

Designing a pilot sequential multiple assignment randomized trial for developing an adaptive treatment strategy

Daniel Almirall,^{a,*†} Scott N. Compton,^b
Meredith Gunlicks-Stoessel,^c Naihua Duan^d and
Susan A. Murphy^e

There is growing interest in how best to adapt and readapt treatments to individuals to maximize clinical benefit. In response, adaptive treatment strategies (ATS), which operationalize adaptive, sequential clinical decision making, have been developed. From a patient's perspective an ATS is a sequence of treatments, each individualized to the patient's evolving health status. From a clinician's perspective, an ATS is a sequence of decision rules that input the patient's current health status and output the next recommended treatment. Sequential multiple assignment randomized trials (SMART) have been developed to address the sequencing questions that arise in the development of ATSS, but SMARTs are relatively new in clinical research. This article provides an introduction to ATSS and SMART designs. This article also discusses the design of SMART pilot studies to address feasibility concerns, and to prepare investigators for a full-scale SMART. We consider an example SMART for the development of an ATS in the treatment of pediatric generalized anxiety disorders. Using the example SMART, we identify and discuss design issues unique to SMARTs that are best addressed in an external pilot study prior to the full-scale SMART. We also address the question of how many participants are needed in a SMART pilot study. A properly executed pilot study can be used to effectively address concerns about acceptability and feasibility in preparation for (that is, prior to) executing a full-scale SMART. Copyright © 2012 John Wiley & Sons, Ltd.

Keywords: pilot; experimental; adaptive; individualized; intervention; dynamic treatment regimes

1. Introduction

With the establishment of evidence based treatments for many chronic conditions, there is growing interest and need for research on how to adapt and readapt treatments to maximize clinical benefit. That is, there is growing interest in developing health interventions that are individualized to the patient and which respond over time to the needs (successes, benefits) of the patient. In the field of mental health, for example, the director of the National Institute of Mental Health (NIMH) recognizes the current state of interventions research: 'To improve outcomes we will need to [develop treatments which]... personalize care based on individual responses' [1].

Effective clinical management of chronic conditions (such as psychiatric disorders) often requires a sequence of treatments, each adapted to individual response, and hence multiple treatment decisions

^aSurvey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, USA

^bDepartment of Psychiatry & Behavioral Sciences, Duke University Medical Center, Durham, NC, USA

^cDepartment of Psychiatry, University of Minnesota, Twin Cities, MN, USA

^dNew York State Psychiatric Institute, Columbia University, New York, NY, USA

^eDepartment of Statistics & Institute for Social Research, University of Michigan, Ann Arbor, MI, USA

*Correspondence to: Daniel Almirall, 2204 Institute for Social Research, University of Michigan, 426 Thompson Street, University of Michigan, Ann Arbor, MI, USA.

†E-mail: dalmiral@umich.edu

throughout the course of an individual's clinical care. For example, in child and adolescent mental health, the American Academy of Child and Adolescent Psychiatry practice parameters for pediatric depressive disorders recommend antidepressants following a nonresponse to initial psychotherapy [2]. Similarly, for pediatric anxiety disorders, an augmentation strategy of medication is recommended for children who show partial response to first-line psychotherapy [3]. Sequential treatments, in which treatments are adapted over time, are often necessary because treatment outcomes are heterogeneous across patients, treatment goals change over time, and in the long-term it is necessary to balance benefits (e.g., symptom reduction) with observed and potential risks (e.g., unwanted side-effects, patient burden) [4, 5]. As a result, clinicians often find themselves implicitly engaging in a sequence of treatments with the goal of optimizing both short-term and long-term outcomes.

Adaptive treatment strategies (ATSS) [4–7] formalize such sequential clinical decision making. An ATS individualizes treatment via decision rules that specify whether, how, and when to alter the intensity, type, or delivery of treatment at critical clinical decisions. Examples of critical decisions include, which treatment to provide initially, how long to wait for the initial treatment to work, how to determine whether the initial treatment worked or not, and which treatment to provide next if the initial treatment is or is not working. Treatments at each critical decision may include medications, behavioral interventions, or some combination of these two. The following is an example of an ATS following an initial diagnosis of pediatric generalized anxiety disorder:

ATS 1: 'First treat with the medication sertraline (SERT) for 12 weeks. If the child has not achieved an adequate response to initial SERT (at week 12), *augment* by initiating a combination of sertraline + individual cognitive behavioral therapy (CBT) for 12 additional weeks; otherwise, if the child shows adequate response, *maintain* SERT alone for another 12 weeks'.

ATSS should be explicit; for instance, in ATS1, 'adequate response' may be defined as the child exhibiting a value on a symptom scale beyond a prespecified cut-off (more on this topic in Section 4 below). From the perspective of the child and his/her parent(s), the ATS is a sequence of treatments: for example, SERT for 12 weeks, followed by CBT for another 12 weeks (assuming an inadequate response to SERT alone). From the perspective of the clinician, the ATS is a clinical decision rule that guides treatment both initially and also following an assessment of the 12-week response status. ATSS are also known as adaptive interventions and dynamic treatment regimes [7–22].

Sequential multiple assignment randomized trials (SMARTs, discussed in more detail below) have been proposed to facilitate or accelerate the development of ATSS and represent an important advancement in clinical research methodology [5, 23–25]. SMARTs can be used: (1) to discover which treatments work together in a sequence to lead to improved outcomes; (2) to investigate the interplay between trajectories of change in illness and the development of sequences of treatments; (3) to compare different sequences of medication, behavioral treatments or treatment tactics (e.g., treatment delivery methods); and (4) to investigate the clinical utility of both biological (e.g., genetic information) and clinician-observable data for individualizing treatment sequences. A central aim of SMARTs is to inform the construction of an optimized ATS, that is, to develop the sequence of treatments that lead to optimal outcomes in the long-term. Furthermore, because at their core SMARTs are concerned with the identification and use of which treatments work best for whom, when, and under what circumstances, mental health research that employs SMART designs fits squarely within the domain of comparative effectiveness research, which is another national research priority [26].

Sequential multiple assignment randomized trials are gaining rapid acceptance in the clinical and health services research community. The seminal NIMH-funded trials CATIE [27] and STAR*D [28, 29] in schizophrenia and depression, respectively, were early precursors to SMARTs and represent important trials in terms of encouraging researchers to consider the development of ATSS. In oncology, trials similar to SMARTs were conducted in the early 1990s [30]; see the work of Thall *et al.* [22] for more recent work. In mental health and substance abuse research, a number of SMARTs have been completed or are currently on-going, including trials in alcohol dependence (D. Oslin, personal communication), attention deficit hyperactivity disorder (W. Pelham, personal communication), cocaine and alcohol dependence (J. McKay, personal communication), substance abuse problems in pregnant women (H. Jones, personal communication), and autism (C. Kasari, personal communication).

Despite their promise, increasing popularity, high quality fit for research in individualized medicine, and adoption by some clinical trialists, SMARTs are relatively new in clinical research. Because of their novelty, and because SMARTs represent a departure from the standard two-arm randomized clinical trial, questions remain about the use and execution of SMARTs by clinical investigators. Primary

among these concerns is the issue of feasibility. Study sections, grant-funding review boards, and other stakeholders often want to see evidence that the proposed SMART study design is feasible. This includes showing evidence that investigators have the experience to execute the SMART design properly.

External pilot studies have long been used in all areas of health research to address concerns about feasibility. The primary aim of this article is to provide guidance on executing a SMART pilot study in preparation for a full-scale SMART. In Section 2 we review the goals of pilot studies. In Section 3 we review the SMART design, and we introduce a motivating example SMART in pediatric anxiety disorders. In Section 3, we also review the types of scientific questions that a full-scale SMART can examine. In Section 4 we propose and discuss design considerations within the context of a pilot study in preparation for a full-scale SMART. In Section 5, we provide general guidance on how to choose the sample size for a SMART pilot study. To make ideas concrete, we focus on the pediatric anxiety disorders example SMART throughout the article; however, all of the main ideas presented in this manuscript extend readily to SMARTs used to develop and optimize ATSS for other chronic disorders.

2. Pilot studies

Following the lead of a diverse set of researchers, statisticians, and methodologists [31–36], we define a pilot study as a small-scale version of the larger study with the aim of fine-tuning the study design, evaluating its feasibility and acceptability, and preparing the research team for a future ‘full scale’ randomized trial. In this manuscript, ‘feasibility’ means both the ability of the investigators to execute the SMART, and the ability to treat participants with the ATSS that comprise (i.e., that are embedded in) the SMART (see below). By ‘acceptability’ we mean the tolerability or appropriateness of the ATSS (including assessment procedures that make up the ATSS, see Section 4 below) being studied from both the perspective of study participants and clinicians.

There is a distinction in the statistical literature concerning internal versus external pilot studies; the definition given above (which we use throughout) is consistent with the definition of an external pilot study. A common aim of internal pilot studies, on the other hand, is to improve sample size calculations during the execution of the already-developed full-scale trial. In this case, they use the first prespecified number of subjects to recalculate the sample size needed for the remainder of the full-scale trial. As such, internal pilot studies are not concerned with examining feasibility and acceptability; for more information on internal pilot studies, consult the work of Wittes and colleagues [37–39].

A well-designed and executed external pilot study helps answer questions such as: ‘Are we able to deliver properly the interventions we are proposing to compare?’ ‘What is the level of staff understanding and fidelity to the research protocol?’ ‘Are the proposed adaptive interventions acceptable to participants?’ ‘Should we devise special quality control measures and procedures to improve and maintain fidelity during the large scale trial?’

Pilot studies also offer the opportunity to fine-tune proposed interventions and may provide preliminary knowledge about the direction of its effect. Pilot studies can also inform whether a proposed intervention study is worth pursuing in its current form. For example, if feasibility and acceptability are found to be lacking beyond what can be achieved by fine-tuning, the outcome of a pilot study may be that a second pilot or a new study is necessary. Under this definition, the primary role of a pilot study is not to test hypotheses about the potency of a given intervention, nor to obtain information about effect sizes with any certainty. For example, this idea is shared in the NIMH’s pilot study program announcement (R34; <http://grants.nih.gov/grants/guide/pa-files/PAR-09-173.html>), which explicitly states that ‘...conducting formal tests of outcomes or attempting to obtain an estimate of an effect size is often not justified.’

The primary aim of this article is to provide guidance on executing a SMART pilot study in preparation for a full-scale SMART. Beginning in Section 4, we discuss scientific, statistical, and logistical issues specific to executing a SMART that should be considered in a SMART pilot. This article does not provide guidance on executing pilot studies in general; rather, we focus on the unique aspects of SMART designs. Scientists preparing to execute a randomized trial should also refer to the literature on executing external pilot studies for pointers on more general uses of pilot studies [34–36].

3. Sequential multiple assignment randomized trials

The overarching aim of a SMART is to inform the construction of an optimized ATS. The key feature of a SMART is that it allows investigators to evaluate the timing, sequencing, and adaptive selection of

treatments in a principled fashion by use of randomized data. In a SMART, participants can move through multiple stages of treatment; each stage corresponds to a critical decision, and participants are randomized at each stage/critical decision. Randomized treatment options at each critical decision include appropriate single-component or multicomponent treatment alternatives.

An example of a SMART is shown in Figure 1. This SMART can be used to develop an ATS for the management of pediatric generalized anxiety disorder involving the medication sertraline (SERT), cognitive behavioral therapy (CBT), and a combination of both (COMB). This SMART provides data that help investigators address two critical decisions: (1) ‘Which treatment to provide first?’ and (2) ‘Which treatment to provide to non-responding participants?’ Because the answer to each question may depend on the answer to the other question, the SMART involves two stages of treatment, one per critical decision. Participating children are first randomly assigned to either 12-weeks of SERT or 12-weeks of CBT as first-line treatment. After the end of the initial 12 weeks of treatment, each child’s response to treatment is evaluated and classified as either a treatment responder or treatment nonresponder. This binary indicator is the primary tailoring variable used as part of the SMART: children who are not responding at the end of 12 weeks are rerandomized between an augmentation of their initial treatment or a switch in treatment, whereas children who do respond continue with their initial treatment. As indicated in Figure 1, the primary outcome could be a longitudinal measure of anxiety over the 48-week trial period.

By using sequenced randomizations, SMARTs ensure that at each critical decision, the groups of participants assigned to each of the treatment alternatives are balanced in terms of both observed and unobserved participant characteristics. This includes time-varying characteristics and outcomes experienced during prior treatment such as symptom levels, side effects, and adherence.

All SMART designs have multiple ATSs embedded within them. For example, in addition to the ATS described in the Introduction (ATS 1, sub-group A + B), the example SMART shown in Figure 1 also includes the following three additional ATSs:

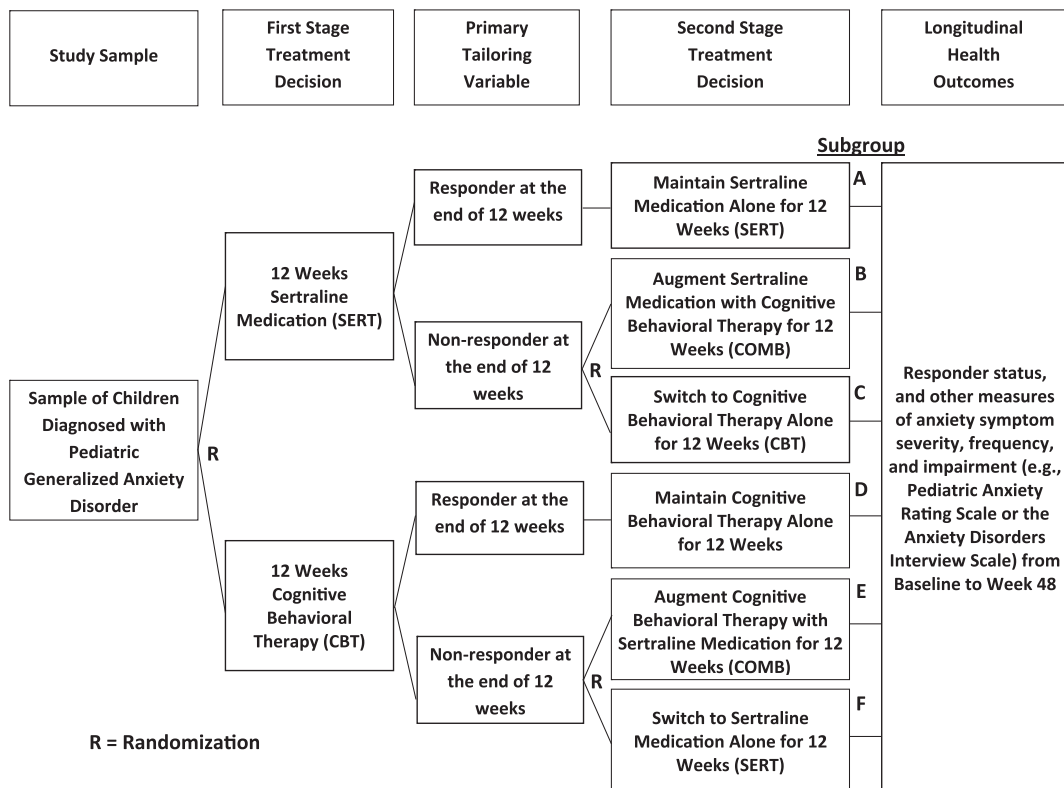


Figure 1. This example SMART can be used to develop an adaptive treatment strategy involving sertraline medication (SERT), cognitive behavioral therapy (CBT), and their combination (COMB) for the management of pediatric anxiety disorders.

- ATS 2:** First treat with SERT only for 12 weeks. Then, if the child does not respond to initial SERT, switch to CBT alone for 12 additional weeks; otherwise, if the child responds to initial SERT, maintain on SERT alone for another 12 weeks (subgroup A + C)
- ATS 3:** First treat with CBT only for 12 weeks. Then, if the child does not respond to initial CBT, augment treatment by initiating a combination strategy (COMB) of sertraline + CBT for 12 additional weeks; otherwise, if the child responds to initial CBT, maintain on CBT alone for another 12 weeks (subgroup D + E).
- ATS 4:** First treat with CBT only for 12 weeks. Then, if the child does not respond to initial CBT, switch to SERT medication for 12 additional weeks; otherwise, if the child responds to initial CBT, maintain on CBT alone for another 12 weeks (subgroup D + F)

The four ATSs embedded in the SMART shown in Figure 1 are also described in Table I. These are the only ATSs embedded in this example SMART.

A full-scale SMART can be used to evaluate a variety of primary and secondary scientific questions to inform the development of an optimal ATS. SMARTs are factorial designs in a sequential setting [21]; thus the primary aims usually involve main effects. One example of a primary aim is ‘What is the main effect of first-line treatment?’ For example, in Figure 1, this involves a comparison of first-stage SERT versus first-stage CBT (subgroup A+B+C versus subgroup D+E+F). Note that this main effect comparison averages over the second stage treatments. A SMART can also be used to contrast two or more ATSs. For example, in Figure 1, an investigator may be interested in examining which of the four ATSs leads to the most rapid reduction in symptoms over the course of 48 weeks because of initial diagnosis. Because some participants in the SMART simultaneously are consistent with multiple embedded ATSs (e.g., participants in subgroup A are consistent with both ATS1 and ATS2), specialized methods that account for the multiple use of subjects are used to estimate and compare mean outcomes under different ATSs [14, 15]. A full-scale SMART also provides investigators the opportunity to investigate the use of time-invariant (e.g., patient characteristics such as baseline severity, demographic variables or genetic information) and other time-varying (e.g., adherence with treatment) tailoring variables for improving sequential treatment (these may be the exploratory aims of a SMART). For instance, the example childhood anxiety SMART could be used to explore if treatment adherence and treatment satisfaction over the course of the initial 12 weeks are important predictors of second stage treatment response such that they may be useful additional tailoring measures. This can be done using standard treatment-by-covariate moderator analyses of the impact of second-line treatments on subsequent outcomes — for example, ‘Among nonresponders, does adherence during the initial 12 weeks of treatment moderate the impact of subsequently augmenting versus switching treatments during the next 12 weeks?’ — or using more sophisticated data analytic methods such as dynamic regime marginal structural model estimation [14], iterative minimization and G-estimation of the structural nested mean model [10, 12, 16], or Q-Learning ([4, 40] for software, see <http://methcenter.psu.edu>). In other SMARTs, the duration of the first stage of treatment may differ between trial participants; for example, this may occur when the second stage

Table I. The four adaptive treatment strategies embedded as part of the example SMART design shown in Figure 1.

Label	Adaptive treatment strategy			Type of strategy*	Subgroup**
	First stage treatment	Primary tailoring variable	Second stage treatment		
ATS1	SERT	Responder Nonresponder	SERT SERT+CBT	Maintain-Augment	A + B
ATS2	SERT	Responder Nonresponder	SERT CBT	Maintain-Switch	A + C
ATS3	CBT	Responder Nonresponder	CBT SERT+CBT	Maintain-Augment	D + E
ATS4	CBT	Responder Nonresponder	CBT SERT	Maintain-Switch	D + F

* All of the ATSs embedded in the SMART in Figure 1 maintain (or continue) the first-stage treatment as a second-stage treatment if the subject is an early responder to first-stage treatment.

** These are the subgroups from Figure 1 containing participants who are consistent with the ATS.

begins at an event time (as in the *ExTEND* trial example below). In such cases, the duration of first-stage treatment could also be examined in secondary analyses for its potential usefulness as a tailoring variable. Hence, SMARTs are used not only to test treatment options at particular stages of treatment and to contrast embedded ATs, but also yield high-quality data to inform the clinical utility of individualizing treatment using additional tailoring.

Consistent with the overarching aim of a SMART, data analyses associated with these primary and secondary aims avoid choosing the best treatment at each critical decision point using separate one-at-a-time optimizations. To appreciate why this is important, consider that a first-line treatment leading to poorer outcomes in the short-term may lead to better outcomes in the long-term when considered as part of a whole ATs. For instance, in the SMART in Figure 1, it is possible for first-stage CBT to lead to poorer (or similar) symptom relief at the end of 12 weeks relative to first-stage SERT; yet when evaluated at 48 weeks, beginning with CBT results in lower symptoms than beginning with SERT. This can happen, for example, if the initial CBT sets the stage for a more pronounced response to COMB among nonresponders in the second stage (say, by priming the individual to take advantage of the subsequent CBT) than initial SERT. Or to consider another case, suppose that the CBT provided as part of COMB includes components designed to increase adherence to medication. Here, SERT may be the better initial treatment when considered part of a sequence; this can happen if initial SERT reveals prescriptive information: initial treatment by SERT may be better than CBT at identifying individuals who are poor adherers, and thus indicate who needs COMB (which, in our example, includes components to improve adherence). Other conjectures are possible. The key point here is that developing an ATs by piecing together the treatments that are best myopically (i.e., work best at the end of each decision point) may be a suboptimal way to proceed in terms of long-term outcomes.

It is important to understand that the SMART is not an adaptive trial design [41,42]. Just as the standard two-arm RCT is a fixed trial design, the SMART is also a fixed trial design that does not change during the course of the trial. What is adaptive about the SMART are the treatment strategies embedded within the SMART (e.g., ATs1–ATs4, described above). It is conceivable, of course, to conduct a SMART using an adaptive trial design by allowing some design parameters (e.g., sample size or sample selection) to vary with interim data; this topic, however, is outside the scope of this article.

Sequential multiple assignment randomized trials can be seen as developmental trials used to construct and optimize an ATs. Following the successful completion of a SMART, the constructed ATs can be tested against a usual care treatment (or other state-of-the-art intervention) in a standard RCT. In such a confirmatory trial, participants would be randomly assigned to either the state-of-the-art intervention or to the SMART-optimized ATs to test which treatment strategy is more effective. SMARTs can also be sized to have other roles besides that of a developmental trial. For example one could size a SMART to conduct a comparison of the embedded ATs; this may be particularly attractive when one of the embedded ATs represents ‘usual care.’

4. Considerations to address in a SMART pilot

In this section we discuss nine topics specific to executing a SMART that can be considered in a SMART pilot. Table II provides examples of issues/concerns that may arise (specific to each topic) and how they may lead to changes in the full-scale SMART protocol.

4.1. The primary tailoring variable

One area unique to a SMART is the importance placed on tailoring variables. In preparation for a SMART, a key exercise is to undertake a thorough discussion about the primary tailoring variable. (Note that in the ATs embedded within a SMART, the primary tailoring variable is used to adaptively determine the next treatment; in contrast, in the SMART design, the primary tailoring variable determines the set of randomized treatment options.) This involves brainstorming about how to determine early signs of nonresponse and when this determination should be made. As part of this discussion, investigators will need to decide how to assess early response/nonresponse (e.g., the Clinical Global Impression-Improvement Scale (CGI-I) [31]?), and what criterion or cut-score should be used (e.g., less than 3 on the CGI-I?). Other considerations include: how sensitive is the measure to treatment change? Is there an established precedent in the literature that can be used to justify the measure as a tailoring variable? Would the measure be feasible in real-world settings?

Table II. Feasibility or acceptability issues/concerns and subsequent changes in the full-scale SMART protocol. Examples are given by topic addressed in Section 4.

Topic	Examples of issues discovered in the SMART pilot	Possible changes to the SMART or in preparation for the SMART
Primary tailoring variable	A long paper and pencil assessment, more appropriate for research than for practice, was used. Mid-way through the pilot, investigators noticed that it was too long and burdensome for both study participants and first-stage clinicians. This led to missing or delays in items used to assess early non/responder status and, as a result, an inability to randomize second-stage treatment in a timely manner.	A new version of the primary tailoring variable embedded within the electronic medical record (EMR), which was more appropriate for clinical practice, was implemented. Feedback from attending clinicians in the pilot study was used to develop the new assessment, which was tested during the second-half of the pilot study prior to rolling it out in the full-scale SMART.
Randomization procedure	A real-time sequentially randomized allocation procedure was first used. During the pilot, many nonresponders to first-stage treatment who were not adherent ended up (by chance) in the second-stage augmentation arms. Furthermore, there were difficulties in communication between the attending clinician and study coordinators to obtain second-stage treatment allocation at the end of the 12 week clinic visit, resulting in some subjects who left the clinic without an assigned second-stage treatment.	The allocation procedure was modified for the full-scale SMART to include adherence as a stratification measure in the second randomization. Furthermore, additional steps were taken to make it easier to obtain treatment allocations at the end of the 12 week clinic visit based on early non/responder status; and an automated reminder from the EMR was added where second-stage clinicians would be notified that the treatment for the participant has changed.
Missing primary tailoring variable	Children who missed the 12 week clinic visit but subsequently returned for treatment, were all classified as nonresponders, and re-randomized at the time of the post 12-week visit to second-stage treatment. At the time, it was intuitive to define the embedded ATSS this way because investigators thought children who missed clinic visits did so because they were having more difficulties. However, this led to problems as missed visits were, in fact, not always related to how poorly the child was doing.	For the full-scale trial, a different approach was used whereby children who missed the 12 week clinic visit were labeled as non/responder based on fixed rule which involved the last known non/responder status. This new rule for handling missed visits was a better reflection of actual clinical practice.
Other potential tailoring variables	Investigators piloted a measure of adherence to first-stage treatment. This measure was to be examined as a tailoring variable for more refined treatment tailoring in secondary analyses. Adherence measurements were taken every 6 weeks (hence, by the time children were ready to move to second-stage treatment, there were 2 adherence measures). However, many participants would stop adhering prior to the first adherence measurement at 6 weeks.	For the full-scale trial, adherence was collected more frequently. This will allow investigators in the full scale trial to examine 'time until nonadherence' as a potential tailoring variable leading to more refined ATSS.
Identifying unanticipated tailoring variables	During the pilot study, clinicians noted (qualitatively, that is, based on intuition) that families of nonresponders who were more difficult to manage seem to have benefitted more from switch rather than augmentation. This came up in pilot study meetings between investigators and study clinicians. Some clinicians argued that this was a due to participant dissatisfaction with first-stage treatment; that is, these families were more difficult to manage because they were not satisfied with first-stage treatment.	In the full-scale trial, satisfaction, preference, and additional process data was collected by study coordinators such as tardiness to clinic visits and number of re-schedules. These data will be used in secondary, hypothesis-generating analyses of the full-scale SMART to determine if it is useful (in conjunction with other clinical measures) for developing a more refined ATSS.

Table II. (Continued)

Evaluation assessments versus treatment assessments	During the course of the SMART pilot, research staff — not fully understanding the distinction between the SMART study versus the ATSS embedded within it — used evaluation assessments to supplement clinician assessments in the determination of early non/responder status.	SMART quality-control procedures were put in place to prevent evaluation assessments from being used to determine early non/responder status.
Staff acceptability and fidelity to changes in treatment	In a small number of cases, psychiatrists (psychologists) treating with first-stage SERT (CBT) held-off on labeling cases as nonresponders to delay a possible switch to CBT (SERT). Beyond being a violation of the SMART study protocol, this was possible indication that defining the embedded ATSS on the basis of a 12 week early non/response criteria was not appropriate.	SMART quality-control procedures were enacted for the full-scale trial to prevent this from happening. Furthermore, based on feedback from clinicians that 12 weeks was insufficient time in first-stage treatment, in the full-scale trial early non/response criteria leading to second-stage treatment was defined at 14 weeks, instead.
Participant concerns about changes in treatment	During the pilot study, investigators learned that among nonresponders initially receiving CBT, subsequent switch to SERT (medication) led to substantial treatment nonadherence as well as drop-out. Focus group discussions revealed that there were parental concerns about medication, in general, and with medicating children at this young age, in particular. Parents reported that there was no time or forum during clinic visits where they could bring up these concerns about medication. The problem was mainly in the participants who switched from CBT to SERT, but group discussions revealed that similar problems were also occurring anytime SERT was being offered (including as first stage treatment).	A manualized medical management (MM) module, successfully used in a previous trial, was to be included in conjunction with SERT. MM included a brief clinician-administered educational intervention (using motivational enhancement techniques, as well as print materials) for improving adherence, discussing the benefits/costs/side effects/rationale for medication, and a format for families to raise their concerns over medications. The new MM+SERT (which was to replace SERT) was piloted with the remaining pilot study participants and later used in the full-scale SMART.
Ethical considerations and consent procedures	In the SMART pilot, participants consented to be part of the study up-front and it was explained that this meant the participant could receive one of various treatment sequences. Focus group discussion revealed that some participants did not find second-stage treatments helpful/acceptable and dropped out of the study shortly after learning about their second-stage treatment assignment. Research staff incorrectly understood this to mean that participants were not consenting to the randomly allocated second-stage treatment. As a result, these participants were not followed up for subsequent outcome assessments.	Consent procedures were revised prior to the full-scale SMART. A quality control procedure was implemented to remind research staff that study consent occurs up-front only (even though a real-time randomized allocation procedure was being used) and that failure to accept second-stage treatment assignment did not preclude research staff from attempting to obtain outcome assessments.

In addition to identifying how to assess the tailoring variable, investigators must determine how frequently the tailoring variable needs to be assessed. That is, how often should early response/nonresponse be evaluated? This will depend in large part on the domain being studied and historical precedent. In our example in Figure 1, response/nonresponse is assessed at one point in time (12 weeks) and a score ≤ 2 on the CGI-I is the criterion. Participants who are not responding at 12 weeks are rerandomized to the next stage treatment. In this example — which uses a primary tailoring variable fixed at a prespecified number of weeks after the initiation of first-stage treatment — it is necessary for investigators to operationalize what they mean by ‘end of 12 weeks’. This is because, as with any study, it is not always feasible to schedule clinic visits at exactly ‘the end of 12 weeks’ because of scheduling conflicts. Often, for example, a prespecified window of time around 12 weeks would be used. Indeed, this issue is not unique to SMARTs. However, in a SMART the width of the window deserves particular attention because different

window lengths imply different operationalizations (definitions) of the ATSS, which are embedded in the SMART. The related issue of how to define the primary tailoring variable when non/responder status is missing is discussed below.

In other SMARTs, multiple assessments for early response and/or nonresponse can occur and the primary tailoring variable is a summary of these multiple assessments. Or the primary tailoring variable may be defined as a ‘time until’ outcome of first-stage treatment. For example, in a SMART concerning alcohol dependence (the ExTEND clinical trial, D. Oslin, personal communication), counts of heavy drinking days are used to measure response/nonresponse and participants are assessed weekly to ascertain the number of heavy drinking days occurring over the prior week. Here the participant is deemed to be an early nonresponder as soon as ≥ 2 heavy drinking days occur. Note that in the ExTEND example, the primary tailoring variable — defined as a ‘time until’ measure — requires more frequent monitoring times (compared with a primary tailoring variable assessed at the end of first-stage treatment, as in Figure 1) adding logistical complexity. The pilot study can be used to examine the feasibility of assessing and using such a measure.

In other example SMARTs, the primary tailoring variable need not be dichotomous (e.g., it may be a trichotomous variable measuring responder, nonresponder, and partial responder status). However, the more complicated the primary tailoring variable, then the more complex the trial design becomes because the randomized treatment options may differ by values of the primary tailoring variable. In general, the choice of primary tailoring variable should be driven by a primary, parsimonious scientific question. Another advantage of parsimony in the choice of primary tailoring variable is that it allows for observed variability on other measures, which may be useful for building ATSS that are more refined (i.e., which may offer more individually-tailored treatments) than those embedded in the SMART by design. These additional measures would be considered in secondary analyses of the data arising from a SMART. That is, other, possibly more interesting, scientific questions involving more refined tailoring (such as considering how to tailor treatment using a less coarse/continuous tailoring variable, or such as using adherence to first-stage treatment to decide how to treat nonresponders in the second-stage) can be addressed as part of secondary analyses; for more on this topic, see Section 4.4.

The SMART pilot should allow the investigative team ample opportunities to train in applying the approach for assessing and using the primary tailoring variable, assessing whether the approach is clinically feasible, and refining the measurement of the primary tailoring variable and refining the criterion for determining early response/nonresponse during the full-scale trial.

4.2. *Randomization procedure*

In a SMART, participants are randomized at multiple critical decisions over the course of the trial. Investigators can choose between two randomization procedures: an up-front approach or a real-time approach. In the up-front approach, participants are randomized at the beginning of the trial to the different ATSS that are embedded in the SMART design. In our example, this means randomizing participants at baseline to one of the four ATSS described in Section 2. In the real-time approach, participants are randomized sequentially at each critical decision point as described in Figure 1. In both approaches, participating families will be informed during the consent process of the possible treatment sequences to which they might be randomized. Compared with the up-front approach, the real-time approach has at least one important advantage: it allows investigators to capitalize on information (including time-varying covariates) available at time of randomization to ensure balance in assigned treatment options at each critical decision stage. For instance, if we use a real-time approach in our example, the second randomization among nonresponders (to initial SERT or initial CBT) can be stratified on adherence to treatment, symptom severity, or other important outcomes observed during the first 12 weeks of treatment. This is advantageous because if, by chance, the composition of groups differs by these variables and these variables are prognostic for subsequent study outcomes, then differences between the groups can be attributed to these compositional differences as opposed to differences between second stage treatments. The up-front approach does not afford investigators this level of control over compositional balance. A SMART pilot study will give the research team’s analyst an opportunity to develop and evaluate the randomization procedure and check for unanticipated errors.

4.3. *Missing the primary tailoring variable*

In our example, the primary tailoring variable is response status (responder/nonresponder) at the end of acute treatment (week 12). Ideally, all participants in a SMART will have a measure of the primary

tailoring variable available at the end of 12 weeks that can be used to guide subsequent treatment assignments. However, this ideal situation might not hold. That is, the assessment of a given participant's response to the first 12 weeks of treatment may be missing, either because the participant dropped-out of the study prior to week 12, or because the participant was unavailable for the week 12 assessment (e.g., the participant might be ill or on vacation). The problematic situation for purposes of executing the SMART is the latter one, in which a participant is unavailable for the 12-week assessment yet returns to the study at some later point when a decision must be made concerning response/nonresponse status and next step in treatment.

All randomized trials must contend with missing data, but this problem is unique in SMARTs because the interventions (e.g., the ATs) that make up a SMART are adaptive. The critical issue here is how to manage missingness for purposes of offering/assigning subsequent treatment, as opposed to how to handle missing evaluation outcomes for purposes of data analysis. As in standard randomized trials, the pilot can be used to prepare and practice low burden approaches that can be used to facilitate the collection of the missing evaluation outcomes (e.g., via telephone assessments). Beyond this, a satisfactory solution to the 'missing tailoring data problem' (as opposed to 'missing evaluation data problem') is one that recognizes that this type of missingness should be part and parcel of the definition of the ATs embedded in the SMART, just as in typical clinical practice the clinician has to decide next treatment when faced with missed visits. Therefore, the solution to this problem should be guided by what would be achieved in clinical practice. Clinical investigators should ask 'How do I treat a patient when s/he returns after a missed clinic visit and what do I need to know concerning the missed visit to make this decision?' Investigators may need to differentiate between excused (e.g., family could not find childcare for younger siblings) and unexcused missed assessments, how long the participant was missing, how many sessions the participant attended in the first treatment phase, how well or poorly the participant was doing prior to missing, or how well the participant is doing when s/he reappears. The actual approach used will depend on the particular research question(s) being investigated and the types of disorders and treatments being studied. Consistent with these ideas, the solution to this problem involves having a fixed, prespecified way to determine subsequent treatment in the presence of a missed clinic visit. This can be operationalized in at least two ways in the SMART: First, missingness could be made part of the definition of early nonresponse. This approach could be taken if in actual practice, a missed clinic visit is clinically viewed as nonresponse (this is often the case in substance abuse treatment). One way to operationalize this is to classify all participants with missing response status at any given decision point as nonresponders, then assign the participant their randomized treatment option at the next clinic visit. This approach could be labeled 'nonresponding until proven responding'. The opposite approach, or 'responding until proven nonresponding,' could also be employed whereby a participant missing the responder/nonresponder status is classified as a responder for purposes of subsequent treatment assignment. Another option in the case of a missing 12-week visit is to devise an approach that relies on data that would be readily available to the clinician in practice, including previous response to treatment and current clinical status up to the point of missingness to determine the responder status. This option requires investigators to decide how to summarize the observed history of treatment response, including the decision of how much historical data to use. A second way to operationalize missingness as part of the ATs embedded in the SMART is to treat it separately from non/responder status and offer a separate treatment altogether. This approach may be more appropriate than the above approach if the second stage treatment options are simply not feasible for subjects exhibiting this type of missingness. Importantly, no matter what approach is used, the choice of subsequent treatments in the presence of missingness should be well-specified and fixed *prior* to the trial. The SMART pilot will allow the investigative team to train in applying the chosen approach and assess whether it is clinically feasible and scientifically relevant.

4.4. *Other potential tailoring variables*

The investigative team must also decide which additional potential tailoring variables should be collected. Potential tailoring variables include both baseline patient characteristics and time-varying measures (e.g., treatment adherence or side effects) that might be useful in tailoring the treatment to the patient. In the childhood anxiety example, investigators may want (in the full-scale SMART) to explore whether patient characteristics and baseline measures might be used to tailor initial treatment, and whether patient characteristics, baseline measures, and outcomes because of initial treatment (but collected prior to the subsequent critical decision point) may be used to tailor the second treatment. A time-varying tailoring variable can be measured at a single point (e.g., the week 12 clinic visit) or may

be a cumulative summary of treatment response up to that point in time. Potential tailoring variables should be simple, easy to use (minimal burden) in actual clinical practice — for example, short-form of instruments — and able to be collected by the treating clinician. The SMART pilot can be used to pilot such instruments, items, or questions under consideration for tailoring. Although these variables will not impact the full-scale trial (unlike the primary tailoring variable), they could lead to the creation of new variables useful for tailoring treatment in future studies or found to be important when the data are analyzed.

The use of tailoring variables represent an important departure from standard randomized trials, even when designed with an interest in understanding moderators which predict treatment response [44], because they are rarely considered outcomes to initial treatment, such as symptom, side effect, and adherence measures, which may predict later outcomes to second-stage treatments. This is one important reason why SMARTs more closely mirror clinical practice and will ultimately lead to information that is more clinically relevant to the practicing clinician.

4.5. Identifying unanticipated tailoring variables

Just as the SMART pilot can be used to practice the measurement of new tailoring variables identified *a priori*, as described above, it may also be helpful in identifying unanticipated variables that can be useful for tailoring treatment and can then be measured in more detail in a full-scale trial. Focus groups or structured exit interviews scheduled during and after the SMART pilot will likely be helpful in uncovering new and potentially important tailoring variables. Such focus groups can also be used with nonclinician members of the treatment team (e.g., research assistants, project coordinators). For example, in the context of our example SMART, families who are more difficult to schedule frequently, arrive late to visits, require repeated reminders, rush through paperwork, and more challenging to work with may benefit more from one treatment over another when compared with families who are highly compliant and easier to manage.

4.6. Evaluation (or SMART) assessments versus treatment (or ATS) assessments

In a SMART, a clear distinction is made between research assessments for purposes of data analysis to evaluate the effectiveness of ATSS (data used in evaluation) versus assessments made as part of the ATS to inform subsequent treatment assignment (data used in tailoring). Indeed, keeping these assessments distinct is not entirely unique to SMARTs; in standard two-arm RCTs it is equally important to differentiate between information gathered and shared as part of treatment versus information gathered for purposes of evaluating treatment effectiveness. However, in SMARTs it is important to further highlight this distinction because of the realistic possibility that ATSS embedded within the SMART could unknowingly become ill-defined if evaluation data is implicitly or explicitly used in the determination of early non/response during the conduct of the SMART.

In our example, the week 12 response status is assessed by the treating clinician as part of the ATS (data used in tailoring). Because the aim of the SMART study is to inform actual clinical practice, it is acceptable for the week 12 response status (used to inform subsequent treatment randomizations) to be an unblinded clinician evaluation. However, other assessment measures are also collected at the week 12 visit and are used to determine the effectiveness of treatment(s) (data used in evaluation). The key difference is that the latter assessments are not part of the embedded ATSS. If possible, it is important to use blinded independent evaluators (clinicians not involved in the provision of treatment) to collect the outcome measures that will be used to evaluate the effectiveness of the ATS or its components. It is equally important to ensure that only data used in tailoring (i.e., as part of the embedded ATSS) is used to determine subsequent treatment changes. The SMART pilot will provide staff an opportunity to prepare and practice these assessment methods and strategies for maintaining these two types of assessments separate and distinct. Fundamentally, this distinction is about understanding the distinction between SMART versus the ATSS embedded within the SMART.

4.7. Staff acceptability and fidelity to changes in treatment

A properly executed SMART requires careful staff fidelity to changes in treatment provided over time as dictated by the study design. This may be challenging because (1) clinical researchers accustomed to participating in standard randomized trials may have little experience with sequenced treatments that are an explicit part of the SMART research protocol and (2) following the SMART protocol may

limit the use of clinical judgment. Prior to a full-scale SMART, a pilot SMART can be used to identify concerns clinicians may have about the sequence of treatments offered and the assessment of early response/nonresponse. The pilot can be used to develop training procedures to enhance clinician fidelity to both the research protocol and treatment strategies and to ensure that the clinical team has the required training and expertise needed to successfully carryout the SMART. For instance, in our example SMART, suppose that a child is randomized to receive SERT as first stage treatment and that prior to week 12, say at week 10, the treating clinician is concerned that the child is worsening and insists that the child be immediately moved to the next stage of treatment. Is this an indication that the definition and timing of nonresponse should be revised prior to the full-scale SMART? Do staff members need training in how to manage these emergent clinical situations in a consistent manner? Can something be learned from this situation that will refine and improve the sequence of treatments? A SMART pilot can be used to identify when and where staff flexibility is warranted, to develop fidelity measures for its continued assessment, and to receive staff feedback about the timing of treatment switches and augmentations.

4.8. *Participant concerns about changes in treatment*

The SMART pilot can also assess whether treatment changes specified in the SMART are acceptable to participants and whether the new treatment option(s) being offered are clinically feasible. Understanding participant concerns about the treatment sequences may lead to modifications to enhance its efficacy, acceptability, and feasibility. To inform these concerns, the SMART pilot may include additional survey items, exit interviews, or focus groups with participants to better understand from their perspective what was useful about the sequence of treatments offered, the transitions, and concerns about acceptability. Questions may include: ‘How was your experience when you transitioned from a psychiatrist to a psychologist?’ ‘How was your experience when you participated in the CBT sessions after having come off your medication?’ ‘Did you find that the concerns you expressed during your sessions with the psychiatrist were also understood by the psychologist?’ ‘Was the rationale for the treatment change adequate?’ This information will aid in the execution of the full-scale trial, and also inform treatment delivery and refinement of the treatment strategies. It may also likely lead to additional measures to investigate for use as tailoring variables.

4.9. *Ethical considerations and consent procedures*

Sequential multiple assignment randomized trials studies may be more acceptable to participants than RCTs of a fixed, nonadaptive treatment. With a few exceptions when there are adverse events, in standard RCTs of fixed nonadaptive treatments there is usually no alternative treatment within the context of the trial for participants who are not responding well. In contrast, consider the SMART in Figure 1 in which, if a participant is not responding well then a second treatment is offered. Of course, guarantees cannot be made prior to a SMART, that a change or augmentation in treatment (among first-stage nonresponders) will necessarily result in improved outcome; however, in a SMART such changes or augmentations are examined (and could possibly lead to improved outcomes), whereas in fixed-treatment RCTs this is usually not an alternative. Furthermore, as in Figure 1, a SMART can involve potentially less burdensome maintenance or step-down treatment options for responding participants.

From the investigator’s point of view, participants in a SMART are randomized to a number of pre-specified ATSS (see Table I). From the study participant’s point of view, they are offered a sequence of treatments over time. A participant who is randomized to ATSS1 in the SMART in Figure 1, for example, may receive either the treatment sequence (SERT, SERT) or the treatment sequence (SERT, SERT+CBT), depending on their early non/response status. Because interventions offered to SMART participants are treatment sequences — rather than fixed nonadaptive treatments, as in most standard RCTs — this aspect of the SMART design may require consent procedures or language different from those used in a standard RCT. A SMART pilot study can be used to practice these changes in the language typically used in standard RCT consent forms.

Rerandomization does not imply that reconsent is necessary. As described in Section 4.2, the actual allocation procedure used may be one that performs the randomizations up-front or in real-time (sequentially). Regardless of the allocation procedure used, SMART participants provide consent up-front to be part of the entire study and to be assigned to one of the embedded ATSS (and therefore receive one of the embedded treatment sequences), just as in a standard RCT participants consent up-front to be part of the study and to be assigned to one of the fixed treatments. A key part of this, of course, is that

participants have a clear understanding (as part of the up-front consent procedures) of the types of treatment sequences which may be offered during the course of the SMART. Another key part of this is that investigators understand that a SMART is not a combination/packaging of separate substudies, one per randomization; rather, a SMART is itself one study with multiple randomizations.

Indeed, reconsenting SMART participants at the second stage may conflate second-stage treatment drop-out with study consent and have unintended consequences in terms of study drop-out. A SMART participant may, in fact, not find second-stage treatments acceptable/helpful and as a consequence drop-out of treatment; this does not mean the SMART participant drops out of the study. This is no different from a participant in a standard RCT who stops attending CBT sessions after week 5, for example, but continues to provide outcome assessments (that is, a participant who is a treatment drop-out but not a study drop-out). The problem with reconsenting SMART participants prior to second-stage treatment initiation is that — not only is it unnecessary, as described above — it may have the unintended consequence of encouraging study drop-out (leading to missing outcomes) among participants who would otherwise just have been treatment drop-outs (and possibly continued to provide outcome assessments). The SMART pilot study can be further used to ensure that study consent is done up-front rather than sequentially to avoid these concerns.

5. How many participants are necessary for a SMART pilot?

As discussed above, the primary aim of a pilot study is to examine the feasibility of carrying out a future larger-scale trial, rather than examining the clinical impact of a proposed set of treatments. Correspondingly, the sample size for a SMART pilot study should be based on a feasibility aim, rather than on detecting an effect size.

To ensure that the investigative team can assess feasibility is to ensure that sufficient number of participants appear in all of the subgroups. One way to accomplish this goal is to size the pilot study so that with fixed probability k , at least m participants will fall into the nonresponder subgroups B and E in Figure 1. More formally, the total sample size N can be chosen such that

$$\Pr(M_B > m \text{ and } M_E > m) > k, \tag{1}$$

where the probability is over repeated pilots of size N , and M_B and M_E are random variables denoting the number of subjects who fall into nonresponder subgroup B and E, respectively. Assuming that one-half (1/2) of the available participants are allocated to each first-stage and second-stage treatment, we have that

$$M_j = \frac{1}{2} \sum_{i=1}^{N/2} (1 - R_{ji}) \text{ for } j = B, E, \tag{2}$$

where R_{ji} is a dummy indicator which equals 1 if subject i is a responder and 0 if subject i is a nonresponder. (To ensure approximately equal numbers of participants at each randomization during the execution of both the pilot and full-scale SMART, investigators may consider permuted-block randomizations (i.e., AABB-AB-BA-BAAB-...) or more sophisticated minimization allocation procedures during the conduct of a SMART.) Note that the random variable $V_j = \sum_{i=1}^{N/2} (1 - R_{ji})$ has a binomial distribution of size $N^* = N/2$ with probability q_j , which is the true (unknown) rate of nonresponse to first-stage treatment j . For simplicity (and to be conservative, see below), we further assume $q = q_B = q_E$ (equal rates of nonresponse to either first-stage treatment option) so that display (1) is equivalent to

$$\Pr(V > 2m)^2 > k, \tag{3}$$

where V has a binomial distribution of size N^* with probability q . Note that subgroups B and C will have the same number of subjects, as will subgroups E and F, which is why it suffices to focus on subgroups B and E alone rather than all four nonresponder subgroups. Furthermore, we focus on just the nonresponder subgroups B and E because in practice — with nonresponse rates in the typical range of 35% to 65% for SMARTs — if nonresponder subgroups B and E contain at least m participants with high probability, then so will subgroups A and D, respectively; intuitively, this is because responders are not rerandomized and therefore not further split.

Three steps are required to use display (3) to determine sample size for the pilot study: First, investigators supply m , a guess at the common nonresponse rate q , and the desired k (say 80% or 90%).

Table III. The table shows sample sizes required for piloting SMART studies of the type shown in Figure 1. N is chosen such that with probability k and early nonresponse rate q , a minimum number of participants m will fall into the nonresponder subgroups B and E (and therefore C and F) in Figure 1.

		q						
		0.35	0.40	0.45	0.50	0.55	0.60	0.65
$k = 0.80$								
	$m = 2$	42	36	32	28	26	22	20
	$m = 3$	56	48	42	38	34	30	28
	$m = 4$	70	60	52	46	42	38	34
	$m = 5$	82	72	62	56	50	46	42
$k = 0.85$								
	$m = 2$	44	38	34	30	26	24	22
	$m = 3$	58	50	44	40	36	32	28
	$m = 4$	72	62	54	48	44	40	36
	$m = 5$	86	74	66	58	52	48	42
$k = 0.90$								
	$m = 2$	48	40	36	32	28	26	22
	$m = 3$	62	54	46	42	38	34	30
	$m = 4$	76	66	58	52	46	42	38
	$m = 5$	90	78	68	60	54	50	44

Second, binomial cumulative distribution functions of different sizes (in increasing order, beginning with $N^* = 1$) with probability q are evaluated (upper tail) at the quantile $2m$. This produces a list of non-decreasing probabilities. Third, choose N^* corresponding to the first (the smallest) probability larger than k , and then set $N = 2N^*$. These calculations are easily accomplished using any statistical software package. An R [45] function, which performs these calculations, is available for download on the Penn State University Methodology Center website <http://methcenter.psu.edu>.

Table III shows suggested values of N for values of $k = 0.80, 0.85, 0.90, m = 2, 3, 4, 5$, and q varying from 0.35 to 0.65. As an example, suppose that the investigative team agrees that $m = 3$ children in one of the small nonresponder subgroups is a sufficient number to ensure familiarity with the research protocol, treatment delivery, identify potential problems, and to address the concerns described in Section 4 above. Suppose the team would like to see this happen in the pilot with $k = 90\%$ probability. If the team expects a nonresponse rate of $q = 50\%$ at the end of week 12, then a pilot study for the SMART in Figure 1 would require approximately 42 participants. The key variable needed is the estimated non-response rate; this can be estimated based on existing studies in the relevant topic area.

The nonresponse rate q is unknown prior to a SMART pilot. Although the investigators may be able to find somewhat similar studies with somewhat similar treatments and participants, it is likely that the participants and the first stage treatments were not identical to those in previous studies. Thus, it may be useful to use a value of q smaller than the actual guess so that N is chosen conservatively. In addition, in the calculations above we assume the nonresponse rate q is identical for both first-line treatments. In practice, it is likely this assumption will not hold (indeed, investigators may hypothesize that one of the first-line treatments will lead to better short-term outcomes). In this case, we recommend investigators set q to the smaller of the two anticipated nonresponse rates; again this is recommended to be conservative in the choice of sample size for the pilot. Furthermore, as in any study (pilot or full), it is useful to inflate N by a guess of the study drop-out/attrition rate. For instance, in the example above, if the team expects an overall 10% study drop-out/attrition rate by the end of the study, then the pilot study sample size should be $47 = 42/(1 - 0.10)$ instead of 42.

6. Summary

Adaptive treatment strategies hold much promise in operationalizing and informing the type of adaptive, sequential treatment decisions made to address chronic conditions in ‘real-world’ clinical settings. SMARTs have been developed explicitly for the purpose of developing such ATs. A small number of SMARTs have been designed, completed or are currently being used in clinical research. Despite

this, SMARTs are still new to many researchers and some questions remain concerning how to design SMARTs appropriately. In this article, we presented a number of design considerations, unique to SMARTs, which are best addressed within the context of a small, but useful, pilot study in preparation for a full-scale SMART. To motivate and illustrate these considerations, we discussed an example SMART that addresses how to treat children and adolescents with anxiety disorder using medication, CBT, and/or the combination of both (including how to treat them following nonresponse to an initial first-line treatment). This article can be a useful guide to clinical trial investigators who are interested in planning a SMART study to develop or optimize an ATS.

Acknowledgements

Funding was provided by NIMH grants K23-MH-075843-04 (Compton), R01-MH-080015 (Murphy), a K23-MH-090216 and New York State Psychiatric Institute Research Associate Award (Gunlicks-Stoessel), and NIDA grant P50-DA-010075 (Murphy). Drs. Almirall, Compton, Gunlicks-Stoessel, Duan, and Murphy report no financial or potential conflicts of interest. The authors would like to thank John Walkup and Joel Sherrill for thoughtful comments on an earlier draft of this manuscript. In addition, we would like to thank three reviewers for comments which helped improve this manuscript significantly, including to help it reach a broader audience of clinicians and biostatisticians alike.

References

- Insel TR. Translating scientific opportunity into public health impact: a strategic plan for research on mental illness. *Archives of General Psychiatry* Feb 2009; **66**(2):128–133.
- Birmaher B. Practice parameter for the assessment and treatment of children and adolescents with depressive disorders. *Journal of the American Academy of Child and Adolescent Psychiatry* Nov 2007; **46**(11):1503–1526.
- Connolly SD, Bernstein GA, Work Group on Quality Issues. Practice parameter for the assessment and treatment of children and adolescents with anxiety disorders. *Journal of the American Academy of Child and Adolescent Psychiatry* Feb 2007; **46**(2):267–283.
- Murphy SA, Lynch KG, Oslin D, McKay JR, TenHave T. Developing adaptive treatment strategies in substance abuse. *Drug and Alcohol Dependence* May 2007; **88**:S24–S30.
- Murphy SA, Collins LM, Rush AJ. Customizing treatment to the patient: Adaptive treatment strategies. *Drug and Alcohol Dependence* May 2007; **88**:S1–S3.
- Dawson R, Lavori PW. Placebo-free designs for evaluating new mental health treatments: the use of adaptive treatment strategies. *Statistics in Medicine* Nov 2004; **23**(21):3249–3262.
- Lavori PW, Dawson R, Rush AJ. Flexible treatment strategies in chronic disease: clinical and research implications. *Biological Psychiatry* Sep 2000; **48**(6):605–614.
- Murphy SA, Almirall D. Dynamic Treatment Regimens. In *Encyclopedia of Medical Decision Making*, Kattan MW (ed.). Sage Publications: Thousand Oaks, CA, 2009.
- Lavori PW, Dawson R. Dynamic treatment regimes: practical design considerations. *Clinical Trials* Feb 2004; **1**(1):9–20.
- Murphy SA. Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society Series B* 2003; **65**:331–355.
- Hernán MA, Lanoy E, Costagliola D, Robins JM. Comparison of Dynamic Treatment Regimes via Inverse Probability Weighting. *Basic & Clinical Pharmacology & Toxicology* 2006; **98**:237–242.
- Moodie EEM, Richardson TS, Stephens DA. Demystifying optimal dynamic treatment regimes. *Biometrics* 2007; **63**:447–455.
- Murphy SA, van der Laan MJ, Robins JM, CPPRG. Marginal Mean Models for Dynamic Regimes. *Journal of the American Statistical Association* 2001; **96**:1410–1423.
- Orellana L, Rotnitzky A, Robins JM. Dynamic Regime Marginal Structural Mean Models for Estimation of Optimal Dynamic Treatment Regimes, Part I: Main Content. *The International Journal of Biostatistics* 2010; **6**(2). Article No. 8, <http://www.bepress.com/ijb/vol6/iss2/8>.
- Orellana L, Rotnitzky A, Robins JM. Dynamic Regime Marginal Structural Mean Models for Estimation of Optimal Dynamic Treatment Regimes, Part II: Proofs of Results. *The International Journal of Biostatistics* 2010; **6**(2). Article No. 9, <http://www.bepress.com/ijb/vol6/iss2/9>.
- Robins JM. Optimal structural nested models for optimal sequential decisions. In *Proceedings of the Second Seattle Symposium on Biostatistics*, Lin DY, Heagerty PJ (eds). Springer: New York, 2004.
- van der Laan MJ, Petersen ML, Joffe MM. History-Adjusted Marginal Structural Models and Statically-Optimal Dynamic Treatment Regimes. *The International Journal of Biostatistics* 2005; **1**(1). Article No. 4. <http://www.bepress.com/ijb/vol1/iss1/4>.
- Wahed AS, Tsiatis AA. Optimal estimator for the survival distribution and related quantities for treatment policies in two-stage randomization designs in clinical trials. *Biometrics* 2004; **60**:124–133.
- Bembon O, van der Laan MJ. Analyzing sequentially randomized trials based on causal effect models for realistic individualized treatment rules. *Statistics in Medicine* Aug 2008; **27**(19):3689–3716.
- Cain LE, Robins JM, Lanoy E, Logan R, Costagliola D, Hernán MA. When to Start Treatment? A Systematic Approach to the Comparison of Dynamic Regimes Using Observational Data. *The International Journal of Biostatistics* 2010; **6**(2). Article No. 18.

21. Murphy SA, Bingham D. Screening Experiments for Developing Dynamic Treatment Regimes. *Journal of the American Statistical Association* 2009; **114**:391–408.
22. Thall PF, Millikan RE, Sung HG. Evaluating multiple treatment courses in clinical trials. *Statistics in Medicine* Apr 2000; **19**(8):1011–1028.
23. Murphy SA. An experimental design for the development of adaptive treatment strategies. *Statistics in Medicine* May 2005; **24**(10):1455–1481.
24. Lavori PW, Dawson R. Improving the efficiency of estimation in randomized trials of adaptive treatment strategies. *Clinical Trials* 2007; **4**(4):297–308.
25. Lavori PW, Dawson R. Adaptive treatment strategies in chronic disease. *Annual Review of Medicine* 2008; **59**:443–453.
26. Institute of Medicine. 100 Initial Priority Topics for Comparative Effectiveness Research. <http://www.iom.edu/Reports/2009/ComparativeEffectivenessResearchPriorities.aspx>.
27. Schneider LS, Tariot PN, Lyketsos CG, Dagerman KS, Davis KL, Davis S, Hsiao JK, Jeste DV, Katz IR, Olin JT, Pollock BG, Rabins PV, Rosenheck RA, Small GW, Lebowitz B, Lieberman JA. National Institute of Mental Health Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE): Alzheimer disease trial methodology. *American Journal of Geriatric Psychiatry* Fall 2001; **9**(4):346–360.
28. Lavori PW, Rush AJ, Wisniewski SR, et al. Strengthening clinical effectiveness trials: equipoise-stratified randomization. *Biological Psychiatry* Nov 2001; **50**(10):792–801.
29. Rush AJ, Crismon ML, Kashner TM, Toprac MG, Carmody TJ, Trivedi MH, Suppes T, Miller AL, Biggs MM, Shores-Wilson K, Witte BP, Shon SP, Rago WV, Altschuler KZ, TMAP Research Group. Texas Medication Algorithm Project, phase 3 (TMAP-3): rationale and study design. *Journal of Clinical Psychiatry* Apr 2003; **64**(4):357–369.
30. Stone RM, Berg DT, George SL, Dodge RK, Paciucci PA, Schulman P, Lee EJ, Moore JO, Powell BL, Schiffer CA. Granulocyte-macrophage colony-stimulating factor after initial chemotherapy for elderly patients with primary acute myelogenous leukemia. Cancer and Leukemia Group B. *New England Journal of Medicine* Jun 1995; **332**(25):1671–1677.
31. Vogt WP. *Dictionary of statistics and methodology: a nontechnical guide for the social sciences*. Sage Publications: Newbury Park, California, 1993.
32. Collins LM, Murphy SA, Nair VN, Strecher VJ. A strategy for optimizing and evaluating behavioral interventions. *Annals of Behavioral Medicine* Aug 2005; **30**(1):65–73.
33. Kraemer HC, Mintz J, Noda A, Tinklenberg J, Yesavage JA. Caution regarding the use of pilot studies to guide power calculations for study proposals. *Archives of General Psychiatry* May 2006; **63**(5):484–489.
34. Thabane L, Ma J, Chu R, Cheng J, Ismaila A, Rios LP, Robson R, Thabane M, Giangregorio L, Goldsmith CH. A tutorial on pilot studies: the what, why and how. *BMC Medical Research Methodology* Jan 2010; **10**(1). <http://www.biomedcentral.com/1471-2288/10/1>.
35. Lancaster GA, Dodd S, Williamson PR. Design and analysis of pilot studies: recommendations for good practice. *Journal of Evaluation in Clinical Practice* May 2004; **10**(2):307–312.
36. Leon AC, Davis LL, Kraemer HC. The role and interpretation of pilot studies in clinical research. *Journal of Psychiatric Research* May 2011; **45**(5):626–629.
37. Wittes J, Brittain E. The role of internal pilot studies in increasing the efficiency of clinical trials. *Statistics in Medicine* 1990; **9**:65–72.
38. Wittes JT, Schabenberger O, Zucker DM, Brittain E, Proschan M. Internal pilot studies I: type I error rate of the naïve *t*-test. *Statistics in Medicine* 1999; **18**:3481–3491.
39. Zucker DM, Wittes JT, Schabenberger O, Brittain E. Internal pilot studies II: comparison of various procedures. *Statistics in Medicine* 1999; **18**:3493–3509.
40. Nahum-Shani I, Qian M, Almirall D, Pelham WE, Gnagy B, Fabiano G, Waxmonsky J, Yu J, Murphy SA. Q-Learning: a data analysis method for constructing adaptive interventions. *Pennsylvania State University Methodology Center Technical Report Series*. No. 10-107.
41. Berry SM, Carlin BP, Lee JJ, Muller P. *Bayesian adaptive methods for clinical trials*. CRC Press: Boca Raton, FL, 2011.
42. The Methodology Center Perspective. *Ask a Methodologist* Spring 2011; Methodology Center, Pennsylvania State University, <http://methcenter.psu.edu>.
43. Guy W. *ECDEU Assessment Manual for Psychopharmacology*. National Institute of Mental Health: Rockville, MD, 1976.
44. Kraemer HC, Stice E, Kazdin A, Offord D, Kupfer D. How do risk factors work together? Mediators, moderators, and independent, overlapping, and proxy risk factors. *American Journal of Psychiatry* Jun 2001; **158**(6):848–856.
45. R Development Core Team. *R: A language and environment for statistical computing*. Vienna, Austria. <http://www.R-project.org/>.