtain information concerning the second-stage intervention option for nonresponders who do not adhere to behavioral intervention.

Third, subjects who enroll and remain in studies in which there are no options for nonresponders (or responders) might be different than those who enroll and remain in a SMART trial. For example, previous studies might not offer subsequent intervention options to children who are not improving, and hence children in these studies might exhibit differing patterns of adherence and/or might be more prone to drop out relative to children in a SMART study who know that their intervention can be altered. Thus the choice of first-stage intervention option based on the single-stage approach might be based on poorer quality data (due to nonadherence) relative to the SMART. Moreover, in the single-stage approach, the investigator usually selects the first-stage and second-stage intervention options based on data from different samples of participants. Because the samples might differ due to changing schooling practices, changing diagnostic criteria, and so on, bias can be introduced into the results (see Murphy et al., 2007, for a detailed discussion of these *cohort* effects). In a SMART, on the other hand, conclusions concerning the best first and second-stage intervention options can be made in a synergetic manner using the same sample of participants throughout the stages of the decision process.

Analysis Methods

Denote the observable data for a subject in a two-stage SMART by (O_1, A_1, O_2, A_2, Y), where O_1 and O_2 are vectors of pretreatment information and intermediate outcomes, respectively, A_1 and A_2 are the randomly assigned first- and second-stage intervention options, and Y is the primary outcome of an individual. For example, in the adaptive ADHD study (see Figure 1), O_1 might include severity of ADHD symptoms, whether the child received medication during the previous school year, or other baseline measures; O_2 might include the subject's response status and adherence to the first-stage intervention; and Y might be the teacher's evaluation of the child's school performance at the end of the school year. Let A_1 denote the indicator for the first-stage intervention ($1 =$ low-intensity behavioral intervention; $-1 =$ low-dose medication) and A_{2NR} denote the indicator for the second-stage intervention for nonresponders to the first-stage intervention ($1 =$ increase the initial intervention; $-1 =$ augment the initial intervention with the other type of intervention).

The multiple randomizations in the SMART allow the investigator to estimate a large variety of causal effects important in the development of adaptive interventions. Here we consider three primary research questions (a) concerning the difference between first-stage intervention options, (b) concerning the difference between second-stage intervention options, and (c) concerning the comparison of adaptive interventions that are embedded within the SMART design.

Comparing First-Stage Intervention Options

Consider, for example, the following question concerning the difference between low-dose medication versus low-intensity behavioral intervention: In the context of the specified second-stage intervention options, does starting with low-intensity behavioral intervention result in a better long-term outcome relative to starting with low-dose medication? This question is addressed by pooling Y from Subgroups 1 through 3 and comparing the resulting average to the pooled Y from Subgroups 4 through 6 (in Figure 1). This is the main effect of the first-stage intervention, that is, the difference between the two first-stage intervention options, averaging over the second-stage intervention options. Notice that responders to the first-stage intervention are included in this comparison because a particular first-stage intervention might lead to good initial response but the performance of responders might deteriorate over time (see Oetting et al., 2007, for more details).

Comparing Second-Stage Intervention Options

Consider, for example, the following question concerning the difference between the two second-stage intervention options: Among those who do not respond to their initial intervention, is there a difference between intensifying the initial intervention versus augmenting the initial intervention? This question is addressed by pooling Y from Subgroups 2 and 5 and comparing the resulting average to the pooled Y from Subgroups 3 and 6 (in Figure 1). This is the main effect of the second-stage intervention options for nonresponding children, that is, the difference between the two second-stage intervention options, averaging over the first-stage intervention options.

Comparing Adaptive Interventions That Are Embedded Within the SMART Design

To understand how to compare adaptive interventions that are embedded in the SMART design, consider first estimating the mean outcome of only one of the four adaptive interventions, say $(1, 1)$. In adaptive intervention $(1, 1)$ children are initially provided low-intensity behavioral intervention and then the intensity of the behavioral intervention is increased only if the child does not respond. To estimate the mean outcome under this adaptive intervention, a natural (yet incorrect) approach might be to average the outcomes of all children in the study who are consistent with this adaptive intervention (Subgroups 4 and 5 in Figure 1). Observe that the outcomes of all children who responded to the low-intensity behavioral intervention are included in this sample average, whereas the outcomes of only half of all children who did not respond to low-intensity behavioral intervention are included in this sample average. It turns out that this sample average is a biased estimator for the mean outcome that would occur if all children in the population were provided the adaptive intervention $(1, 1)$. This bias occurs because, by design, nonresponding children

are rerandomized and thus split into two subgroups (5 and 6 in Figure 1), whereas the responding children are not rerandomized and thus not split into two subgroups. Therefore, due to the restricted randomization scheme, the above sample average contains an overrepresentation of outcomes from responding children and an underrepresentation of outcomes from nonresponding children. To accommodate this over/underrepresentation, weights can be used. Since this over/underrepresentation occurs by design, we know the value of the weights to use in order to counteract this over/underrepresentation. In particular, the weights are the inverse of the randomization probability, that is, $W = 2$ is the weight for responders and $W = 4$ is the weight for nonresponders.[1] Informally, each participant receives a weight that is inversely proportional to his/her probability of receiving his/her own adaptive intervention. This is similar to the inverse-probability-of-treatment weights used in the estimation of marginal structural models (MSM) for making causal inferences concerning the effects of time varying treatments (see S. R. Cole & Hernán, 2008; Hernán, Brumback, & Robins, 2000).

In this estimation procedure, the weight assigned to each child depended on whether the child was a responder or a nonresponder to the first-stage intervention. Therefore, the distribution of the weights depends on the observed response rate in the sample, a statistic that varies from one sample to the next. In order to account for the sample-to-sample variance in the distribution of the weights, robust (sandwich) standard errors can be used to make appropriate inferences (e.g., confidence intervals, p values) with the weighted averages (see Hernán et al., 2000, for more details).

Following the estimation procedure described above, one can obtain the average weighted outcome, separately, for each of the four adaptive interventions that are embedded in the SMART study. As indicated above, for each adaptive intervention, the investigator will need to restrict the data set and estimate the average weighted outcome only for those participants who are consistent with that adaptive intervention. However, investigators are typically interested in carrying out one data analysis (i.e., using one fitted regression model) that enables them to (a) estimate the mean outcome under all four adaptive interventions simultaneously; and (b) estimate all possible mean-comparisons between the four adaptive interventions simultaneously, while (c) providing appropriate inferences (e.g., confidence intervals, p values) for these estimates. Apart from convenience, doing this has the added advantage of allowing investigators to control for baseline measures that may be highly correlated with the outcome. This has the potential to improve effect estimates by reducing error variance (i.e., leading to smaller standard errors and therefore increased statistical power).

To utilize the entire sample using standard software, the data set needs to be restructured. This is because some participants' observations are consistent with more than one of the embedded adaptive interventions. In the ADHD study these are the responding participants. A participant who responds, say to low-intensity behavioral intervention, has outcomes that are consistent with two adaptive interventions: Begin with low-intensity behavioral intervention and then add medication only if the child does not respond $(1, -1)$, and begin with low-intensity behavioral intervention and then increase the intensity of the behavioral intervention only if the child does not respond $(1, 1)$. Thus, in order to simultaneously estimate the mean outcome under adaptive interventions $(1, -1)$

and $(1, 1)$ using the ADHD SMART data, each responder's outcome would have to be used twice. This point was first made by Robins and colleagues (Orellana, Rotnitzky, & Robins, 2010; Robins, Orellana, & Rotnitzky, 2008) in their approach for use in observational studies to compare adaptive interventions. In this article, we generalize their approach for use with SMART study data, as follows.

In the ADHD study we restructure the data set such that instead of one observation per responder, the new data set includes two identical observations per responder. Then, for each responder, we set A_{2NR} to 1 in one of the two replicated observations and A_{2NR} to -1 in the other replicated observation. For example, if the original data set included 58 responders with no value for the indicator A_{2NR} (in the original data set this happens because there is no variation in second-stage intervention options among responders since they were not rerandomized), the new data set includes 116 responders, half in which $A_{2NR} = 1$ and half in which $A_{2NR} = -1$. The number of observations for nonresponders (who were rerandomized with probability .5) remains the same. Such replication of observations enables investigators to conveniently "reuse" the observed primary outcome from each responding child in estimating the average outcome of two adaptive interventions using one fitted regression model. As before, the new data set should contain a variable providing the weights for each observation that is the value 2 for observations that belong to responders and 4 for observations that belong to nonresponders.

Note that the replication of observations and the use of weights discussed above serve different purposes. The weights are used to accommodate for over/underrepresentation of outcomes resulting from the randomization scheme. This can occur when (a) the number of randomizations vary across participants (e.g., some participants are rerandomized while others are not), and/or (b) randomization probabilities are not equal (e.g., at the first stage, participants are randomized with probability .6/.4 to medication vs. behavioral intervention, respectively), and/or (c) participants are randomized to differing numbers of intervention options (e.g., responding children are randomized to two second-stage intervention options, while nonresponders are rerandomized to three intervention options). The replication of observations on the other hand, is done in order to enable investigators utilize standard software to conduct simultaneous estimation and comparison of the effects of all four adaptive interventions. Such replication is

---

[1] The average of the weighted outcomes of all children in the study who are consistent with this adaptive intervention (Subgroups 4 and 5 in Figure 1) equals $\frac{\sum_{i=1}^{N(1,1)} (w_i Y_i)}{\sum_{i=1}^{N(1,1)} w_i}$, where w_i is the weight assigned to each child who is consistent with adaptive intervention $(1, 1)$, Y_i is the outcome associated with that child, and $N(1, 1)$ is the number of children consistent with adaptive intervention $(1, 1)$. Investigators may, instead, choose to use normalized weights, such as $w_i^* \frac{N(1,1) w_i}{\sum_{i=1}^{N(1,1)} w_i}$, which averages to 1 (i.e., $\frac{\sum_{i=1}^{N(1,1)} w_i^*}{N(1,1)} = 1$), and still obtain the same average of weighted outcome. Further, the robust standard error calculations used by SAS PROC GENMOD are also a function of weighted averages $\frac{\sum_{i=1}^{N(1,1)} (w_i f_i)}{\sum_{i=1}^{N(1,1)} w_i}$, where f_i is a function of only the ith child's observations. Thus, the robust standard errors also do not change if one uses normalized weights.

required whenever there are participants whose observations are consistent with more than one embedded adaptive intervention. A technical explanation of this approach (including weighting and replication) as well as a justification for the validity of the robust standard errors is provided in Appendix A.

Using the new data set, the following model can be estimated

$$E[Y|A_1, A_{2NR}, O_1] = \beta_0 + \beta_1 A_1 + \beta_2 A_{2NR} + \beta_3 A_1 A_{2NR} + \boldsymbol{\gamma}^T O_1,$$

(1)

where Y is the school performance at the end of the school year; O_1 is a vector of baseline measures (obtained prior to the first-stage intervention and included in the regression model as *mean-centered covariates*); β_0 is the intercept; $\beta_1 - \beta_3$ are the regression coefficients expressing the effects of the first-stage intervention, the effects of the second-stage intervention offered to nonresponders, and the interaction between them (respectively); and $\boldsymbol{\gamma}$ is the vector of regression coefficients expressing the effects of the baseline measures O_1. Accordingly, Equation 1 models the average school performance for each adaptive intervention, such that $E[Y|1, -1] = \beta_0 + \beta_1 - \beta_2 - \beta_3$ is the average school performance for children following adaptive intervention $(1, -1)$; $E[Y| -1, -1] = \beta_0 - \beta_1 - \beta_2 + \beta_3$ is the average school performance for children following adaptive intervention $(-1, -1)$; $E[Y|1, 1] = \beta_0 + \beta_1 + \beta_2 + \beta_3$ is the average school performance for children following adaptive intervention $(1, 1)$; and $E[Y| -1, 1] = \beta_0 - \beta_1 + \beta_2 - \beta_3$ is the average school performance for children following adaptive intervention $(-1, 1)$.

Notice that because the aim of this analysis is to compare the four adaptive interventions embedded in the SMART design, Equation 1 does not include an interaction between the intervention options (A_1, A_{2NR}) and the baseline measures, or intermediate outcomes (O_1, O_2). O_1 is included in Equation 1 only to reduce error variance. However, interactions between the baseline measures and the intervention options could be included in this analysis to explore whether these intervention options should be more deeply tailored. Unfortunately the inclusion of intermediate outcomes in this analysis either by themselves or as interactions with intervention options can lead to bias; see the companion article (Nahum-Shani et al., 2012), for discussion and for a method that can be used to jointly consider tailoring with baseline measures and with intermediate outcomes (e.g., whether and how the first-stage intervention should be adapted to the child's baseline severity measures O_1, and whether and how the second-stage intervention for nonresponders should be adapted to the child's intermediate outcomes O_2).

In order to estimate the model in Equation 1 we use weighted regression with the SAS GENMOD procedure (SAS Institute, 2008); this procedure minimizes

$$\sum_{j=1}^{M} W_j(Y - \beta_0 - \beta_1 A_{1j} - \beta_2 A_{2j} - \beta_3 A_{1j} A_{2NRj} - \boldsymbol{\gamma}^T O_{1j})^2$$

(2)

to estimate the regression coefficients. The summation is over the observations in the new data set, M is the number of observations (i.e., rows) in the new data set, and W is the weight assigned to each observation in the new data set. Because the weights (depending on response/nonresponse) are random and observations are replicated, it is not immediately obvious that the estimator of

the mean of the primary outcome for each adaptive intervention is consistent (unbiased in large samples). In Appendix A we provide an intuitive demonstration of how estimating this model with the SAS GENMOD procedure based on the new data set results in consistent estimators of the population averages for the regression coefficients. To estimate the standard errors, we use the robust (sandwich) standard errors provided by the SAS GENMOD procedure (SAS Institute, 2008). Justification for the use of the robust standard errors is also included in Appendix A. In Appendix B we provide the syntax for estimating the regression coefficients and robust standard errors, using SAS GENMOD procedure. SAS code for comparing adaptive interventions that are embedded this this SMART design (including generating the weights and replicating observations) is available at the following website: http://methodology.psu.edu/ra/adap-treat-strat/smartcodeex

Data Analysis Methods for Other Forms of SMART Designs

Above, we discussed data analysis methods for addressing primary research questions using data from a SMART design in which response/nonresponse determines whether a participant is rerandomized (e.g., in the ADHD study only nonresponders were rerandomized to second-stage intervention options). Here, we illustrate how the ideas discussed above can be used to compare the embedded adaptive interventions when data are obtained from other common types of SMART designs. More specifically, we focus on SMART designs with no embedded tailoring (e.g., the design in Figure 2); a SMART design in which participants are rerandomized to different second-stage intervention options, depending on an intermediate outcome (e.g., the design in Figure 3); and a SMART design in which whether a participant is rerandomized depends on both an intermediate outcome and the prior treatment (e.g., the design in Figure 4). To address questions concerning the main effects of the first-stage or second-stage intervention option using data from these SMART designs, one can use the same analysis methods discussed with respect to the ADHD SMART study. However, to compare adaptive interventions that are embedded in the SMART design, the primary difference between the analysis methods for the ADHD design and these designs concern the form of the weights and which, if any, participant observations need to be replicated. Also, the model used to compare the embedded adaptive interventions might differ.

Data analysis for SMART designs with no embedded tailoring variables. Consider the SMART design illustrated in Figure 2 and assume that at the first stage, participants were randomized with probability .5 to two intervention options: B or C. After 12 weeks, all participants were rerandomized with probability .5 to two second-stage intervention options: D or E, regardless of any information observed prior to the second stage. Let A_1 denote the indicator for the first-stage intervention options (coded –1 for B, and 1 for C). Let A_2 denote the indicator for the second-stage intervention options (coded –1 for D, and 1 for E). Note that because there are no embedded tailoring variables, the embedded intervention sequences are nonadaptive unlike those in the ADHD study. There are four embedded nonadaptive intervention sequences (see below).

Weights are not required in this setting because all participants are randomized twice, each with equal probability, and all participants were randomized to the same number of intervention options (i.e., two). Furthermore because each participant's observations are consistent with only one embedded nonadaptive intervention sequence there is no need to replicate participant observations in order to use common statistical software such as the SAS GENMOD procedure. Because no observations are replicated and weights are not assigned, traditional estimates of the standard errors can be used to conduct inference.

A model that can be used to compare the 4 embedded nonadaptive interventions is the usual analysis of variance (ANOVA) model:

$$E[Y|A_1, A_2, O_1] = \beta_0 + \beta_1 A_1 + \beta_2 A_2 + \beta_3 A_1 A_2 + \gamma^T O_1,$$

where Y is the primary outcome; O_1 is a vector of baseline measures (obtained prior to the first-stage intervention and included in the regression model as *mean-centered covariates*); β_0 is the intercept; and $\beta_1 - \beta_3$ are the regression coefficients expressing the effects of the first-stage intervention, the effects of the second-stage intervention, and the interaction between them, respectively. γ is the vector of regression coefficients expressing the effects of the baseline measures O_1. Accordingly, $E[Y| 1, -1] = \beta_0 + \beta_1 - \beta_2 - \beta_3$ is the average outcome for participants following a nonadaptive intervention that begins with C and then offers D (1, -1); $E[Y| -1, -1] = \beta_0 - \beta_1 - \beta_2 + \beta_3$ is the average outcome for participants following a nonadaptive intervention that begins with B and then offers D ($-1, -1$); $E[Y| 1, 1] = \beta_0 + \beta_1 + \beta_2 + \beta_3$ is the average outcome for participants following a nonadaptive intervention that begins with C and then offers E (1, 1); and $E[Y| -1, 1] = \beta_0 - \beta_1 + \beta_2 - \beta_3$ is the average outcome for participants following a nonadaptive intervention that begins with B and then offers E ($-1, 1$).

Data analysis for SMARTs in which rerandomization to different second-stage intervention options depends on an intermediate outcome. Consider the SMART design illustrated in Figure 3 and assume that at the first stage, participants were randomized with probability .5 to either B or C. After 12 weeks, responders were rerandomized with probability .5 to one of two maintenance interventions, M or M+. Nonresponders, on the other hand, were rerandomized with probability .5 to either switch to a third intervention E or to the combined intervention B + C. Let A_1 denote the indicator for the first-stage intervention options (coded -1 for B and 1 for C); let A_{2R} denote the indicator for the second-stage intervention options for responders (coded -1 for M and 1 for M+); and let A_{2NR} denote the indicator for the second-stage intervention options for nonresponders (coded -1 for E and 1 for B + C). There are eight embedded adaptive interventions in this study (see below).

Weights are not required in this setting because all participants are randomized twice, each with equal probability, and all participants were randomized to the same number of intervention options (i.e., two). Next each participant's observations are consistent with two adaptive interventions; hence the new data set should contain two replicates of each participant's observations. As written above, the replication enables the reuse of the observed outcome from each participant in estimating the mean outcome of the two adaptive interventions that are consistent with this participant's observations. In particular for each participant who responded set A_{2NR} (i.e., the indicator for the second-stage intervention options for nonresponders) to -1 in one of the two replicated observations and A_{2NR} to 1 in the other replicated observation, and for each participant who did not respond set A_{2R} (i.e., the indicator for the second-stage intervention options for responders) to -1 in one of the two replicated observations and A_{2R} to 1 in the other replicated observation.

Finally, using the new data set, standard regression can be applied using a model such as

$$E[Y|A_1, A_{2R}, A_{2NR}, O_1] = \beta_0 + \beta_1 A_1 + \beta_2 A_{2R} + \beta_3 A_{2NR} + \beta_4 A_1 A_{2R}$$
$$+ \beta_5 A_1 A_{2NR} + \beta_6 A_{2R} A_{2NR} + \beta_7 A_1 A_{2R} A_{2NR} + \gamma^T O_1,$$

where Y is the primary outcome; O_1 is a vector of baseline measures (obtained prior to the first-stage intervention and included in the regression models as *mean-centered covariates*); β_0 is the intercept; and $\beta_1 - \beta_7$ are the regression coefficients expressing the effect of the first-stage intervention option, the effects of the second-stage intervention options (offered to responders and nonresponders, respectively), and the interactions between them, respectively. γ is the vector of regression coefficients expressing the effects of the baseline measures O_1. Accordingly, $E[Y| -1, -1, -1] = \beta_0 - \beta_1 - \beta_2 - \beta_3 + \beta_4 + \beta_5 + \beta_6 - \beta_7$ is the average outcome for participants following the adaptive intervention ($-1, -1, -1$); $E[Y| -1, -1, 1] = \beta_0 - \beta_1 - \beta_2 + \beta_3 + \beta_4 - \beta_5 - \beta_6 + \beta_7$ is the average outcome for participants following the adaptive intervention ($-1, -1, 1$); $E[Y| -1, 1, -1] = \beta_0 - \beta_1 + \beta_2 - \beta_3 - \beta_4 + \beta_5 - \beta_6 + \beta_7$ is the average outcome for participants following the adaptive intervention ($-1, 1, -1$); $E[Y| -1, 1, 1] = \beta_0 - \beta_1 + \beta_2 + \beta_3 - \beta_4 - \beta_5 + \beta_6 - \beta_7$ is the average outcome for participants following the adaptive intervention ($-1, 1, 1$); $E[Y|1, -1, -1] = \beta_0 + \beta_1 - \beta_2 - \beta_3 - \beta_4 - \beta_5 + \beta_6 + \beta_7$ is the average outcome for participants following the adaptive intervention (1, $-1, -1$); $E[Y|1, -1, 1] = \beta_0 + \beta_1 - \beta_2 + \beta_3 - \beta_4 + \beta_5 - \beta_6 - \beta_7$ is the average outcome for participants following the adaptive intervention (1, $-1, 1$); $E[Y|1, 1, -1] = \beta_0 + \beta_1 + \beta_2 - \beta_3 + \beta_4 - \beta_5 - \beta_6 - \beta_7$ is the average outcome for participants following the adaptive intervention (1, 1, -1); and $E[Y|1, 1, 1, O_1] = \beta_0 + \beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5 + \beta_6 + \beta_7$ is the average outcome for participants following the adaptive intervention (1, 1, 1).

Data analysis for SMARTs in which the decision whether to rerandomize or not depends on an intermediate outcome and prior treatment. Consider the SMART design illustrated in Figure 4 and assume that at the first stage, participants were randomized with probability .5 to either B or C. After 12 weeks, only nonresponders to first-stage intervention option B were rerandomized with probability .5 to either B+ or B + C. Let A_1 denote the indicator for the first-stage intervention options (coded -1 for B and 1 for C). Let A_{2NRB} denote the indicator for the second-stage intervention options for nonresponders to B (coded -1 for B+ and 1 for B + C). There are four embedded adaptive interventions (see below for a list).

In this study only a subset of participants are randomized twice: the participants who do not respond to first-stage intervention B. As a result, weights are needed. Specifically, $W = 4$ is the weight for nonresponders to B because these participants are randomized twice, each randomization with probability .5. $W = 2$ is the weight for responders to B, as well as for responders and nonresponders

to C, because these participants are randomized only once with probability .5. In terms of replicating observations, only the participants who responded to first-stage intervention B have observations that are consistent with more than one adaptive intervention. In particular responders to intervention B have observations consistent with the two adaptive interventions: Begin with intervention B and then intensify to B+ only if the participant does not respond, and begin with intervention B and then augment to B + C only if the participant does not respond. In other words, replicating observations for responders to B enables the reuse of the observed outcome from each responder to B in estimating the average outcome of two adaptive interventions $(-1, 1)$ and $(-1, -1)$. Accordingly, for each participant who responded to intervention B, set A_{2NRB} to -1 in one of the two replicated observations and A_{2NRB} to -1 in the other replicated observation.

Next, to form a model, note that because only participants who are initially randomized to intervention B $(A_1 = -1)$ might potentially be rerandomized, A_{2NRB} is nested within $A_1 = -1$. To represent this nested structure, create a new indicator, Z, that equals 0 when $A_1 = 1$ (the first-stage intervention option is C) and equals 1 when $A_1 = -1$ (the first-stage intervention option is B). Finally, using the new data set, standard weighted regression can be applied based on a model of the form

$$E[Y|A_1, A_2Z, O_1] = \beta_0 + \beta_1 A_1 + \beta_2 A_{2NRB}Z + \gamma^T O_1,$$

where Y is the primary outcome; O_1 is a vector of baseline measures (obtained prior to the first-stage intervention and included in the regression models as *mean-centered covariates*); β_0 is the intercept; β_1 is the regression coefficient expressing the effect of the first-stage intervention options; and β_2 is the regression coefficient that expresses the effect of the second-stage intervention options offered to nonresponders to B. Notice that because A_{2NRB} is nested within $A_1 = -1$ (or equivalently $Z = 1$), A_{2NRB} occurs only in the product with Z in this model. γ is the vector of regression coefficients expressing the effects of the baseline measures O_1. Accordingly, $E[Y| 1, 0] = \beta_0 + \beta_1$ is the average outcome for participants following the adaptive intervention that begins with C and then offers C for responders and C+ for nonresponders; $E[Y| -1, -1] = \beta_0 - \beta_1 - \beta_2$ is the average outcome for participants following the adaptive intervention that begins with B and then offers B for responders and B+ for nonresponders; and $E[Y| -1, 1] = \beta_0 - \beta_1 + \beta_2$ is the average outcome for participants following the adaptive intervention that begins with B and then offers B for responders and B + C for nonresponders. The model above can be estimated with SAS GENMOD procedure, using a robust standard error for inference.

Data Analysis Results Based on the Adaptive ADHD Data

Sample

One hundred and forty-nine children (75% boys) between the ages of 5 and 12 (mean 8.6 years) participated in the study. Due to dropout and missing data,[2] the effective sample used in the current analysis was 139. At the first stage (A_1), 71 children were randomized to receive low-dose medication, and 68 were randomized to receive low-intensity behavioral intervention. By the end of the school year, 81 children were classified as nonresponders and rerandomized to one of the two second-stage intervention options, with 40 children assigned to increasing the dose of the first-stage intervention and 41 children assigned to augmenting the first-stage intervention with the other type of intervention. Overall, 46 children were consistent with adaptive intervention $(1, -1$; starting with low-intensity behavioral intervention and then adding low-dose medication to nonresponders and offering the same intensity of behavioral intervention to responders), 53 children were consistent with adaptive intervention $(-1, -1$; starting with low dose of medication and then adding low-intensity behavioral intervention to nonresponders and offering the same dose of medication to responders), 44 children were consistent with adaptive intervention $(1, 1$; starting with low-intensity behavioral intervention and then enhancing the intensity of the behavioral intervention to nonresponders and offering the same intensity of behavioral intervention to responders), and 54 children were consistent with adaptive intervention $(-1, 1$; starting with low-dose medication and then enhancing the dose of medication to nonresponders and offering the same dose of medication to responders).

Measures

Primary outcome (Y). We consider a measure of children's school performance based on the Impairment Rating Scale (IRS; Fabiano et al., 2006) after an 8-month period as our primary outcome. This primary outcome ranges from 1 to 5, with higher values reflecting better school performance. Because the current analysis is for illustrative, rather than substantive, purposes, we use this measure as a primary outcome despite limitations relating to its distribution and reliability.

Baseline measures. (a) Medication prior to first-stage intervention, reflecting whether the child did (coded as 1) or did not (coded as 0) receive medication at home during the previous school year (i.e., prior to the first-stage intervention); (b) ADHD symptoms at the end of the previous school year, which is the mean of teacher's evaluation on 14 ADHD symptoms (the Disruptive Behavior Disorders Rating Scale; Pelham et al., 1992), each ranging from 0 to 3. We reverse-coded this measure such that larger values reflect fewer symptoms (i.e., better school performance); (c) oppositional defiant disorder (ODD) diagnosis, reflecting whether the child was (coded as 1) or was not (coded as 0) diagnosed with ODD before the first-stage intervention.

Results

Table 2 presents parameter estimates for Model 1 (obtained using the SAS GENMOD procedure described in Appendix B); Table 3 presents the estimated means and robust standard error for each adaptive intervention based on the estimated coefficients in Model 1 (as well as the sample size and the sample mean for each

---

[2] In a full analysis one would want to use a modern missing data method such as multiple imputation to avoid bias. Such methods would be valid under the missing at random assumption. Further discussion can be found in Shortreed, Laber, Pineau, and Murphy (2010).

Table 2
Results (Parameter Estimates) for Model 1

Parameter	Estimate	Robust *SE*	95% confidence limit		Z	Pr > |Z|
			LL	UL		
Intercept	3.43	0.23	2.97	3.89	14.63	<.0001
Baseline: ODD diagnosis	0.37	0.18	0.02	0.72	2.07	.0384
Baseline: ADHD symptoms	0.57	0.14	0.29	0.85	3.95	<.0001
Baseline: Medication before Stage 1	−0.61	0.25	−1.10	−0.13	−2.47	.0134
A1	0.07	0.09	−0.11	0.24	0.75	.4555
A2	0.02	0.08	−0.13	0.18	0.26	.7924
A1 * A2	−0.12	0.08	−0.27	0.04	−1.46	.1436

Note. LL = lower limit; UL = upper limit; ODD = oppositional defiant disorder; ADHD = attention-deficit/hyperactivity disorder.

adaptive intervention); and Table 4 presents the estimated differences and the associated robust standard error for each possible comparison of the four adaptive interventions.

As shown in Tables 3 and 4, school performance was higher for those children following adaptive intervention (1, −1; estimated $M = 4.36$, $SE = 0.15$) relative to those following adaptive intervention (−1, −1; estimated $M = 4.00$, $SE = 0.15$). The difference between adaptive interventions (1, −1) and (−1, −1) was found to be marginally significant (estimated difference = 0.36, $SE = 0.22$, $p < .10$). No other significant differences were found between the four adaptive interventions.

Accordingly, this analysis indicates that starting with low-intensity behavioral intervention and then adding medication to nonresponders and offering the same intensity behavioral intervention to responders is better than starting with low-dose medication and then adding low-intensity behavioral intervention to nonresponders and offering the same dose of medication to responders.

Discussion

Adaptive interventions are both conceptually and scientifically appealing. Here, we reviewed the ways in which decision rules can be used to operationalize adaptation and individualization of interventions in the social and behavioral sciences. We also reviewed SMART studies, emphasizing their advantages and arguing for an increase in their use for informing the construction of high-quality adaptive interventions. Finally, we presented a data analysis method for addressing primary research questions concerning the comparison of intervention options at different stages of the adaptive intervention, as well as the comparison of adaptive interventions that are embedded in the SMART design. An adaptive

intervention for children with ADHD SMART study was used to illustrate this method.

SMART designs have the potential to advance research in many areas in the behavioral and social sciences. For example, in the area of goal setting, although goal-setting theorists assert that a goal will result in higher performance if it is difficult, attainable, and specific (Locke & Latham, 1990), more recent approaches to goal setting (see Barlas & Yasarcan, 2006; Fried & Slowik, 2004) suggest that the contribution of goal-setting theory has not been fully realized because of the failure to systematically capture its dynamic and sequential nature. For example, with respect to the difficulty of the goal, individuals might pursue sequential goals, which differ in their level of difficulty: They might function well with less challenging goals if they perceive the accomplishment of these goals as the basis for pursuing more difficult and challenging goals in the future (Fried & Slowik, 2004). Another example relates to the concept of floating goals (see Barlas & Yasarcan, 2006), a more sophisticated and realistic goal-related structure in which the difficulty of the goal is adjusted depending on current conditions (e.g., if the performance of an individual is persistently poor in approaching the original goal, then he/she lowers the goal, but if he/she exhibits surprisingly good performance, then the goal is pushed higher). These new approaches raise questions concerning the optimal sequence of goals and the way in which the goals should be adjusted over time in order to maximize long-term performance. A SMART design in which each participant can be sequentially randomized multiple times might be appropriate for addressing such research questions by allowing within-individual variation in the extent to which the goal is difficult, attainable, and specific over time. However, capturing the dynamic and sequential

Table 3
Estimated Mean and Standard Error for Each Adaptive Intervention

Adaptive intervention			Responders		Nonresponders		Estimated weighted mean	Robust *SE*
	Stage 1	Stage 2	Sample size	Sample mean	Sample size	Sample mean		
(1, −1)	BMOD	AUGMENT	22	4.64	24	4.08	4.36	0.15
(−1, −1)	MED	AUGMENT	36	4.39	17	3.47	4.00	0.15
(1, 1)	BMOD	INTENSIFY	22	4.64	22	3.96	4.17	0.22
(−1, 1)	MED	INTENSIFY	36	4.39	18	4.22	4.27	0.13

Table 4

Estimated Differences Between the Four Adaptive Interventions Based on the Estimated Regression Coefficients in Table 2

Label	Estimate	95% confidence limit		Robust SE	χ^2	Significance
		LL	UL			
Difference between $(1, -1)$ and $(-1, -1)$	0.36	-0.06	0.79	0.22	2.82	.0932
Difference between $(1, 1)$ and $(-1, 1)$	-0.10	-0.60	0.40	0.26	0.16	.6920
Difference between $(1, -1)$ and $(1, 1)$	0.19	-0.32	0.70	0.26	0.55	.4600
Difference between $(1, -1)$ and $(-1, 1)$	0.09	-0.31	0.49	0.20	0.20	.6563
Difference between $(-1, -1)$ and $(1, 1)$	-0.17	-0.70	0.35	0.27	0.42	.5161
Difference between $(-1, -1)$ and $(-1, 1)$	0.27	-0.08	0.63	0.18	2.26	.1328

Note. LL = lower limit; UL = upper limit.

nature of goal setting (in which both self-set goals and environmental feedback are incorporated into a system that monitors performance relative to the desired state and repeatedly adjusts subsequent goals; Campion & Lord, 1982) might require a SMART design with a high number of randomizations per participant. Accordingly, an important direction for future research would be developing methods for analyzing data from such designs.

Several other issues warrant further discussion in future research. First, the outcome of interest in the current article was a single measurement that expresses long-term school performance. However, researchers might be interested in other types of outcomes, such as noncontinuous outcomes (e.g., binary or count measures reflecting the severity of ADHD symptoms), multiple outcomes (e.g., a set of outcome measures reflecting different dimensions of school performance, such as peer relations, academic performance, and classroom performance), or longitudinal outcomes (e.g., improvement in children's classroom performance over the school year). The model used in the current study can be easily generalized via a generalized linear model to estimate the difference between adaptive interventions with respect to binary, categorical, or count outcomes. The SAS GENMOD procedure discussed in the current article allows the specification of various types of distributions and link functions. Still, additional research attention should be given to developing data analysis methods for the comparison of adaptive interventions with respect to other types of outcome measures. It is especially important to generalize our approach to estimate the difference between adaptive interventions with respect to time-varying outcomes by utilizing random effects as in growth-curve modeling (Hu, Goldberg, Hedeker, Flay, & Pentz, 1998).

Second, in a SMART, like other trials (Montgomery, Peters, & Little, 2003), there is concern about the loss of balance and subsequent loss of power due to study dropout and item missingness. Modern-day missing-data techniques will be needed in the analysis of data from SMART designs as in the case of any randomized trial. Like any other trials, these techniques will lead to unbiased results when the data are missing at random (MAR); that is, the probability that an observation is missing depends on information that is observed (i.e., characteristics of the child or baseline severity of ADHD symptoms). If data are missing not at random (MNAR)—that is, the probability that an observation is missing depends on information that is not observed—valuable information is lost from the data, and there is no universal method of handling the missing data properly (see Donders, van der Heijden, Stijnen, & Moons, 2006, for review).

Finally, in the current study we discussed primary research questions relating to the construction of adaptive interventions, but investigators often collect information concerning other potential moderators of the intervention options, aiming to use these measures to address secondary research questions concerning the best adaptive sequence of intervention options. For example, researchers might be interested in assessing whether both the first-stage intervention and the second-stage intervention should be more deeply tailored to the individual (perhaps the choice of the first-stage intervention option should depend on baseline severity, whereas the choice of the second-stage intervention option should depend on the participant's willingness to adhere to the first-stage intervention option). In this sense, additional information concerning potential moderators can be used to explore ways to more deeply tailor the adaptive intervention while using the advantages of the multiple randomizations in the SMART to eliminate confounding. In the companion article (Nahum-Shani et al., 2012), we provide a secondary data analysis method that can be used to find the best sequence of decision rules while evaluating potential tailoring variables at each intervention stage. This data analysis method takes a holistic approach—one that considers future intervention stages when assessing the best way to tailor the intervention at each specific stage.

References

Abele, A. E., & Wiese, B. S. (2008). The nomological network of self-management strategies and career success. *Journal of Occupational and Organizational Psychology, 81,* 733–749. doi:10.1348/096317907X256726

Baltes, P. B., & Baltes, M. M. (1990). Psychological perspectives on successful aging: The model of selective optimization with compensation. In P. B. Baltes & M. M. Baltes (Eds.), *Successful aging: Perspectives from the behavioural sciences* (pp. 1–34). Cambridge, MA: Cambridge University Press. doi:10.1017/CBO9780511665684.003

Barlas, Y., & Yasarcan, H. (2006). Goal setting, evaluation, learning and revision: A dynamic modeling approach. *Evaluation and Program Planning, 29,* 79–87. doi:10.1016/j.evalprogplan.2005.08.003

Beck, J. S., Liese, B. S., & Najavits, L. M. (2005). Cognitive therapy. In

R. J. Frances, S. I. Miller, & A. Mack (Eds.), *Clinical textbook of addictive disorders* (3rd ed., pp. 474–501). New York, NY: Guilford Press.

Bierman, K. L., Nix, R. L., Maples, J. J., & Murphy, S. A. (2006). Examining clinical judgment in an adaptive intervention design: The Fast Track Program. *Journal of Consulting and Clinical Psychology, 74,* 468–481. doi:10.1037/0022-006X.74.3.468

Campion, M. A., & Lord, R. G. (1982). A control systems conceptualization of the goal-setting and changing process. *Organizational Behavior and Human Performance, 30,* 265–287. doi:10.1016/0030-5073(82)90221-5

Cole, M. B. (2005). *Group dynamics in occupational therapy: The theoretical basis and practice application of group treatment* (3rd ed.). Thorofare, NJ: Slack.

Cole, S. R., & Hernán, M. A. (2008). Constructing inverse probability weights for marginal structural models. *American Journal of Epidemiology, 168,* 656–664. doi:10.1093/aje/kwn164

Collins, L. M., Dziak, J. J., & Li, R. (2009). Design of experiments with multiple independent variables: A resource management perspective on complete and reduced factorial designs. *Psychological Methods, 14,* 202–224. doi:10.1037/a0015826

Collins, L. M., Murphy, S. A., & Bierman, K. A. (2004). A conceptual framework for adaptive preventive interventions. *Prevention Science, 5,* 185–196. doi:10.1023/B:PREV.0000037641.26017.00

Conduct Problems Prevention Research Group. (1992). A developmental and clinical model for the prevention of conduct disorders: The FAST Track Program. *Development and Psychopathology, 4,* 509–527. doi:10.1017/S0954579400004855

Connell, A. M. (2009). Employing complier average causal effect analytic methods to examine effects of randomized encouragement trials. *The American Journal of Drug and Alcohol Abuse, 35,* 253–259. doi:10.1080/00952990903005882

Connell, A. M., & Dishion, T. J. (2008). Reducing depression among at-risk early adolescents: Three-year effects of a family-centered intervention embedded within schools. *Journal of Family Psychology, 22,* 574–585. doi:10.1037/0893-3200.22.3.574

Connell, A. M., Dishion, T. J., Yasui, M., & Kavanagh, K. (2007). An adaptive approach to family intervention: Linking engagement in family-centered intervention to reductions in adolescent problem behavior. *Journal of Consulting and Clinical Psychology, 75,* 568–579. doi:10.1037/0022-006X.75.4.568

Cook, T. D., & Campbell, D. T. (1979). *Quasiexperimentation: Design and analysis issues for field settings.* Chicago, IL: Rand McNally.

Dishion, T. J., & Kavanagh, K. (2003). *Intervening in adolescent problem behavior: A family-centered approach.* New York, NY: Guilford Press.

Donders, A. R. T., van der Heijden, G. J. M. G., Stijnen, T., & Moons, K. G. M. (2006). A gentle introduction to imputation of missing values. *Journal of Clinical Epidemiology, 59,* 1087–1091. doi:10.1016/j.jclinepi.2006.01.014

Dziak, J. J., Nahum-Shani, I., & Collins, L. M. (2012). Multilevel factorial experiments for developing behavioral interventions: Power, sample size, and resource considerations. *Psychological Methods, 17,* 153–175. doi:10.1037/a0026972

Efron, B. (1971). Forcing a sequential experiment to be balanced. *Biometrika, 58,* 403–417. doi:10.1093/biomet/58.3.403

Fabiano, G. A., Pelham, W. E., Waschbusch, D. A., Gnagy, E. M., Lahey, B. B., Chronis, A. M., . . . Burrows-MacLean, L. (2006). A practical measure of impairment: Psychometric properties of the impairment rating scale in samples of children with attention-deficit/hyperactivity disorder and two school-based samples. *Journal of Clinical Child and Adolescent Psychology, 35,* 369–385. doi:10.1207/s15374424jccp3503_3

Fava, M., Rush, A. J., Trivedi, M., Nierenberg, A. A., Thase, M. E., Sackeim, H., . . . Kupfer, D. J. (2003). Background and rationale for the sequenced treatment alternatives to relieve depression (STAR*D) study.

Psychiatric Clinics of North America, 26, 457–494. doi:10.1016/S0193-953X(02)00107-7

Fried, Y., & Slowik, L. H. (2004). Enriching goal-setting theory with time: An integrated approach. *The Academy of Management Review, 29,* 404–422.

Gopher, D. (1993). The skill of attention control: Acquisition and execution of attention strategies. In D. Meyer & S. Kornblum (Eds.), *Attention and Performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 299–322). Cambridge, MA: MIT Press.

Gunter, L., Zhu, J., & Murphy, S. A. (2011). Variable selection for qualitative interactions. *Statistical Methodology, 8,* 42–55. doi:10.1016/j.stamet.2009.05.003

Hedden, S. L., Woolson, R. F., & Malcolm, R. J. (2006). Randomization in substance abuse clinical trials. *Substance Abuse Treatment, Prevention, and Policy, 1,* 1–17. doi:10.1186/1747-597X-1-6

Hernán, M. A. (2004). A definition of causal effect for epidemiological research. *Journal of Epidemiology and Community Health, 58,* 265–271. doi:10.1136/jech.2002.006361

Hernán, M. A., Brumback, B., & Robins, J. M. (2000). Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology, 11,* 561–570. doi:10.1097/00001648-200009000-00012

Hernán, M. A., Brumback, B., & Robins, J. M. (2002). Estimating the causal effect of zidovudine on CD4 count with a marginal structural model for repeated measures. *Statistics in Medicine, 21,* 1689–1709. doi:10.1002/sim.1144

Hirschi, A., & Vondracek, F. W. (2009). Adaptation of career goals to self and opportunities in early adolescence. *Journal of Vocational Behavior, 75,* 120–128. doi:10.1016/j.jvb.2009.05.005

Hu, F. B., Goldberg, J., Hedeker, D., Flay, B. F., & Pentz, M. A. (1998). Comparison of population-averaged and subject-specific approaches for analyzing repeated binary outcomes. *American Journal of Epidemiology, 147,* 694–703. doi:10.1093/oxfordjournals.aje.a009511

Kernan, W. N., Viscoli, C. M., Makuch, R. W., Brass, L. M., & Horwitz, R. I. (1999). Stratified randomization for clinical trials. *Journal of Clinical Epidemiology, 52,* 19–26. doi:10.1016/S0895-4356(98)00138-3

Laurenceau, J.-P., Hayes, A. M., & Feldman, G. C. (2007). Statistical and methodological issues in the study of change in psychotherapy. *Clinical Psychology Review, 27,* 682–695. doi:10.1016/j.cpr.2007.01.007

Lavori, P. W., & Dawson, R. (2000). A design for testing clinical strategies: Biased individually tailored within-subject randomization. *Journal of the Royal Statistical Society, Series A: Statistics in Society, 163,* 29–38. doi:10.1111/1467-985X.00154

Lavori, P. W., & Dawson, R. (2004). Dynamic treatment regimes: Practical design considerations. *Clinical Trials, 1,* 9–20. doi:10.1191/1740774S04cn002oa

Lavori, P. W., Rush, A. J., Wisniewski, S. R., Alpert, J., Fava, M., Kupfer, D. J., . . . Trivedi, M. (2001). Strengthening clinical effectiveness trials: Equipoise-stratified randomization. *Biological Psychiatry, 50,* 792–801. doi:10.1016/S0006-3223(01)01223-9

Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika, 73,* 13–22. doi:10.1093/biomet/73.1.13

Locke, E. A., & Latham, P. G. (1990). *A theory of goal setting and task performance.* Englewood Cliffs, NJ: Prentice-Hall.

Lunceford, J. K., Davidian, M., & Tsiatis, A. A. (2002). Estimation of survival distributions of treatment policies in two-stage randomization designs in clinical trials. *Biometrics, 58,* 48–57. doi:10.1111/j.0006-341X.2002.00048.x

Marlowe, D. B., Festinger, D. S., Arabia, P. L., Dugosh, K. L., Benasutti, K. M., Croft, J. R., & McKay, J. R. (2008). Adaptive interventions in drug court: A pilot experiment. *Criminal Justice Review, 33,* 343–360. doi:10.1177/0734016808320325

Martocchio, J. J., & Webster, J. (1992). Effects of feedback and cognitive playfulness on performance in microcomputer software training. *Personnel Psychology, 45,* 553–578. doi:10.1111/j.1744-6570.1992.tb00860.x

McKay, J. R. (2005). Is there a case for extended interventions for alcohol and drug use disorders? *Addiction, 100,* 1594–1610. doi:10.1111/j.1360-0443.2005.01208.x

Montgomery, A. A., Peters, T. J., & Little, P. (2003). Design, analysis and presentation of factorial randomized controlled trials. *BMC Medical Research Methodology, 3,* 26. doi:10.1186/1471-2288-3-26

Moore, R., & Blackburn, I. (1997). Cognitive therapy in the treatment of nonresponders to antidepressant medication: A controlled pilot study. *Behavioural and Cognitive Psychotherapy, 25,* 251–259. doi:10.1017/S1352465800018543

Murphy, S. A. (2005). An experimental design for the development of adaptive treatment strategies. *Statistics in Medicine, 24,* 1455–1481. doi:10.1002/sim.2022

Murphy, S. A., Collins, L. M., & Rush, A. J. (2007). Customizing treatment to the patient: Adaptive treatment strategies [Editorial]. *Drug and Alcohol Dependence, 88*(Suppl. 2), S1–S3. doi:10.1016/j.drugalcdep.2007.02.001

Murphy, S. A., van der Laan, M. J., Robins, J. M., & Conduct Problems Prevention Research Group. (2001). Marginal mean models for dynamic regimes. *Journal of the American Statistical Association, 96,* 1410–1423. doi:10.1198/016214501753382327

Nahum-Shani, I., Qian, M., Almirall, D., Pelham, W. E., Gnagy, B., Fabiano, G. A., . . . Murphy, S. A. (2012). Q-learning: A data analysis method for constructing adaptive interventions. *Psychological Methods,17,* 478–494. doi:10.1037/a0029373

Oetting, A. I., Levy, J. A., Weiss, R. D., & Murphy, S. A. (2007). Statistical methodology for a SMART design in the development of adaptive treatment strategies. In P. E. Shrout (Ed.), *Causality and psychopathology: Finding the determinants of disorders and their cures* (pp. 179–205). Arlington, VA: American Psychiatric Publishing.

Orellana, L., Rotnitzky, A., & Robins, J. M. (2010). Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, Part II: Proofs of results. *The International Journal of Biostatistics, 6*(2). doi:10.2202/1557-4679.1242

Pelham, W. E., Evans, S. W., Gnagy, E. M., & Greenslade, K. E. (1992). Teacher ratings of *DSM–III–R* symptoms for the disruptive behavior disorders: Prevalence, factor analyses, and conditional probabilities in a special education sample. *School Psychology Review, 21,* 285–299.

Pelham, W. E., Hoza, B., Pillow, D. R., Gnagy, E. M., Kipp, H. L., Greiner, A. R., . . . Fitzpatrick, E. (2002). Effects of methylphenidate and expectancy on children with ADHD: Behavior, academic performance, and attributions in a summer treatment program and regular classroom setting. *Journal of Consulting and Clinical Psychology, 70,* 320–335. doi:10.1037/0022-006X.70.2.320

Robins, J. (1986). A new approach to causal inference in mortality studies with sustained exposure periods: Application to control of the healthy worker survivor effect. *Computers and Mathematics with Applications, 14,* 1393–1512.

Robins, J., Orellana, L., & Rotnitzky, A. (2008). Estimation and extrapolation of optimal treatment and testing strategies. *Statistics in Medicine, 27,* 4678–4721. doi:10.1002/sim.3301

Robins, J. M., & Wasserman, L. (1997). Estimation of effects of sequential treatments by reparameterizing directed acyclic graphs. In D. Geiger & P. Shenoy (Eds.), *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence* (pp. 409–420). San Francisco, CA: Kaufmann.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and non-randomized studies. *Journal of Educational Psychology, 66,* 688–701. doi:10.1037/h0037350

Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *Annals of Statistics, 6,* 34–58. doi:10.1214/aos/1176344064

Sackett, D. L., Rosenberg, W. M. C., Gray, J. A. M., Haynes, R. B., & Richardson, W. S. (1996). Evidence-based medicine: What it is and what it isn't. *British Medical Journal, 312,* 71–72. doi:10.1136/bmj.312.7023.71

SAS Institute. (2008). *SAS/STAT 9.2 user's guide.* Cary, NC: Author.

Schneider, L. S., Tariot, P. N., Lyketsos, C. G., Dagerman, K. S., Davis, K. L., Davis, S., . . . Lieberman, J. A. (2001). National Institute of Mental Health Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE): Alzheimer disease trial methodology. *The American Journal of Geriatric Psychiatry, 9,* 346–360.

Shadish, W. R. (2002). Revisiting field experimentation: Field notes for the future. *Psychological Methods, 7,* 3–18. doi:10.1037/1082-989X.7.1.3

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston, MA: Houghton Mifflin.

Shani, G. (2007). *A training approach to bridge the gap between wizards and macro users when working with computers* (Unpublished master's thesis). Department of Industrial Engineering and Management, Technion: Israel Institute of Technology, Haifa, Israel.

Shortreed, S. M., Laber, E., Pineau, J., & Murphy, S. A. (2010). *Imputation methods for the Clinical Antipsychotic Trials of Intervention and Effectiveness study* (Technical Report No. SOCS-TR-2010.8). School of Computer Science, McGill University, Montreal, Quebec, Canada.

Simon, R. (1979). Restricted randomization designs in clinical trials. *Biometrics, 35,* 503–512. doi:10.2307/2530354

Splawa-Neyman, J. (1990). On the application of probability theory to agricultural experiments: Essay on principles. Section 9. *Statistical Science, 5,* 465–480. (Reprinted from *Roczniki Nauk Rolniczych Tom X,* pp. 1–51, by D. M. Dabrowska & T. P. Speed, Eds., 1923)

Stone, R. M., Berg, D. T., George, S. L., Dodge, R. K., Paciucci, P. A., Schulman, P., . . . Schiffer, C. A. (1995). Granulocyte-macrophage colony-stimulating factor after initial chemotherapy for elderly patients with primary acute myelogenous leukemia. *The New England Journal of Medicine, 332,* 1671–1677. doi:10.1056/NEJM199506223322503

Stroup, T. S., McEvoy, J. P., Swartz, M. S., Byerly, M. J., Glick, I. D., Canive, J. M., . . . Lieberman, J. A. (2003). National Institute of Mental Health Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE): Schizophrenia trial design and protocol development. *Schizophrenia Bulletin, 29,* 15–31. doi:10.1093/oxfordjournals.schbul.a006986

Thall, P. F., Logothetis, C., Pagliaro, L. C., Wen, S., Brown, M. A., Williams, D., & Millikan, R. E. (2007). Adaptive therapy for androgen-independent prostate cancer: A randomized selection trial of four regimens. *Journal of the National Cancer Institute, 99,* 1613–1622. doi:10.1093/jnci/djm189

Thall, P. F., Millikan, R. E., & Sung, H. G. (2000). Evaluating multiple treatment courses in clinical trials. *Statistics and Medicine, 19,* 1011–1028. doi:10.1002/(SICI)1097-0258(20000430)19:8<1011::AID-SIM414>3.0.CO;2-M

Thall, P. F., Sung, H. G., & Estey, E. H. (2002). Selecting therapeutic strategies based on efficacy and death in multicourse clinical trials. *Journal of the American Statistical Association, 97,* 29–39. doi:10.1198/016214502753479202

Thall, P. F., & Wathen, J. K. (2005). Covariate-adjusted adaptive randomization in a sarcoma trial with multi-stage treatments. *Statistics in Medicine, 24,* 1947–1964. doi:10.1002/sim.2077

Wahed, A. S., & Tsiatis, A. A. (2004). Optimal estimator for the survival distribution and related quantities for treatment policies in two-stage randomization designs in clinical trials. *Biometrics, 60,* 124–133. doi:10.1111/j.0006-341X.2004.00160.x

Wahed, A. S., & Tsiatis, A. A. (2006). Semiparametric efficient estimation of survival distribution for treatment policies in two-stage randomization

designs in clinical trials with censored data. *Biometrika, 93,* 163–177. doi:10.1093/biomet/93.1.163

Walsh, B. T., Agras, W. S., Devlin, M. J., Fairburn, C. G., Wilson, G. T., Kahn, C., & Chally, M. K. (2000). Fluoxetine for bulimia nervosa following poor response to psychotherapy. *The American Journal of Psychiatry, 157,* 1332–1334. doi:10.1176/appi.ajp.157.8.1332

Weisz, J. R., Chu, B. C., & Polo, A. J. (2004). Treatment dissemination and evidence-based practice: Strengthening interventions through clinician–researcher collaboration. *Clinical Psychology: Science and Practice, 11,* 300–307. doi:10.1093/clipsy.bph085

Yalom, I. D. (1995). *The theory and practice of group psychotherapy* (4th ed.). New York, NY: Basic Books.

Appendix A

Consistency of Estimators and Justification for Using Robust Standard Errors

In Equation 1 we provided a method for estimating the mean of the primary outcome, conditional on baseline measures, for each adaptive intervention (a_1, a_2). This method employs random weights (e.g., the weight depends on response/nonresponse) and also replicates the data of the responders. Thus it is not immediately obvious that the estimator of the mean of the primary outcome for each adaptive intervention is consistent (unbiased in large samples). To demonstrate that this is indeed the case we define the mean of primary outcome, conditional on baseline measures, for each adaptive intervention (a_1, a_2) using the language of potential outcomes. Subsequently we demonstrate that this conditional mean can be consistently estimated by minimizing the sum of squared error given in Equation 2. Last, we verify that the robust standard errors provided by SAS GENMOD are indeed unbiased (consistent) in large samples.

Recall that (O_1, A_1, O_2, A_2, Y) denotes a participant's data in a SMART study. Here and following, uppercase letters represent random variables, and lowercase letters represent particular values, or realizations, of random variables. Recall that O_1 denotes a vector of baseline measures, (A_1, A_2) denotes the randomized intervention options, and O_2 is a vector of intermediate outcomes. One intermediate outcome is R, the indicator for response (other intermediate outcomes might be measures of adherence to first-stage intervention, other measures of severity, etc.). Only nonresponding participants have an A_2. A particular adaptive intervention is denoted by $(A_1, A_2) = (a_1, a_2)$ (in the ADHD example both a_1 and a_2 are either 1 or -1). We write that a participant in a SMART study "follows" adaptive intervention (a_1, a_2) if this participant's first-stage intervention is a_1 and either this participant responds throughout the study ($R = 1$) or, if the participant does not respond, then the participant's second-stage intervention is a_2. Thus, for example, in the ADHD study, participants who respond to first-stage intervention a_3 follow both adaptive interventions $(a_1, 1)$ and $(a_1, -1)$.

In the potential outcomes framework (Robins, 1986; Rubin, 1978; Splawa-Neyman, 1923/1990) and for each fixed sequence of interventions (a_1, a_2), we conceptualize potential outcomes denoted by $O_2(a_1)$ and $Y(a_1, a_2)$. $O_1(a_1)$ is the vector of intermediate observations that a participant would have had at the second stage if he/she had followed a_1 at the first-stage intervention; $O_2(a_1)$

includes the indicator of response $R(a_1)$. $Y(a_1, a_2)$ is the primary outcome that would have been observed had the participant followed the sequence (a_1, a_2). In the ADHD study responding participants do not enter the second-stage intervention; thus, we have that when $R(a_1) = 1$, $Y(a_1, 1) = Y(a_1, -1)$. $O_2(a_1)$ and $Y(a_1, a_2)$ are called *potential* outcomes because they describe an individual outcome value that would have been observed under a potential exposure to the intervention that the individual might or might not have actually experienced. The assumption is that each intervention option has the potential to be observed until assignment to intervention options is made (for a more detailed definition of potential outcomes, see Rubin, 1974, as well as Hernán, 2004). The set of all potential outcomes is

$$O = \{(O_2(a_1), Y(a_1, a_2)): a_1 \in \{-1, 1\}, a_2 \in \{-1, 1\}\}.$$

By definition, the multivariate distribution of $(O_1, O_2(a_1), Y(a_1, a_2))$ is the multivariate distribution of (O_1, O_2, Y) when all individuals follow adaptive intervention (a_1, a_2). Denote $E[Y(a_1, a_2)|O_1]$ by $\mu(a_1, a_2|O_1)$; we know that $\mu(a_1, a_2|O_1)$ minimizes the total mean squared error for the four adaptive interventions:

$$E[(Y(-1, -1) - \mu(-1, -1|O_1))^2]$$
$$+ E[(Y(-1, 1) - \mu(-1, 1|O_1))^2] + E[(Y(1, 1) - \mu(1, 1|O_1))^2]$$
$$+ E[(Y(1, -1) - \mu(1, -1|O_1))^2].$$

Assume that $\mu(a_1, a_2|O_1) = \beta_0 + \beta_1 a_1 + \beta_2 a_2 + \beta_3 a_1 a_2 + \gamma^T O_1$. The above can be rewritten as

$$E[(Y(-1, -1) - \beta_0 + \beta_1 + \beta_2 - \beta_3 - \gamma^T O_1)^2]$$
$$+ E[(Y(-1, 1) - \beta_0 + \beta_1 - \beta_2 + \beta_3 - \gamma^T O_1)^2] + E[(Y(1, 1) - \beta_0$$
$$- \beta_1 - \beta_2 - \beta_3 - \gamma^T O_1)^2] + E[(Y(1, -1) - \beta_0 - \beta_1 + \beta_2 + \beta_3$$
$$- \gamma^T O_1)^2]. \tag{3}$$

Our goal is to show that the weighted least squares estimator obtained by minimizing Equation 2, $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \widehat{\gamma^T})$ is a consistent (unbiased in large samples) estimator of $(\beta_0, \beta_1, \beta_2, \beta_3, \gamma^T)$. We do this by showing that Equation 3 is the large-sample version of Equation 2 up to a constant factor.

(Appendices continue)

Since $Y(a_1, a_2) = R(a_1)Y(a_1, a_2) + (1 - R(a_1))Y(a_1, a_2)$, Equation 3 can be rewritten as

$$\sum_{a_1 \in \{1,-1\}} \sum_{a_2 \in \{1,-1\}} E[(R(a_1)Y(a_1, a_2) + (1 - R(a_1))Y(a_1, a_2) - \beta_0$$
$$- \beta_1 a_1 - \beta_2 a_2 - \beta_3 a_1 a_2 - \boldsymbol{\gamma}^T O_1)^2].$$

Splitting the squared terms by the indicators, $R(a_1)$ and $1 - R(a_1)$, we obtain that Equation 3 is equal to

$$\sum_{a_1 \in \{1,-1\}} \sum_{a_2 \in \{1,-1\}} E[R(a_1)(Y(a_1, a_2) - \beta_0 - \beta_1 a_1 - \beta_2 a_2 - \beta_3 a_1 a_2$$
$$- \boldsymbol{\gamma}^T O_1)^2]$$

$$+ \sum_{a_1 \in \{1,-1\}} \sum_{a_2 \in \{1,-1\}} E[(1 - R(a_1))(Y(a_1, a_2) - \beta_0 - \beta_1 a_1 - \beta_2 a_2$$
$$- \beta_3 a_1 a_2 - \boldsymbol{\gamma}^T O_1)^2]. \quad (4)$$

Since A_1, A_2 are randomized independently of the potential outcomes, Equation 4 can be expressed as

$$\sum_{a_2 \in \{1,-1\}} E[2R(A_1)(Y(A_1, a_2) - \beta_0 - \beta_1 A_1 - \beta_2 a_2 - \beta_3 A_1 a_2$$
$$- \boldsymbol{\gamma}^T O_1)^2] + E[4(1 - R(A_1))(Y(A_1, A_2) - \beta_0 - \beta_1 A_1 - \beta_2 A_2$$
$$- \beta_3 A_1 A_2 - \boldsymbol{\gamma}^T O_1)^2].$$

(To see this, note that if we average out over A_1 in the first two terms, the weight 2 cancels out the randomization probability .5 and we get an additional sum over a_1; next if we average out over both A_1, A_2 in the last term, the weight 4 cancels out the product of the randomization probabilities .5*.5 and we get a double sum over a_1, a_2.)

Next we express Equation 4 and hence Equation 3 in terms of the observed data (O_1, A_1, O_2, A_2, Y) (recall that O_2 includes the response indicator R). We assume that Robins's consistency assumption holds, namely, that the observed outcome for a participant equals the participant's potential outcome, when the intervention is set to the observed intervention option (see Robins & Wasserman, 1997). Accordingly, the potential outcomes are connected to the SMART data by $R = R(A_1)$ and $= Y(A_1, A_2)$; because responders are not provided a second-stage intervention, A_2, $R(A_1)Y(A_1, 1) = R(A_1)Y(A_1, -1)$. The consistency assumption implies that Equation 4 is equal to

$$E[2R(Y - \beta_0 - \beta_1 A_1 - \beta_2 - \beta_3 A_1 - \boldsymbol{\gamma}^T O_1)^2] + E[2R(Y - \beta_0$$
$$- \beta_1 A_1 + \beta_2 + \beta_3 A_1 - \boldsymbol{\gamma}^T O_1)^2] + E[4(1 - R)(Y - \beta_0 - \beta_1 A_1$$
$$- \beta_2 A_2 - \beta_3 A_1 A_2 - \boldsymbol{\gamma}^T O_1)^2].$$

However, the above display is the large-sample version of Equation 2 up to a constant factor. The number 2 in the first two terms is the weight given to responders and the number 4 in the third term is the weight given to nonresponders in Equation 2. Thus under mild regularity conditions, $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \widehat{\boldsymbol{\gamma}^T})$ is a consistent estimator (unbiased in large samples) of the regression coefficients in $E[Y(a_1, a_2)|O_1]$. That is, $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \widehat{\boldsymbol{\gamma}^T})$ is a consistent estimator of $(\beta_0, \beta_1, \beta_2, \beta_3, \boldsymbol{\gamma}^T)$.

Last, we provide a sketch of the proof that the robust standard error (sandwich standard error) provided by the SAS GENMOD procedure is a consistent estimator of the variance-covariance of the asymptotic normal distribution of $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \widehat{\boldsymbol{\gamma}^T})$. Note that Equation 2 is equal to

$$\sum_{i=1}^{N} [2R_i(Y_i - \beta_0 - \beta_1 A_{1i} - \beta_2 - \beta_3 A_{1i} - \boldsymbol{\gamma}^T O_{1i})^2]$$

$$+ \sum_{i=1}^{N} [2R_i(Y_i - \beta_0 - \beta_1 A_{1i} + \beta_2 + \beta_3 A_{1i} - \boldsymbol{\gamma}^T O_{1i})^2]$$

$$+ \sum_{i=1}^{N} [4(1 - R_i)(Y_i - \beta_0 - \beta_1 A_{1i} + \beta_2 A_{2i} + \beta_3 A_{1i} A_{2i} - \boldsymbol{\gamma}^T O_{1i})^2],$$

where N is the number of subjects in the data.

Since $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \widehat{\boldsymbol{\gamma}^T})$ minimizes Equation 2, the derivative of Equation 2 evaluated at $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \widehat{\boldsymbol{\gamma}^T})$ is zero. That is, $\sum_{i=1}^{N} U_i(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \widehat{\boldsymbol{\gamma}^T}) = 0$, where

$$U_i(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \widehat{\boldsymbol{\gamma}^T}) = 4R_i(Y_i - \hat{\beta}_0 - \hat{\beta}_1 A_{1i} - \hat{\beta}_2 - \hat{\beta}_3 A_{1i}$$
$$- \widehat{\boldsymbol{\gamma}^T} O_{1i})(1, A_{1i}, 1, A_{1i}, O_{1i}^T)^T + 4R_i(Y_i$$
$$- \hat{\beta}_0 - \hat{\beta}_1 A_{1i} + \hat{\beta}_2 + \hat{\beta}_3 A_{1i} - \widehat{\boldsymbol{\gamma}^T} O_{1i})$$
$$\times (1, A_{1i}, -1, -A_{1i}, O_{1i}^T)^T + 8(1 - R_i)(Y_i$$
$$- \hat{\beta}_0 - \hat{\beta}_1 A_{1i} - \hat{\beta}_2 A_{2i} - \hat{\beta}_3 A_{1i} A_{2i}$$
$$- \widehat{\boldsymbol{\gamma}^T} O_{1i}) \times (1, A_{1i}, A_{2i}, A_{1i} A_{2i}, O_{1i}^T)^T$$

for $i = 1, \ldots, N$.

Recall that responders are replicated and get a weight of 2, and nonresponders get a weight of 4. The sum of the weights is $\sum_{i=1}^{N} [2R_i + 2R_i + 4(1 - R_i) = 4N]$. Define $Z_N = 4N (\hat{\beta}_0 - \beta_0, \hat{\beta}_1 - \beta_1, \hat{\beta}_2 - \beta_2, \hat{\beta}_3 - \beta_3, \boldsymbol{\gamma}^T - \boldsymbol{\gamma}^T)^T$. A Taylor series for $0 = \sum_{i=1}^{N} U_i(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \widehat{\boldsymbol{\gamma}^T})$ in $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \widehat{\boldsymbol{\gamma}^T})$ around $(\beta_0, \beta_1, \beta_2, \beta_3, \boldsymbol{\gamma}^T)$ results in

$$0 = \frac{1}{\sqrt{4N}} \sum_{i=1}^{N} U_i(\beta_0, \beta_1, \beta_2, \beta_3, \boldsymbol{\gamma}^T) + \frac{1}{4N}$$
$$\sum_{i=1}^{N} \frac{\partial U_i(\beta_0, \beta_1, \beta_2, \beta_3, \boldsymbol{\gamma}^T)}{\partial(\beta_0, \beta_1, \beta_2, \beta_3, \boldsymbol{\gamma}^T)} Z_N.$$

Solve for Z_N to obtain

$$Z_N = \left(-\frac{1}{4N} \sum_{i=1}^{N} \frac{\partial U_i(\beta_0, \beta_1, \beta_2, \beta_3, \boldsymbol{\gamma}^T)}{\partial(\beta_0, \beta_1, \beta_2, \beta_3, \boldsymbol{\gamma}^T)} \right)^{-1}$$
$$\frac{1}{\sqrt{4N}} \sum_{i=1}^{N} U_i(\beta_0, \beta_1, \beta_2, \beta_3, \boldsymbol{\gamma}^T).$$

(*Appendices continue*)

Thus, the large-sample variance–covariance matrix of Z_N is given by $I_0^{-1} I_1 I_0^{-1}$ where

$$I_0 = 4E[R(1, A_1, 1, A_1, O_1^T)^T (1, A_1, 1, A_1, O_1^T) + R(1, A_1, -1,$$
$$-A_1, O_1^T)^T (1, A_1, -1, -A_1, O_1^T) + (1-R)$$
$$\times (1, A_1, A_2, A_1 A_2, O_1^T)^T (1, A_1, A_2, A_1 A_2, O_1^T)]$$

$$\left(I_0 = -E\left[\frac{\partial U_i(\beta_0, \beta_1, \beta_2, \beta_3, \boldsymbol{\gamma}^T)}{\partial (\beta_0, \beta_1, \beta_2, \beta_3, \boldsymbol{\gamma}^T)} \right] \right)$$

and

$$I_1 = 16E[R(Y - \beta_0 - \beta_1 A_1 - \beta_2 - \beta_3 A_1$$
$$- \boldsymbol{\gamma}^T O_1)^2 (1, A_1, 1, A_1, O_1^T)^T (1, A_1, 1, A_1, O_1^T) + R(Y - \beta_0$$
$$- \beta_1 A_1 + \beta_2 + \beta_3 A_1 - \boldsymbol{\gamma}^T O_1)^2 (1, A_1, -1,$$
$$- A_1, O_1^T)^T (1, A_1, -1, -A_1, O_1^T) + R(Y - \beta_0 - \beta_1 A_1$$
$$- \beta_2 - \beta_3 A_1 - \boldsymbol{\gamma}^T O_1)(Y - \beta_0 - \beta_1 A_1 + \beta_2 + \beta_3 A_1 - \boldsymbol{\gamma}^T O_1)$$
$$\times [(1, A_1, 1, A_1, O_1^T)^T (1, A_1, -1, -A_1, O_1^T) + (1, A_1, -1,$$
$$- A_1, O_1^T)^T (1, A_1, 1, A_1, O_1^T)] + 4(1-R)(Y - \beta_0 - \beta_1 A_1$$
$$- \beta_2 A_2 - \beta_3 A_1 A_2 - \boldsymbol{\gamma}^T O_1)^2$$
$$\times (1, A_1, A_2, A_1 A_2, O_1^T)^T (1, A_1, A_2, A_1 A_2, O_1^T)],$$

where I_1 is the variance–covariance matrix of $U_i(\beta_0, \beta_1, \beta_2, \beta_3, \boldsymbol{\gamma}^T)$. Replacing the expectation with the weighted sample average (i.e., replacing E with $\frac{1}{4N} \Sigma_{i=1}^{N}$.) and plugging in the weighted least squares estimator, we obtain an estimator of the variance–covariance matrix of $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \widehat{\boldsymbol{\gamma}^T})$, namely, $\frac{1}{4N} \hat{I}_0^{-1} \hat{I}_1 \hat{I}_0^{-1}$, where

$$\hat{I}_0 = \frac{1}{N} \sum_{i=1}^{N} [R_i(1, A_{1i}, 1, A_{1i}, O_{1i}^T)^T (1, A_{1i}, 1, A_{1i}, O_{1i}^T) + R_i(1, A_{1i},$$
$$- 1, -A_{1i}, O_{1i}^T)^T (1, A_{1i}, -1, -A_{1i}, O_{1i}^T) + 2(1-R_i)$$
$$\times (1, A_{1i}, A_{2i}, A_{1i} A_{2i}, O_{1i}^T)^T (1, A_{1i}, A_{2i}, A_{1i} A_{2i}, O_{1i}^T)]$$

and

$$\hat{I}_1 = \frac{4}{N} \sum_{i=1}^{N} [R_i(Y_i - \hat{\beta}_0 - \hat{\beta}_1 A_{1i} - \hat{\beta}_2 - \hat{\beta}_3 A_{1i}$$
$$- \widehat{\boldsymbol{\gamma}^T} O_{1i})^2 (1, A_{1i}, 1, A_{1i}, O_{1i}^T)^T (1, A_{1,i}, 1, A_{1i}, O_{1i}^T) + R_i(Y_i - \hat{\beta}_0$$
$$- \hat{\beta}_1 A_{1i} + \hat{\beta}_2 + \hat{\beta}_3 A_{1i} - \widehat{\boldsymbol{\gamma}^T} O_{1i})^2 (1, A_{1i}, -1,$$
$$- A_{1i}, O_{1i}^T)^T (1, A_{1i}, -1, -A_{1i}, O_{1i}^T) + R_i(Y_i - \hat{\beta}_0 - \hat{\beta}_1 A_{1i} + \hat{\beta}_2$$
$$- \hat{\beta}_3 A_{1i} - \widehat{\boldsymbol{\gamma}^T} O_{1i})(Y_i - \hat{\beta}_0 - \hat{\beta}_1 A_{1i} + \hat{\beta}_2 + \hat{\beta}_3 A_{1i} - \widehat{\boldsymbol{\gamma}^T} O_{1i})$$
$$\times [(1, A_{1i}, 1, A_{1i}, O_{1i}^T)^T (1, A_{1i} - 1, -A_{1i}, O_{1i}^T) + (1, A_{1i}, -1,$$
$$- A_{1i}, O_{1i}^T)^T (1, A_{1i}, 1, A_{1i}, O_{1i}^T)] + 4(1 - R_i(Y_i - \hat{\beta}_0 - \hat{\beta}_1 A_{1i}$$
$$- \hat{\beta}_2 A_{2i} - \hat{\beta}_3 A_{1i} A_{2i} - \widehat{\boldsymbol{\gamma}^T} O_{1i})^2$$
$$\times (1, A_{1i}, A_{2i}, A_{1i} A_{2i}, O_{1i}^T)^T (1, A_{1i}, A_{2i}, A_{1i} A_{2i}, O_{1i}^T)].$$

(*Appendices continue*)

Appendix B

SAS Syntax for Estimating Equation 1

The following SAS syntax can be used to estimate the weighted regression in Equation 1 (Hernán, Brumback, & Robins, 2000, 2002):

```
PROC GENMOD data=new;

CLASS id;

    MODEL outcome = a1 a2 a1*a2 o1;

    SCWGT weights;

    REPEATED SUBJECT=id / TYPE=ind;

    ESTIMATE 'difference 1, −1 and −1, −1' a1 2 a2 0 a1*a2 −2;

    ESTIMATE 'difference 1, 1 and −1, 1' a1 2 a2 0 a1*a2 2;

    ESTIMATE 'difference 1, −1 and 1, 1' a1 0 a2 −2 a1*a2 −2;

    ESTIMATE 'difference 1, −1 and −1, 1' a1 2 a2 −2 a1*a2 0;

    ESTIMATE 'difference −1, −1 and 1, 1' a1 −2 a2 −2 a1*a2 0;

    ESTIMATE 'difference −1, 1 and −1, −1' a1 0 a2 −2 a1*a2 2;

RUN;
```

The first line calls the GENMOD procedure and specifies the working data set. The second line specifies the child's identification number (id) as a categorical variable. The third line specifies the model to be estimated, identifying the outcome and the independent variables. The fourth line specifies the variable weights in the data set as the variable that contains the weight for each observation, using the SCWGT statement. The program will provide robust standard errors for the model parameters when the option REPEATED is included, with the child's identification number (SUBJECT=id) and the independent working correlation matrix (TYPE=ind) being specified (see Liang & Zeger, 1986). The ESTIMATE statements specify the comparisons between the adaptive interventions based on the estimated coefficients.