Published in final edited form as:

Am J Eval. 2016 September ; 37(3): 344–363. doi:10.1177/1098214015617013.

Adaptive Interventions and SMART Designs: Application to child behavior research in a community setting

Kelley M. Kidwell and

Department of Biostatistics, University of Michigan

Luke W. Hyde

Department of Psychology, Center for Human Growth and Development, Institute for Social Research, University of Michigan

Abstract

Heterogeneity between and within people necessitates the need for sequential personalized interventions to optimize individual outcomes. Personalized or adaptive interventions (AIs) are relevant for diseases and maladaptive behavioral trajectories when one intervention is not curative and success of a subsequent intervention may depend on individual characteristics or response. AIs may be applied to medical settings and to investigate best prevention, education, and community-based practices. AIs can begin with low-cost or low-burden interventions and followed with intensified or alternative interventions for those who need it most. AIs that guide practice over the course of a disease, program, or school year can be investigated through sequential multiple assignment randomized trials (SMARTs). To promote the use of SMARTs, we provide a hypothetical SMART in a Head Start program to address child behavior problems. We describe the advantages and limitations of SMARTs, particularly as they may be applied to the field of evaluation.

Keywords

sequential multiple assignment randomized trial; adaptive intervention; Head Start; behavior problems; feedback; parent training

Public health, medical, and behavioral health intervention research is often examined using randomized controlled trials (RCTs) because they control for potential sources of endogenous influence (i.e., by randomizing we have more confidence that the intervention caused any changes measured; Kleijnen, Gotzsche, Kunz, Oxman, & Chalmers, 1997). Though researchers debate the role of RCTs within the field of evaluation (Donaldson, Christie, & Mark, 2009), RCTs provide the strongest scientific evidence when they can be implemented (Chalmers, 2003; Kunz & Oxman, 1998). Furthermore, RCTs can be seen as one important component of an approach to evaluation (Gambrill, 2006) that balances the strengths and weaknesses of various evaluation approaches in terms of internal and external validity (Gone, 2015). One particular issue with many RCTs, however, specifically in public health research, is that RCTs often investigate research problems with a “one-size fits all” approach to universal prevention or treatment. Researchers have increasingly acknowledged that single interventions may be limited in their effectiveness due to heterogeneity in

treatment response between individuals or within an individual over time (Lei, Nahum-Shani, Lynch, Oslin, & Murphy, 2012; O'Brien Caughy, O'Campo, & Brodsky, 1999; Pachter, 2006). Thus, a major next step in medical, educational and behavioral research is to identify ways to rigorously investigate effective interventions over time in a more personalized way that match interventions to individuals through novel trial design (Lavori, Dawson, & Rush, 2004; Murphy, 2005). It is important to understand how advances in RCTs, particularly cutting-edge RCT designs, may offer new avenues for evaluation.

In practice, interventionists and individuals themselves often use multiple interventions one after another to improve health and well-being. For an individual, interventions may change over time based on response to previous interventions or other characteristics or behaviors. Intervention changes may include using different doses or intensities (e.g., if exercising once per week does not lead to weight loss, the individual may increase to three times per week) or switching to another intervention (e.g., adding healthy eating to exercise). Unfortunately, intervention research, particularly as evaluated by RCTs, has typically focused on a single stage of intervention and not the entire sequence of interventions. Standard RCTs generally ignore that, in practice, one intervention may be one part of a chain of interventions and that the entire series of interventions and their order may affect outcomes dramatically (e.g., starting with exercise may increase or decrease the subsequent effectiveness of healthy eating on weight loss). In many "real life" settings, most intervention and program implementation is not done in a vacuum and interventions are often sequenced and tailored to the individual. An RCT design, the *sequential multiple assignment randomized trial* (SMART, Lavori & Dawson, 2000; Murphy, 2005), has been developed to examine tailored sequential interventions that more closely mimic real-world practice. SMARTs can be used to examine effective sequences of interventions outside of treating diseases and illnesses and test optimal practices in prevention, education, and community-based programs. Thus, a SMART design is an type of RCT applicable to the field of evaluation.

SMARTs investigate sequential strategies where sequential strategies are relevant in many circumstances when individuals vary in their response to interventions, effectiveness of the intervention changes over time, other problems coexist and may need to be addressed in the same intervention process, and when relapse is common or behavior change is difficult to maintain over time (Lei et al., 2012). Sequential strategies or sequences of interventions that are tailored to individuals are referred to as *adaptive interventions* (AIs, also known as *dynamic treatment regimes*, Lavori, Dawson, & Rush, 2004; J. M. Robins, 1986, or *adaptive treatment strategies*, Dawson & Lavori, 2004). AIs were originally applied to areas of mental health (e.g. McKay, 2005; Bierman, Nix, Maples, Murphy, & Conduct Problems Prevention Research Group, 2006; Marlowe et al., 2008) and oncology (e.g. Auyeung et al., 2009; Thall, Millikan, & Sung, 2000). However, the application of AIs is broad in that they can guide any approach that includes two or more interventions used in sequence when subsequent intervention depends on response to the previous intervention or the ordering of interventions may affect overall effectiveness. For example, AIs have also been applied to guide the effective use of a judicial supervision system for drug-abusing defenders (Marlowe, Festinger, Dugosh, Lee, & Benasutti, 2007), effective interventions for weight loss (Almirall, Nahum-Shani, Sherwood, & Murphy, 2014), the timing and type of programs

used to prevent adolescent conduct problems (August, Piehler, & Bloomquist, 2014), and the implementation of a mood disorder program in outpatient clinics (Kilbourne et al., 2014).

An interventionist or program has AIs or guidelines that adapt based on individual characteristics or behaviors in mind over the course of an intervention process. For example, a teacher teaches students in a particular way at the start of the school year, but if many of the students perform poorly on the first test, the teacher may change his or her teaching approach to improve student performance. An AI contains *critical decision points*, *intervention components*, *tailoring variables*, and *decision rules* (Lei et al., 2012). *Critical decisions* include which intervention to begin with, when and how to measure signs of response/non-response, how to maintain the success of the initial intervention, or what intervention is best for individuals who did not initially respond. Critical decisions are made at different points throughout the intervention process. *Intervention components* are the set of intervention options at each critical decision point. A *tailoring variable* is a measurement that indicates which of the possible subsequent interventions is best for a specific individual. The tailoring variable is often a mechanism on the causal pathway of treatment (i.e., a mediator), or an early indicator of the overall outcome (Almirall et al., 2014). For example, tailoring variables could be scores on a test or questionnaire, performance from a session with an interventionist, level of adherence, or any other measure denoting early success or failure of the intervention. Finally, a *decision rule* occurs at each *critical decision point* and links the *tailoring variable* to the *intervention components* providing an intervention guideline based on the characteristics or response of the individual. For example, a decision rule can state that if an individual does not respond (as measured by the tailoring variable) to the first intervention, then that individual will switch to a different intervention.

Often, an AI might start with a relatively universal and low-cost intervention and then target a second tier of more intensive or costly intervention only for individuals who are still in need. For an illustration of a specific AI, consider the following in the area of youth conduct disorder. One AI may be: “First give the Everyday Parenting program for three 60 minute sessions. If youth exhibit a positive response (i.e., they improve to a normative level of behavior) then step down to only monitoring the youth's behavior. If youth do not exhibit response (i.e., they are still in the clinical range of behavior), then use a more intensive version (i.e. more sessions) of the Everyday Parenting program” (August et al., 2014). Thus, in this example, the *critical decisions points* included what intervention to begin with and what to do for those who respond to the initial intervention and what to do for those who do not respond. The *intervention components* include different numbers of sessions of the Everyday Parenting program. The *decision rules* guided by the *tailoring variable* (response) include beginning with three sessions, to add more sessions of the Everyday Parenting program for those who do not respond, and to discontinue the program for those who do respond.

AIs address critical decisions such as “What is the best first line intervention?”, “What should the duration of the first-line intervention be before deciding that it is not beneficial?”, “Can the intervention be modified based on an intermediate outcome?”, “What is the next step for those who did not do well on the first?”, and “Is the second intervention more effective if it was preceded by the first intervention?”. Critical decisions are not always

simple or necessarily intuitive, highlighting the need for evidence-based methods to construct effective AIs. Importantly, AIs are applied in many prevention efforts and other programs in the community; yet we have very few trials that have examined the effectiveness of AIs. That is, AIs and sequenced intervention programs have been employed in the community by combining experts' opinions on how and when to sequence interventions from observational data or by piecing together results from distinct RCTs (for more on alternative approaches to building AIs, see Lei et al., 2012) without rigorous evidence of the effectiveness of the entire AI (i.e., how each stage of the treatment fits together as a whole).

Over the past few decades SMARTs have emerged as tools to unpack the black box of AIs used in the community. SMARTs use two or more stages of randomization along with rigorous statistical methods to understand the components of and build effective AIs, while following the same individuals throughout a single trial. Specifically, a SMART is a type of RCT that follows the same individuals over sequential randomizations (i.e., across ordered interventions). The randomizations occur at *critical decision points* and can depend on information about the individual up to the point of randomization. Most SMARTs have two stages such that individuals are randomized to one of two (or more) interventions at the first stage, followed over a period of time, and potentially re-randomized to an additional intervention in the second stage. Second-stage randomization is based on an intermediate outcome (i.e., a *tailoring variable*) measured between the two stages. SMARTs can have more than two stages or intervention options at each stage, but should be kept parsimonious (see Nahum-Shani et al., 2012a for a detailed description of other possible SMART designs).

The goal of a SMART is to operationalize the *decision rules* used by practitioners and construct effective AIs. Multiple AIs of interest are embedded within a SMART and can be compared systematically to identify the most effective sequential strategies and for whom these strategies work best. That is, we can see which combination of interventions (and in what order) works best for which individuals. To the extent that researchers in the field of evaluation want to test sequential and personalized, multistage programs, a SMART design is a type of RCT that can accommodate personalization and sequencing of a program. Ideally, a SMART produces a promising AI which is then tested in a confirmatory standard RCT against standard practice.

SMARTs have been implemented in a community or classroom setting (August et al., 2014; Kasari et al., 2014; Pelham & Fabiano, 2008), but continue to be lesser known and applied in education and community settings than in the health sciences. Thus, we use the backdrop of a Head Start program to illustrate how a hypothetical SMART could be implemented in a community or education-based structure that might be involved in evaluation efforts. The goal of the hypothetical SMART that we describe would be to prevent and treat early child behavior problems which can prevent more serious behaviors and outcomes (e.g., drug abuse, jail time) in later adolescence and adulthood (Odgers et al., 2007; Shaw & Gross, 2008). Key to the presented example is that SMARTs are helpful not only in “treatment” of disordered medical or psychological states, but also in health promotion efforts, public health prevention efforts, and economic, educational, or vocational programs that may be implemented to promote positive, or protect from negative, outcomes. Here we emphasize the ability to run a SMART within an established community-based center; the proposed

SMART, however, could also be plausible within classrooms or other programs (e.g., women, infant, children centers; community centers).

Motivation for Head Start SMART to prevent early child behavior problems

Antisocial behavior and violence have large costs to society (Foster & Jones, 2005) by disrupting occupational, educational, social, mental and physical health outcomes of perpetrators and victims of crime (Romeo, Knapp, & Scott, 2006; Scott, Knapp, Henderson, & Maughan, 2001). Thus, prevention of behavior problems and antisocial behavior is important for our society broadly. Importantly, a majority of crime and violence is committed by a relatively small number of individuals who have behavior problems starting very early in life (Robins, 1978). Consequently, intervening with emerging behavior problems may be most effective early in life when behavior is more malleable, less rigid/entrenched, and less severe (Baydar, Reid, & Webster-Stratton, 2003; Dishion & Patterson, 1992; Ruma, Burke, & Thompson, 1996).

One difficulty in preventing later problem behaviors is the high desistance of young children from these behaviors. That is, all children have an increase in aggression during the toddler years, but only a few children will have problematic behavior that persists across childhood and into adolescence when their behavior will escalate (Loeber, Tremblay, Gagnon, & Charlebois, 1989; Shaw & Gross, 2008). The key to efficacious and cost-effective prevention is to identify the children at most risk and to begin stepped preventative intervention. Stepped preventative intervention may begin as less intensive, but include continued assessment to identify which children a) desist on their own, b) desist with non-intensive broad interventions, or c) continue to persist and need more intensive approaches. Cutting-edge interventions for behavior problems have targeted children most at risk (e.g., individuals with early risk factors such as those living in poverty, those with parents with mental illness) and have used strategies tailored to each family. These interventions often focus on parent training, an intervention with much empirical support in which parents learn skills to help shape their child's behavior in positive ways through the use of positive play, incentives, and time-out (Dishion et al., 2008; Gill, Hyde, Shaw, Dishion, & Wilson, 2008). Recent tailored parenting training interventions, however, have been short-term and have not pursued if particular interventions that occur before parent training affect children's outcome to parent training. Therefore, as types of preventative intervention programs are being implemented at the state and community level (<http://mchb.hrsa.gov/programs/homevisiting/>), it is important to identify the most cost-effective *sequence* that may be implemented broadly, but which also saves the most intensive (and expensive) interventions for children and families who are most in need. Finally, it is important for these programs to be embedded within community structures that target children and families in need.

Head Start was originally designed to break the cycle of poverty by providing preschoolers of low-income families with comprehensive services to help meet their academic, emotional, social, health, nutritional and psychological needs. Because of its structure and the population it reaches, Head Start is an initial primary prevention strategy for behavior problems (Yoshikawa, 1994) and a health promotion intervention to increase social, educational, and emotional development. As evidenced by several trials that have

implemented behavior problem interventions into Head Start classes (Webster-Stratton & Reid, 2003; Reid, Webster-Stratton, & Baydar, 2004), Head Start is a promising community venue where AIs could be implemented to prevent the escalation of early behavior problems. Moreover, we use Head Start as the setting for the hypothetical SMART because it serves the needs of children during the toddler years when behavior problems may first begin and be successfully treated.

Given the potential that Head Start may help prevent the persistence of early behavior problems, we highlight two additional and sequentially more intensive interventions that have been used to prevent and treat child behavior problems: feedback to parents on their child's behavior and parent training. As a first step, this hypothetical trial would compare the behavior of children who are monitored within the Head Start program without additional intervention to that of children whose families receive feedback about child behavior from a paraprofessional (i.e., is the behavior in a clinical range?). Feedback as an early intervention may help parents identify that their child's behavior is a problem and motivate them to work towards changing their child's behavior. This relatively low-cost step may help improve behavior for many children. If behavior has not improved naturally via developmental processes of desistance, the effect of Head Start, and/or feedback to parents, a more intensive intervention (parent training) would be warranted clinically. Parent training is an effective treatment for child behavior problems (Webster-Stratton & Reid, 2003; Reid, Webster-Stratton, & Baydar, 2004). However, parents may be resistant to this intervention if they have not had time to attempt to improve their child's behavior on their own. Thus, we hypothesize that feedback and parent training may each be effective for treating early behavior problems, but their sequence (i.e., feedback and then parenting training) may be the most effective. Additionally, in moving these potential intervention guidelines to practice, it would be beneficial to identify demographic or risk characteristics that are associated with best response to different sequences of interventions (e.g., do different families respond better to either single intervention or their combination?).

Head Start SMART

The main questions of interest in our hypothetical SMART within Head Start centers would be (1) whether providing feedback to parents on their child's behavior early in the school year is effective in reducing the proportion of children with behavior problems at the end of the year, and (2) if parent training is more effective for children who continue to have behavior problems if parents have already received feedback on their child's behavior. Figure 1 presents the hypothetical SMART design which contains two stages and three intervention components (feedback, parent training, and monitoring child behavior/"control condition"). Within the hypothetical SMART, we have proposed time points that make the trial feasible within a single school year. The estimated time points also allow for teachers to get to know their students and have influence on their behavior, for the first intervention to have an effect on child behavior via parenting behavior, and for the second intervention to be implemented and evaluated.

First-stage Interventions

The first stage in the proposed SMART investigates the usefulness of feedback and sets the stage to examine the extent to which providing feedback will set up better success in a subsequent intervention for children who continue to have behavior problems. In the hypothetical trial, after six weeks of a typical Head Start experience in which teachers get to know children and the Head Start program itself may have a positive effect on child behavior, a first measurement and randomization to the preventative intervention would be made. To measure child behavior, the Head Start teacher would complete the Teacher's version of the Child Behavior Checklist (CBCL), called the Teacher Report Form (TRF). The CBCL/TRF is a widely used and simple questionnaire to assess child behavior problems that has been extensively normed, allowing for comparison within age-group (Achenbach & Ruffle, 2000; Achenbach, 1966). The first CBCL/TRF measurement of child behavior would identify children for entry into the trial so that only children who appear to need further assistance and whose behavior may be disruptive to the classroom would be eligible to enroll. Children who score in the "borderline" or "clinical" range on the externalizing scale of the CBCL/TRF (Achenbach, 1991) would be randomized at six weeks to either 1) have their parents receive feedback on their behavior by a paraprofessional trained in giving normed standardized and motivational feedback or 2) to continue to monitor the child (i.e., do nothing new).

The first tested intervention in this trial would be giving parents feedback on their child's behavior and how it contrasts to other children of the same age. Feedback, particularly when framed in a motivational context, has been shown to be an effective and brief stand-alone intervention (DiClemente, Marinilli, Singh, & Bellino, 2001), as well as a critical component to many empirically supported treatments for a wide array of mental and physical health outcomes (Dishion et al., 2008; Khaylis, Yiaslas, Bergstrom, & Gore-Felton, 2010; Lee, Neighbors, Kilmer, & Larimer, 2010). Children who are randomized to the feedback group enter the preventative intervention where a paraprofessional (e.g., an experienced teacher, a trainee in a mental health or child development profession) would meet briefly in-person with parents. The meeting would provide feedback on the child's behavior (e.g., your child is in the 99th percentile for their age in disruptive behaviors) gathered by teacher ratings on the standardized measure (i.e., the CBCL/TRF). This meeting would also include motivational strategies to encourage parents to engage with community supports or to consider approaches to change child behavior (for a case study example of feedback see Gill et al., 2008). Children whose parents that would not receive feedback would be monitored meaning that their teachers would continue to fill out questionnaires on the child's behavior at each assessment point. The control setting serves as empirical evaluation of study aims with acknowledgement that filling out questionnaires and monitoring child behavior may be an intervention in itself that affects behavior.

Tailoring Variable

Children participating in the SMART (i.e., those who made the cut-off of being in the borderline or clinical range, whether they received feedback or not) would be followed through the program as usual until halfway through the school year (e.g., February). At this intermediate time point, child behavior would be assessed again by teachers using the

CBCL/TRF (the tailoring variable). The tailoring variable in any SMART must be an agreed upon measure from all investigators that is a proximal measure of the outcome (but does not have to be the exact same as the primary outcome), denotes early success or failure, and can be implemented in practice. It does not necessarily need to be dichotomous as we have indicated in this SMART, but it should have few categories for feasibility and practicality. If investigators cannot agree upon such a measure, pilot studies prior to implementing a SMART may inform the choice of a suitable tailoring variable (Almirall, Compton, Gunlicks-Stoessel, Duan, & Murphy, 2012).

Second-stage Interventions

Second-stage interventions depend on the tailoring variable. In this trial, children who are no longer in the clinical or borderline range at the intermediate time point would be monitored throughout the second-stage of the trial and not re-randomized to further intervention (i.e., they are now in the normative range for behavior and do not need further intervention). Children who would continue to be in the clinical or borderline range would be randomized to continue in the program as usual (i.e., monitoring) or for their parents to receive a parent training intervention. Enrollment in the Head Start program alone may reduce behavior problems for some children and feedback may increasingly reduce behavior problems, but we expect at least 35% of children would continue to have behavior problems in February. In general, the category of children that undergoes second-stage randomization should contain at least 20% of the sample. We will refer to the group that undergoes second-stage randomization as non-responders, but this label could differ depending on the design. If the number of expected non-responders is low, the SMART may instead randomize responders to investigate the best strategy to continue to engage participants or maintain success.

Parent training programs in which clinicians engage in individual or group sessions with parents aimed at enhancing parenting strategies for difficult children, have been shown to be robustly effective within Head Start partnerships (Baydar et al., 2003; Reid et al., 2004). In community group settings, parent training has been shown to be effective in as little as 8 weeks (Cunningham, Bremner, & Boyle, 1995). However, parent training is less optimal as a first stage intervention because it requires more intensive resources (a clinician or highly skilled paraprofessional running sessions) that may not be necessary for many children whose behavior problems dissipate on their own or in response to parental feedback. Thus, we propose this intervention as a second and more intensive intervention. After children would be randomized to this second stage and their families engage in the intervention, relative success rates need to be measured for each treatment path. Hence, at the end of the school year in May, the CBCL/TRF would again be completed by the teacher to see how the first- and second-stage interventions affected child behavior.

Embedded Adaptive Interventions

Testing the sequence of feedback followed by parent training is important as this two stage intervention series may be particularly powerful for children who are not otherwise improving. In other clinical trials, we have observed that parents often value autonomy first in their parenting after receiving feedback (i.e., they want to “try it their way” first). Therefore, parents may be more open to the guidance provided in parent training if they

have already received feedback and their child's behavior has not improved (Gill et al., 2008). Investigating AIs embedded within the SMART assesses which tailored sequences of interventions are optimal for particular children and families. In this case, embedded AIs can test whether parent training is more effective if feedback precedes it.

Two AIs are embedded in the hypothetical SMART. The AIs in the Head Start SMART are “First provide feedback to the parent and if the child is in the normative range of the externalizing scale on the CBCL/TRF in February, monitor the child. If the child is in the borderline or clinical range, proceed to an 8 week parent training intervention” and “Monitor the child from the start of school and if the child is in the normative range of the externalizing scale on the CBCL/TRF in February, continue to monitor. If the child is in the borderline or clinical range, proceed to the 8 week parent training intervention.” Both of these AIs are tailored because subsequent treatment depends on behavior as measured by the intermediate CBCL/TRF (i.e. the *tailoring variable*).

Additionally two non-adaptive interventions (NAIs) are embedded in the trial due to one of the interventions at each stage being “to monitor”. The NAIs provide “control” conditions to assess the effects of feedback and parent training. NAIs still include a first-stage intervention followed by a second-stage intervention for responders and a second-stage intervention for non-responders, but the second-stage intervention does not differ by response status. Specifically, the two NAIs in the Head Start SMART are “Provide feedback to the parent and continue to monitor,” and “Monitor the child throughout the program.”

Objectives of a SMART Design

SMARTs can investigate a number of different research questions. The advantage of a SMART design compared to a standard RCT is that it can address similar questions of an RCT and additionally assess tailored intervention sequences. Furthermore, as opposed to a discontinuation or non-responder trial (i.e. a trial where only individuals who have not responded to an initial treatment are followed and others initially enrolled on the trial and responded are excluded from further analysis), a SMART follows the same participants throughout the entire intervention process. By following all participants, non-responders are more likely to be representative of non-responders in the broader population, not just of those who are highly motivated to enroll in a non-responder trial. A SMART can be more efficient than a standard RCT by following all participants through multiple stages and can address a broader range of questions assessing both first and second stage treatments and how they together affect the outcome (Almirall, Compton, Rynn, Walkup, & Murphy, 2012). Furthermore, a SMART provides the ability to test tailoring variables by the collection of baseline and time-varying variables during the trial.

The objectives of a SMART are summarized in Table 1 and may include the comparison or estimation of: 1) first-stage treatment effects; 2) second-stage treatment effects for those who are re-randomized; or 3) AI effects. Each objective requires the analysis of a combination of multiple subgroups within the SMART (adding to the efficiency of the design). Individual subgroups within a SMART are generally not compared (see Caveats to Aims below). The comparison of first or second stage interventions is denoted as a stage-specific objective and

is similar to the objective of standard RCTs. However, in a SMART the interventions are studied within the context of developing AIs (i.e. the first- or second-stage interventions are followed or preceded by interventions which may influence the overall outcome). Objectives involving AIs can be categorized by estimation (ranking the effects to find the best), comparison (hypothesis testing), or optimization of the embedded AIs (Almirall et al., 2014). AI objectives are unique to a SMART and target how tailored sequences of interventions affect the overall outcome. The first and second stage interventions may interact to produce an overall outcome differently than each piece individually; without a SMART design, intervention interactions cannot be rigorously assessed. The optimization aim goes beyond the embedded tailoring variable to further specify for whom each sequence works best. In this aim, baseline and time-varying variables (e.g., child, parent, and teacher characteristics measured at the start of the study and throughout the study) are evaluated to find additional characteristics or behaviors associated with best overall outcomes. We demonstrate each of the aims in the context of the Head Start SMART.

Stage-Specific Objective

In a SMART, a stage-specific objective is similar to the objective of a standard RCT and is relevant if the direct comparison of the first or second stage interventions has not been made previously in the population of interest. Trial size is similar to that of a standard RCT, but a SMART has the benefit of exploring the effects of the AIs. For the Head Start SMART one option for a stage-specific objective is the first-stage comparison. This comparison is to understand the difference in child behavior due to initially providing a feedback session to parents versus only monitoring behavior and compares children who followed each first-stage intervention (i.e., subgroups A, B and C versus D, E, and F in Figure 1). Another option is the second-stage comparison or to understand whether the parent training component was effective for the reduction of problem behavior for children who continued to have behavior issues after time in Head Start (i.e., subgroups B and E versus C and F in Figure 1).

Power and Statistical Analysis Plan for a Stage-Specific Objective

To power a trial comparing first-stage interventions, the sample size can be calculated as usual for a standard two-arm RCT because the second-stage treatments are equated across the initial intervention arms. For example, consider the effect of feedback on the proportion of children with borderline or clinical CBCL/TRF scores at the end of the trial. We aim to detect a difference in the proportion of children who have behavior problems at the end of the trial from the expected 55% of children who were initially monitored versus 40% of children who were initially given feedback. In this case, the SMART requires at least 173 children per initial intervention arm (a total of 346) for 80% power using a two-sided test for proportions at the 5% significance level. The statistical analysis plan could include a standard logistic model (or a longitudinal model) with the binary outcome of CBCL/TRF status at the end of the trial and predictor of interest being the first-stage treatment indicator (feedback versus monitoring). Baseline variables (child, parent, and teacher characteristics) may be controlled for in the model to gain efficiency (i.e., decrease standard errors) in the estimation of the intervention effect, but variables collected after baseline cannot be included since they may be influenced by the initial treatment.

To power a trial comparing second-stage interventions, the sample size of a standard two arm RCT must be modified by the estimated percentage of children who would be re-randomized at the second stage. For instance, if we consider the estimates and difference above as also clinically meaningful between the second-stage interventions, 173 children who have borderline or clinical CBCL/TRF scores in February would be required per group. Since we expect 35% of children to be non-responders in February, the full SMART requires 989 ($346/0.35$) total children at the start of the trial so that 346 children end up in the subgroups B, E, C and F. The statistical analysis plan may implement standard logistic regression (or a longitudinal model) where the variable of interest is an indicator of second-stage treatment and include baseline variables. Covariates measured up to the second-stage randomization may be included in the model to reduce error variance in the intervention effect estimate. Note that we have conditioned on the intermediate outcome of the CBCL/TRF score here (i.e. assumed that children continue to have behavior problems in February), so the intermediate outcome status (CBCL/TRF score in February) does not enter the model.

Embedded Adaptive Intervention Objective

SMARTs are specifically designed to investigate the novel and relevant comparisons of tailored intervention sequences. The key is to determine if the first intervention increases the effectiveness of the second. Embedded AI objectives are commonly secondary (i.e., not powered for) since the goal of a SMART is to develop AIs where many intervention components have not been previously proven effective and interest is often in exploring, but not rigorously proving the effectiveness of the embedded AIs. Embedded AI objectives, however, can be the primary objective of a SMART and adequately powered. The embedded AI objective may be a comparison of any two specific AIs or the estimation of the effects of all AIs with interest in finding the best performing AI(s). In the Head Start SMART, comparison of the embedded AIs attempts to understand if parent training for those with borderline or clinical CBCL/TRF externalizing scores is more effective after feedback. The effects of the two embedded AIs in the trial (one AI includes subgroups A and B versus AI including subgroups D and E; Figure 1) can be estimated and/or formally compared in analysis.

Power and Statistical Analysis Plan for Embedded Interventions Objective

Standard regression analyses cannot be applied to estimate or compare AIs because of the restricted randomization imposed in the design of a SMART. In the Head Start SMART randomization is restricted at the second stage since only non-responding children are re-randomized. A simple average of children who followed each AI is biased by over-representing responding children and under-representing non-responding children. A weighted and replicated approach accounts for the restricted randomization and allows for simultaneous estimation of effects (Nahum-Shani et al., 2012a). The weighted and replicated approach weights individuals according to their probability of receiving their particular intervention sequence (e.g. in the Head Start SMART, those with CBCL/TRF ≥ 60 are randomized twice with $\frac{1}{4}$ chance of following their intervention sequence and those with CBCL/TRF < 60 are randomized once with $\frac{1}{2}$ chance of following their interventions). Once data are weighted, groups that are consistent with more than one AI (or NAI) are replicated

to allow for simultaneous estimation of effects. The responders (CBCL/TRF<60 after the first-stage) are consistent with one AI of interest: initial feedback and monitor if CBCL/TRF<60 and parent training intervention if CBCL/TRF ≥ 60, and one NAI of interest: initial feedback followed by monitoring. Baseline covariates may be used in regression to control for possible differences in initial intervention groups and to reduce error variance and robust standard errors account for sample-to-sample variance in the weights (Nahum-Shani et al., 2012a).

If the estimation or comparison of embedded AIs is of primary interest then special sample size considerations are necessary for adequate power of the study. For continuous outcomes where estimating and ordering or testing the outcomes is the primary aim, sample size may be obtained by using methods from Oetting, Levy, Weiss, & Murphy, 2011. Similar methods are available for the comparison of any two AIs with binary outcomes (manuscript forthcoming by Kidwell) and for survival outcomes (time-to-event analysis, Feng & Wahed, 2009; Li & Murphy, 2011). Sample size calculators are available online (<http://methodologymedia.psu.edu/smart/samplesize>).

To power a comparison of the two AIs in the Head Start SMART, we can apply SMART sample size methods for binary outcomes. For this calculation, assume we want to detect a difference of 15% of children with behavior problems at the end of the trial between the two embedded AIs. Then assume that we expect 50% of the children who follow one AI (children who received feedback at the beginning of the trial and were monitored if they had a CBCL/TRF<60 and received the parent training intervention if they had CBCL/TRF ≥ 60) and 35% of children who follow the other AI (children who were monitored at the beginning of the trial and continued monitoring if they had a CBCL/TRF<60 but received the parent training intervention if they had CBCL/TRF ≥ 60) to have behavior problems at the end of the trial. We also expect that the rate of non-response to the initial treatment to be 35% (for both feedback and monitoring). Then, at least 465 children are required for 80% power. It is important to note that the sample size comparing two AIs of a SMART cannot be directly compared to the sample size comparing two intervention arms from a standard RCT because the questions differ substantially.

Optimization Objective

The optimization objective takes advantage of heterogeneity in the sample to find subsets of patients for which particular AIs are best. During the conceptualization of the study, investigators may pose several variables which are likely to moderate treatment effects. Optimization analyses may also be performed post-hoc as hypothesis generating to inform future trials. For example, one hypothesis to investigate may be that children with parents who are more involved in Head Start or the intervention may benefit most from feedback initially. Another hypothesis is that children who have a younger mother of lower SES, with more stressors and less skills, may benefit less from feedback because the mother has fewer resources to enact to change behavior without additional support. Alternatively, children with younger mothers or those of lower SES may benefit more from the parenting intervention since it provides additional support and education for the mother (i.e., feedback will be less effective and parent training more effective for younger and lower SES parents).

Power and Statistical Analysis Plan for Optimization Objective

Like the embedded AI objective, the optimization objective also requires methods beyond standard (moderator) analyses (Almirall et al., 2014). One effective approach for moderator analysis in the presence of multiple stages of intervention is Q-learning (Watkins, 1989). Q-learning is used to explore baseline and time-varying variables as potential moderators and contrast intervention components at each stage controlling for effects of previous and subsequent components (Nahum-Shani et al., 2012b). This method constructs decision rules which form AIs by employing backward induction through a series of regressions (Almirall et al., 2014; Nahum-Shani et al., 2012b). For the Head Start SMART, we may use Q-learning to see if parent engagement at Head Start moderates the effects of feedback and parent training to provide a more detailed AI at the end of the trial. The first step of Q-learning examines tailoring variables of the second-stage treatment (parent training or monitoring) to find baseline and time-varying moderators that are significantly associated with positive outcomes (i.e., normative behavior). The next step incorporates the effect of making the optimal future treatment decision. The final step assesses tailoring variables of the first-stage treatment (feedback or monitoring) assuming that an optimal treatment decision has been made at the second-stage. The resulting optimal AI for the Head Start SMART may be “For children whose parent(s) are highly engaged, first provide feedback to the parents and continue to monitor children. For children whose parents are not highly engaged, first provide feedback to the parents and if children continue to have behavior issues, engage parents in training, but monitor those who have mitigated problems on own.”

The overall messages in regards to the design and power of a SMART are: a) investigating AIs within a SMART requires specific and different statistical approaches than a standard trial, but also tests more novel and interesting questions, and b) clear and accessible resources exist for those interested in designing and conducting a SMART in terms of powering the study and performing the analysis. Though the approach may be more complex, it is still accessible to researchers implementing tailored sequential interventions in community settings.

Caveats to Objectives

Note, that we have provided power calculations for each type of objective within a SMART. However, a SMART must not be powered on *all* objectives. Similar to a standard RCT, generally, a SMART is powered on a main objective which may or may not provide enough power for other objectives of the trial. For example, if the Head Start trial was powered on the comparison of first-stage interventions (346 children), the second-stage comparison and embedded AI comparison would not be powered at the 80% level. If this trial was powered on the second-stage treatment comparison (989 children), the other aims would be highly powered. If the trial was powered on the embedded AI aim (465 children), the initial intervention comparison would be highly powered, but the second-stage treatment comparison would not. Thus, the main research question of interest and budget may define the primary objective of the trial and therefore the power calculation.

Generally, the objectives of a SMART address research questions by combining subgroups within the SMART, not by comparing individual subgroups. In the Head Start SMART, a

researcher may want to compare subgroups B versus E to investigate the effectiveness of feedback prior to parent training since both are non-responders. Comparing subgroups B and E is not a comparison of AIs because it does not include those who had a CBCL/TRF<60 in February (subgroups A and D in Figure 1). Furthermore, the comparison of subgroups B vs. E is not a causal comparison (i.e. cannot define a cause-effect relationship between intervention and outcome) because the non-responders to feedback may be substantively different from the non-responders to monitoring leading to biased results. The first-stage treatment effect cannot simply be “averaged out” because the effect is unlikely to be equal between both sets of non-responders. Similarly, comparisons between individuals who do and do not respond (e.g., comparing responders to feedback versus non-responders to feedback who are monitored) or between subgroups which begin with different first-stage treatments (e.g., comparing non-responders to feedback to non-responders to monitoring) are not causal comparisons. Thus, individual subgroups are not compared in a SMART unless the groups start with the same first-stage intervention and have the same response status.

Challenges of a SMART design

A SMART is a valuable tool to answer treatment or prevention questions which are not possible to address through standard designs and could lead to new advances in the field of evaluation. However, SMARTs are not always appropriate or feasible. The timing of a SMART is particularly important such that the intermediate outcome must be able to be measured within a relatively short period of time. Any trial should be implemented within a few years to ensure relevance of its findings. Since two or more stages and follow up must occur within an appropriate trial time limit, it may be more difficult to implement a SMART than a standard one-stage trial in certain settings.

Beyond timing which affects all trials, those involved in the design and implementation of a SMART must pay particular attention to obtaining buy-in and fidelity of those implementing the intervention. Fidelity to the interventions is particularly important, as the specific interventions involved in our hypothetical SMART have been shown to be more effective when they are implemented with high fidelity (Chiapa et al., 2015). Thus, an additional parameter to consider for investigators is how to monitor implementation fidelity. To ensure implementation fidelity within our hypothetical trial, we would audio or video-tape a selection of feedback and parent intervention sessions to provide feedback to clinicians (Chiapa et al., 2015). As we expect that changes in parenting are the mediator of effects of the interventions on child behavior (Dishion et al., 2008), we could assess parenting behavior via questionnaire, or ideally via a structured observation (see below for caveats about funding). This assessment would help to assess this potential mediator of treatment effects and a sign of treatment non-compliance. August et al., 2014 discuss 4 options to measure intervention fidelity and minimize the threat to internal validity including up-front training of the research staff from those certified in the interventions, supervisor oversight of delivery using a preliminary client sample, completion of session logs by research staff, and attendance by research staff at regular supervisory meetings.

To increase the success and implementation of any SMART, engagement of investigators, statisticians, and stakeholders (including agency staff if involved) is imperative. Before and during the trial it would be important to meet with stakeholders (e.g., focus groups with parents, teachers, administrators) to tailor the interventions to what would be acceptable for all individuals involved. August et al., 2014 discuss establishing buy-in in the community through the creation of a steering committee with monthly meetings, education and training sessions for agency and research staff, and a pre-pilot study with a small sample of individuals to refine the protocol and gain experience. If intervention implementation is of great concern, a pilot SMART is an excellent approach to gain buy-in, develop fidelity, address feasibility concerns and refine components of the trial (Almirall, Compton, Gunlicks-Stoessel, Duan, & Murphy, 2012). A pilot study does not require many participants and can be published, establishing the investigator's expertise in the literature.

The need for sequential randomizations may also seem complex and a challenge to implementing a SMART. A SMART is defined by sequential randomizations, but the randomizations do not have to occur sequentially. Participants in the trial may be randomized upfront to a specific embedded AI (note that upfront randomization still includes adaptation since AIs guide treatment based on the tailoring variable, but it more closely resembles a multi-arm RCT). Upfront randomization reduces infrastructure, but may induce bias. The advantage to randomizing just after the intermediate outcome is that randomization can stratify on participant characteristics and behaviors for balance in the subsequent stage, so that attrition between the first and second stages will not affect the number randomized into each second-stage intervention (Almirall, Compton, Gunlicks-Stoessel, Duan, & Murphy, 2012; Ko & Wahed, 2012). With a good design at the onset, strict measurement procedures, and a trained staff, re-randomization can become automated and no more difficult mid-trial than initially. If sequential randomization at the individual level still seems daunting and the main research question involves implementation of practices at the community, agency, or center level, randomization may occur at the center level, not at the individual level. For more details and a concrete example of a cluster-randomized trial in the field see Kilbourne et al., 2014.

Attrition and missing response status at the intermediate time point may cause additional challenges to implementing a SMART. Subject attrition is especially important because the resulting AIs should be viable and realistic (Wang, Rotnitzky, Lin, Millikan, & Thall, 2012). Therefore, the definition of the embedded AI should include common contingencies like dropout (Almirall, Lizotte, & Murphy, 2012). In the Head Start SMART attrition is likely with an expected dropout rate of 22% due to families leaving Head Start or families being uninterested in participating in the year-long intervention as shown in a previous trial (Webster-Stratton, 1998). Attrition may best be addressed by including a protocol for the reengagement of participants since the initial treatments in the Head Start SMART are passive (monitor) or a one-time occurrence (feedback). The definition of the AI would then include the first-stage treatment and reengagement strategy in the future rollout. For children who drop out prior to randomization, it is critical to set a clear plan prior to the initiation of the SMART to characterize the children who have unknown intermediate outcome (response) status. The plan may be to classify such participants as non-responders, randomize to a second-stage treatment and treat the outcome as missing (e.g., use multiple

imputation in analysis), classify response by other measures made proximal to dropout, classify as attriters and consider an alternative standard plan for attriters in the definition of an AI, or to continue to attempt to reengage the participant and allow for later entry into the second-stage (Almirall, Compton, et al., 2012). Again, a pilot study may aid in estimating the proportion of participants who dropout, the reasons for noncompliance, and best options for handling missing data.

Some of the possible barriers mentioned above, as well as the need for larger-scale funding may limit the number of researchers interested in implementing a SMART. SMART research may be funded at the federal level through appropriate agencies (e.g., the National Institute on Child Development, the National Science Foundation, the Centers for Disease Control) that are able to fund larger trials. A SMART does not need to be more expensive than a standard trial if embedded within existing structures like Head Start where adaptations occur within a tight time frame and infrastructure exists for the trial. If implemented within existing structures, state, local, and private funders may be interested in funding these cutting-edge approaches to intervention (e.g., see recent funding on home visitation programs <http://mchb.hrsa.gov/programs/homevisiting/>).

In sum, SMARTs contain many of the same challenges as RCTs including challenges to external validity in that they are tested in a controlled manner which may not generalize or be practical or acceptable in all settings. SMARTs, like RCTs, also require more infrastructure and resources to implement than other methods of evaluation. Additionally, SMARTs typically focus on two decision points, but effective interventions may contain many decision points in reality. Though SMARTs are closer than a typical RCT in testing more realistic interventions, it is still difficult to practically test many decision points with many potential intervention options that may reflect the complexity of some community based programs.

Similarities and Differences of an AI, SMART, and Response to Intervention

Though we have described SMARTs as a way to test a stepped approach to intervention, readers of this journal may recognize this approach as being similar to Response to Intervention. A revised version of the Individuals with Disabilities Education Improvement Act of 2004 (IDEIA) included an alternative method to identify children with learning disabilities and provide early intervention to all children at risk for school failure (Fuchs & Fuchs, 2006). The alternative method in IDEIA is known as Response to Intervention (RTI) and has been primarily implemented to enhance literacy in K-12 students, but can be implemented in early childhood targeting language, literacy, mathematics, and social-emotional competence (Division for Early Childhood of the Council for Exceptional Children, National Association for the Education of Young Children, & National Head Start Association, 2014). RTI is similar to an AI such that it is a tiered or stepped approach to treating or preventing problems, focusing on how to obtain best outcomes for children at highest risk, while also saving resource-intensive interventions for children who need them the most. RTI includes a multi-tiered system of teaching and caregiving practices such that all children may first receive the same curriculum and services. After ongoing assessment and progress monitoring, children who are non-responsive, receive increased and potentially

individualized teaching and care. Teachers and family members use evidence-based instruction or caregiving at each tier. Thus, AIs, particularly those similar to the ones embedded within our hypothetical SMART, and RTI approaches provide guidelines for effective sequential interventions that are adapted to individual needs of children.

A SMART differs from RTI because a SMART is a trial which tests the components and ordering of components that make-up an AI (or the sequence that makes up RTI). Therefore, RTI may be effective and is data-driven, but guidance of the sequence is observational without empirical tests of the entire sequenced process. Thus, a SMART could be applied to RTI to provide rigorous evidence of the tiered approach. A SMART may justify the need for extra teachers, caregivers, or training used within RTI.

Conclusion

AIs or the tailored sequencing of (at least) two stages of interventions are commonly used in practice to personalize interventions. However, these intervention sequences are often based on expert experience and piecing together components and results from previous standard trials. Piecing together components may be problematic in many settings because the sequencing itself and combinations of treatments may affect the overall outcome. A SMART unpacks the black box of sequenced intervention components to provide effective evidence-based adaptive, personalized, and sequenced interventions. SMARTs may investigate intervention comparisons at a single stage similar to a standard trial, while also assessing tailored interventions that are interactions between first and second stage interventions based on intermediate response. Furthermore, SMARTs follow the same individuals throughout the intervention process and are thus uniquely able to examine the timing and sequencing of interventions, as well as intervention synergies. SMARTs may require additional infrastructure than a standard RCT and are not appropriate for all types of research questions due to feasibility and requiring reliable tailoring variables and sequential randomizations. However, when appropriate, obtaining buy-in and fidelity through early engagement of investigators, statisticians and stakeholders and the execution of a pilot study can be crucial steps to lead to the successful implementation of a SMART.

Although SMARTs require a strong rationale and rigorous planning (like most trials), the applications of AIs and SMARTs is relevant to community research, prevention, and program evaluation. SMARTs can investigate when more costly or high-resourced interventions are (or are not) necessary, for whom interventions are most effective, and the best follow-up intervention if the initial intervention is not successful. Although our example focused on behavior “problems”, it could easily be mapped on to the development of normative, positive behaviors such as language developmental/literacy, prosocial peer relationships, or civic engagement. Community-based research may provide additional hurdles in the implementation of a SMART, but quality training, incorporating contingency plans into embedded AI definitions, or using cluster-randomization at the center level may reduce obstacles. A SMART is a valuable tool to provide effective evidence-based AIs, which in turn can lead to enhanced interventions, care, prevention, or successful programs. At the most basic level, a SMART provides a design to test intervention questions and

develop adaptive guidelines which are much closer to how interventions are actually implemented in the “real world”.

References

- Achenbach TM. The classification of children's psychiatric symptoms: A factor-analytic study. *Psychological Monographs: General and Applied*. 1966; 80(7):1–37. <http://doi.org/10.1037/h0093906>.
- Achenbach, TM. *Manual for the Child Behavior Checklist/4–18 and 1991 Profile*. Burlington, VT: University of Vermont, Department of Psychiatry; 1991.
- Achenbach TM, Ruffle TM. The Child Behavior Checklist and related forms for assessing behavioral/emotional problems and competencies. *Pediatrics in Review / American Academy of Pediatrics*. 2000; 21(8):265–271.
- Almirall D, Compton SN, Gunlicks-Stoessel M, Duan N, Murphy SA. Designing a pilot sequential multiple assignment randomized trial for developing an adaptive treatment strategy. *Statistics in Medicine*. 2012; 31(17):1887–1902. <http://doi.org/10.1002/sim.4512>. [PubMed: 22438190]
- Almirall D, Compton SN, Rynn MA, Walkup JT, Murphy SA. SMARTer discontinuation trial designs for developing an adaptive treatment strategy. *Journal of Child and Adolescent Psychopharmacology*. 2012; 22(5):364–374. <http://doi.org/10.1089/cap.2011.0073>. [PubMed: 23083023]
- Almirall D, Lizotte DJ, Murphy SA. Comment. *Journal of the American Statistical Association*. 2012; 107(498):509–512. <http://doi.org/10.1080/01621459.2012.665615>. [PubMed: 23543940]
- Almirall D, Nahum-Shani I, Sherwood NE, Murphy SA. Introduction to SMART designs for the development of adaptive interventions: with application to weight loss research. *Translational Behavioral Medicine*. 2014; 4(3):260–274. <http://doi.org/10.1007/s13142-014-0265-0>. [PubMed: 25264466]
- August GJ, Piehler TF, Bloomquist ML. Being “SMART” About Adolescent Conduct Problems Prevention: Executing a SMART Pilot Study in a Juvenile Diversion Agency. *Journal of Clinical Child and Adolescent Psychology: The Official Journal for the Society of Clinical Child and Adolescent Psychology, American Psychological Association, Division*. 2014; 53:1–15. <http://doi.org/10.1080/15374416.2014.945212>.
- Auyeung SF, Long Q, Royster EB, Murthy S, McNutt MD, Lawson D, et al. Musselman DL. Sequential multiple-assignment randomized trial design of neurobehavioral treatment for patients with metastatic malignant melanoma undergoing high-dose interferon-alpha therapy. *Clinical Trials*. 2009; 6:480–490. [PubMed: 19786415]
- Baydar N, Reid MJ, Webster-Stratton C. The role of mental health factors and program engagement in the effectiveness of a preventive parenting program for Head Start mothers. *Child Development*. 2003; 74(5):1433–1453. [PubMed: 14552407]
- Bierman KL, Nix RL, Maples JJ, Murphy SA. Conduct Problems Prevention Research Group. Examining clinical judgment in an adaptive intervention design: The fast track program. *Journal of Consulting and Clinical Psychology*. 2006; 74(3):468–481. <http://doi.org/10.1037/0022-006X.74.3.468>. [PubMed: 16822104]
- Chalmers I. Trying to do more Good than Harm in Policy and Practice: The Role of Rigorous, Transparent, Up-to-Date Evaluations. *The Annals of the American Academy of Political and Social Science*. 2003; 589(1):22–40. <http://doi.org/10.1177/0002716203254762>.
- Chiapa, A., Smith, JD., Kim, H., Dishion, TJ., Shaw, DS., Wilson, MN. The Trajectory of Fidelity in a Multiyear Trial of the Family Check-Up Predicts Change in Child Problem Behavior. *Journal of Consulting and Clinical Psychology*. 2015. <http://doi.org/10.1037/ccp0000034>
- Cunningham CE, Bremner R, Boyle M. Large group community-based parenting programs for families of preschoolers at risk for disruptive behaviour disorders: utilization, cost effectiveness, and outcome. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*. 1995; 36(7): 1141–1159.
- Dawson R, Lavori PW. Placebo-free designs for evaluating new mental health treatments: The use of adaptive treatment strategies. *Statistics in Medicine*. 2004; 23:3249–3262. [PubMed: 15490427]

- DiClemente CC, Marinilli AS, Singh M, Bellino LE. The role of feedback in the process of health behavior change. *American Journal of Health Behavior*. 2001; 25(3):217–227. [PubMed: 11322620]
- Dishion TJ, Nelson SE, Kavanagh K. The family check-up with high-risk young adolescents: Preventing early-onset substance use by parent monitoring. *Behavior Therapy*. 2003; 34(4):553–571. [http://doi.org/10.1016/S0005-7894\(03\)80035-7](http://doi.org/10.1016/S0005-7894(03)80035-7).
- Dishion TJ, Patterson GR. Age effects in parent training outcome. *Behavior Therapy*. 1992; 23(4):719–729. [http://doi.org/10.1016/S0005-7894\(05\)80231-X](http://doi.org/10.1016/S0005-7894(05)80231-X).
- Dishion TJ, Shaw D, Connell A, Gardner F, Weaver C, Wilson M. The Family Check-Up With High-Risk Indigent Families: Preventing Problem Behavior by Increasing Parents' Positive Behavior Support in Early Childhood. *Child Development*. 2008; 79(5):1395–1414. <http://doi.org/10.1111/j.1467-8624.2008.01195.x>. [PubMed: 18826532]
- Division for Early Childhood of the Council for Exceptional Children, National Association for the Education of Young Children, & National Head Start Association. Frameworks for Response to Intervention in Early Childhood: Description and Implications. *Communication Disorders Quarterly*. 2014; 35(2):108–119. <http://doi.org/10.1177/1525740113514111>.
- Donaldson, SI., Christie, CA., Mark, MM. What counts as credible evidence in applied research and evaluation?. Los Angeles, [Calif.]; London: SAGE; 2009. Retrieved from <http://SRMO.sagepub.com/view/what-counts-as-credible-evidence-in-applied-research-and-evaluation-practice/SAGE.xml>
- Feng W, Wahed AS. Sample size for two-stage studies with maintenance therapy. *Statistics in Medicine*. 2009; 28:2028–2041. [PubMed: 19382105]
- Foster EM, Jones DE. The High Costs of Aggression: Public Expenditures Resulting From Conduct Disorder. *American Journal of Public Health*. 2005; 95(10):1767–1772. <http://doi.org/10.2105/AJPH.2004.061424>. [PubMed: 16131639]
- Fuchs D, Fuchs LS. Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly*. 2006; 41(1):93–99. <http://doi.org/10.1598/RRQ.41.1.4>.
- Gambrill E. Evidence-Based Practice and Policy: Choices Ahead. *Research on Social Work Practice*. 2006; 16(3):338–357. <http://doi.org/10.1177/1049731505284205>.
- Gill AM, Hyde LW, Shaw DS, Dishion TJ, Wilson MN. The Family Check-Up in Early Childhood: A Case Study of Intervention Process and Change. *Journal of Clinical Child & Adolescent Psychology*. 2008; 37(4):893–904. <http://doi.org/10.1080/15374410802359858>. [PubMed: 18991138]
- Gone JP. Reconciling evidence-based practice and cultural competence in mental health services: Introduction to a special issue. *Transcultural Psychiatry*. 2015; 52(2):139–149. <http://doi.org/10.1177/1363461514568239>. [PubMed: 25808532]
- Kasari C, Kaiser A, Goods K, Nietfeld J, Mathy P, Landa R, Almirall D. Communication Interventions for Minimally Verbal Children With Autism: A Sequential Multiple Assignment Randomized Trial. *Journal of the American Academy of Child & Adolescent Psychiatry*. 2014; 53(6):635–646. <http://doi.org/10.1016/j.jaac.2014.01.019>. [PubMed: 24839882]
- Kazdin AE. Parent management training: evidence, outcomes, and issues. *Journal of the American Academy of Child and Adolescent Psychiatry*. 1997; 36(10):1349–1356. <http://doi.org/10.1097/00004583-199710000-00016>. [PubMed: 9334547]
- Khaylis A, Yiaslas T, Bergstrom J, Gore-Felton C. A Review of Efficacious Technology-Based Weight-Loss Interventions: Five Key Components. *Telemedicine and E-Health*. 2010; 16(9):931–938. <http://doi.org/10.1089/tmj.2010.0065>. [PubMed: 21091286]
- Kilbourne AM, Almirall D, Eisenberg D, Waxmonsky J, Goodrich DE, Fortney JC, et al. Thomas MR. Protocol: Adaptive Implementation of Effective Programs Trial (ADEPT): cluster randomized SMART trial comparing a standard versus enhanced implementation strategy to improve outcomes of a mood disorders program. *Implementation Science*. 2014; 9(1):132. <http://doi.org/10.1186/s13012-014-0132-x>. [PubMed: 25267385]
- Kleijnen, J., Gotzsche, P., Kunz, RA., Oxman, AD., Chalmers, I. So what's so special about randomisation?. In: Maynard, A., Cochrane, AL., editors. *Non-random reflections on health*

services research: on the 25th anniversary of Archie Cochrane's Effectiveness and efficiency. London: BMJ Publ. Group; 1997.

- Ko JH, Wahed AS. Up-front versus sequential randomizations for inference on adaptive treatment strategies. *Statistics in Medicine*. 2012; 31:812–830. [PubMed: 22362642]
- Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. *BMJ*. 1998; 317(7167):1185–1190. <http://doi.org/10.1136/bmj.317.7167.1185>. [PubMed: 9794851]
- Lavori PW, Dawson R. A design for testing clinical strategies: Biased individually tailored within-subject randomization. *Journal of the Royal Statistical Society, A*. 2000; 163:29–38.
- Lavori PW, Dawson R, Rush AJ. Dynamic treatment regimes: Practical design considerations. *Clinical Trials*. 2004; 1:9–20. [PubMed: 16281458]
- Lee CM, Neighbors C, Kilmer JR, Larimer ME. A brief, web-based personalized feedback selective intervention for college student marijuana use: A randomized clinical trial. *Psychology of Addictive Behaviors*. 2010; 24(2):265–273. <http://doi.org/10.1037/a0018859>. [PubMed: 20565152]
- Lei H, Nahum-Shani I, Lynch K, Oslin D, Murphy SA. A “SMART” design for building individualized treatment sequences. *Annual Review of Clinical Psychology*. 2012; 8:21–48. <http://doi.org/10.1146/annurev-clinpsy-032511-143152>.
- Li Z, Murphy SA. Sample size formulae for two-stage randomized trials with survival outcomes. *Biometrika*. 2011; 98(3):503–518. <http://doi.org/10.1093/biomet/asr019>. [PubMed: 22363091]
- Loeber R, Tremblay RE, Gagnon C, Charlebois P. Continuity and desistance in disruptive boys' early fighting at school. *Development and Psychopathology*. 1989; 1(01):39. <http://doi.org/10.1017/S0954579400000237>.
- Marlowe DB, Festinger DS, Arabia PL, Dugosh KL, Benasutti KM, Croft JR, McKay JR. Adaptive Interventions in Drug Court: A Pilot Experiment. *Criminal Justice Review*. 2008; 33(3):343–360. <http://doi.org/10.1177/0734016808320325>. [PubMed: 19724664]
- Marlowe DB, Festinger DS, Dugosh KL, Lee PA, Benasutti KM. Adapting Judicial Supervision to the Risk Level of Drug Offenders: Discharge and 6-month Outcomes from a Prospective Matching Study. *Drug and Alcohol Dependence*. 2007; 88:S4–S13.
- McKay JR. Is there a case for extended interventions for alcohol and drug use disorders? *Addiction*. 2005; 100(11):1594–1610. <http://doi.org/10.1111/j.1360-0443.2005.01208.x>. [PubMed: 16277622]
- Murphy SA. An experimental design for the development of adaptive treatment strategies. *Statistics in Medicine*. 2005; 24(10):1455–1481. <http://doi.org/10.1002/sim.2022>. [PubMed: 15586395]
- Nahum-Shani I, Qian M, Almirall D, Pelham WE, Gnagy B, Fabiano GA, et al. Murphy SA. Experimental design and primary data analysis methods for comparing adaptive interventions. *Psychological Methods*. 2012a; 17(4):457–477. <http://doi.org/10.1037/a0029372>. [PubMed: 23025433]
- Nahum-Shani I, Qian M, Almirall D, Pelham WE, Gnagy B, Fabiano GA, et al. Murphy SA. Q-learning: A data analysis method for constructing adaptive interventions. *Psychological Methods*. 2012b; 17(4):478–494. <http://doi.org/10.1037/a0029373>. [PubMed: 23025434]
- O'Brien Caughy M, O'Campo P, Brodsky AE. Neighborhoods, families, and children: Implications for policy and practice. *Journal of Community Psychology*. 1999; 27(5):615–633. [http://doi.org/10.1002/\(SICI\)1520-6629\(199909\)27:5<615::AID-JCOP8>3.0.CO;2-F](http://doi.org/10.1002/(SICI)1520-6629(199909)27:5<615::AID-JCOP8>3.0.CO;2-F).
- Oers MHJ, van Kasa R, Marcus RE, Wolf M, Kimby E, Gascoyne RD, et al. Hagenbeek A. Rituximab maintenance improves clinical outcome of relapsed/resistant follicular non-Hodgkin lymphoma in patients both with and without rituximab during induction: results of a prospective randomized phase 3 intergroup trial. *Blood*. 2006; 108(10):3295–301. [PubMed: 16873669]
- Oetting, AL, Levy, JA., Weiss, RD., Murphy, SA. Statistical Methodology for a SMART Design in the Development of Adaptive Treatment Strategies. In: Shrout, P., Keyes, K., Ornstein, K., editors. *Causality and Psychopathology: Finding the Determinants of Disorders and their Cures*. Arlington, VA: American Psychiatric Publishing, Inc; 2011. p. 179–205.

- Oslin, D. Managing Alcoholism in People Who Do Not Respond to Naltrexone (EXTEND). Bethesda, MD: National Institutes of Health; 2005. Retrieved from <http://clinicaltrials.gov/ct2/show/NCT00115037?term=oslin&rank=8>
- Pachter LM. Do Parenting and the Home Environment, Maternal Depression, Neighborhood, and Chronic Poverty Affect Child Behavioral Problems Differently in Different Racial-Ethnic Groups? *Pediatrics*. 2006; 117(4):1329–1338. <http://doi.org/10.1542/peds.2005-1784>. [PubMed: 16585331]
- Pelham WE, Fabiano GA. Evidence-based psychosocial treatments for attention-deficit/hyperactivity disorder. *Journal of Clinical Child and Adolescent Psychology*. 2008; 37:184–214. [PubMed: 18444058]
- Reid MJ, Webster-Stratton C, Baydar N. Halting the development of conduct problems in head start children: the effects of parent training. *Journal of Clinical Child and Adolescent Psychology: The Official Journal for the Society of Clinical Child and Adolescent Psychology, American Psychological Association, Division 53*. 2004; 33(2):279–291. http://doi.org/10.1207/s15374424jccp3302_10.
- Robins JM. A new approach to causal inference in mortality studies with sustained exposure periods-application to control of the healthy worker survivor effect. *Mathematical Modelling*. 1986; 7:1393–1512.
- Robins LN. Sturdy childhood predictors of adult antisocial behaviour: replications from longitudinal studies. *Psychological Medicine*. 1978; 8(4):611–622. [PubMed: 724874]
- Romeo R, Knapp M, Scott S. Economic cost of severe antisocial behaviour in children--and who pays it. *The British Journal of Psychiatry: The Journal of Mental Science*. 2006; 188:547–553. <http://doi.org/10.1192/bjp.bp.104.007625>. [PubMed: 16738345]
- Ruma PR, Burke RV, Thompson RW. Group parent training: Is it effective for children of all ages? *Behavior Therapy*. 1996; 27(2):159–169. [http://doi.org/10.1016/S0005-7894\(96\)80012-8](http://doi.org/10.1016/S0005-7894(96)80012-8).
- Scott S, Knapp M, Henderson J, Maughan B. Financial cost of social exclusion: follow up study of antisocial children into. *BMJ*. 2001; 323(7306):191–191. <http://doi.org/10.1136/bmj.323.7306.191>. [PubMed: 11473907]
- Shaw, DS., Gross, H. Early childhood and the development of delinquency: What we have learned from recent longitudinal research. In: Lieberman, A., editor. *The long view of crime: A synthesis of longitudinal research*. New York: Springer; 2008. p. 79-127.
- Thall PF, Millikan RE, Sung HG. Evaluating multiple treatment courses in clinical trials. *Statistics in Medicine*. 2000; 30:1011–1128.
- Thall PF, Wooten LH, Logothetis CJ, Millikan RE, Tannir NM. Bayesian and frequentist two-stage treatment strategies based on sequential failure times subject to interval censoring. *Statistics in Medicine*. 2007; 26(26):4687–4702. [PubMed: 17427204]
- Thomas X, Raffoux E, Botton Sd, Pautas C, Arnaud P, de Revel T, et al. Dombret H. Effect of priming with granulocyte-macrophage colony-stimulating factor in younger adults with newly diagnosed acute myeloid leukemia: a trial by the Acute Leukemia French Association (ALFA) Group. *Leukemia*. 2007; 21:453–461. [PubMed: 17252021]
- Wang L, Rotnitzky A, Lin X, Millikan RE, Thall PF. Evaluation of Viable Dynamic Treatment regimes in a sequentially randomized trial of advanced prostate cancer. *Journal of the American Statistical Association*. 2012; 107(498):493–508. [PubMed: 22956855]
- Watkins, CJCH. Doctoral Thesis. University of Cambridge; England: 1989. Learning from delayed rewards.
- Webster-Stratton C. Preventing conduct problems in Head Start children: Strengthening parenting competencies. *Journal of Consulting and Clinical Psychology*. 1998; 66(5):715–730. <http://doi.org/10.1037/0022-006X.66.5.715>. [PubMed: 9803690]
- Webster-Stratton, C., Reid, MJ. The incredible years parents, teachers and children training series: A multifaceted treatment approach for young children with conduct problems. In: Kazdin, AE., Weisz, JR., editors. *Evidence-based psychotherapies for children and adolescents*. New York, NY: Guilford Press; 2003. p. 224-240.
- Yoshikawa H. Prevention as cumulative protection: effects of early family support and education on chronic delinquency and its risks. *Psychological Bulletin*. 1994; 115(1):28–54. [PubMed: 8310099]

Figure 1.

Table 1

Potential objectives of a SMART including specific applications to the illustrative Head Start SMART. Subgroups refer to those shown in Figure 1.

Objective	Intervention Question in Head Start SMART	Subgroup Comparison	Method of Analysis
Stage-specific			
First-stage	Is the best initial intervention to decrease the number of children with behavior problems feedback or to monitor in program?	A, B, C vs. D, E, F	Standard methods of two-sample test or regression
Second-stage	If the child is reported to have behavior issues in February, is parent training effective at decreasing problem behavior compared to monitoring?	B, E vs. C, F	Standard methods of two-sample test or regression
Embedded Adaptive Interventions			
Estimation	At the end of the trial, what is the proportion of children with CBCL/TRF<60 for those who receive feedback and parent training if continued behavior issues or monitored if no continued behavior issues?	A,B	Estimation of effects of one specific AI using weighted and replicated estimation
Comparison	Is the proportion of children with CBCL<60 significantly different between children who (1) begin with feedback and receive parent training if continued behavior problems and are monitored if not or (2) children who begin with monitoring and receive parent training if continued behavior problems and are monitored if not?	A,B vs. D,E	Hypothesis test between two AIs using weighted and replicated estimator and Wald test
Optimization	Is feedback more effective prior to in-home intervention for children who have engaged parents and an experienced teacher?	N/A	Q-learning

N/A indicates not applicable; vs. indicates versus.