

Introduction to SMART designs for the development of adaptive interventions: with application to weight loss research

D Almirall,¹ I Nahum-Shani,¹ N E. Sherwood,² S A. Murphy³

Abstract

The management of many health disorders often entails a sequential, individualized approach whereby treatment is adapted and readapted over time in response to the specific needs and evolving status of the individual. Adaptive interventions provide one way to operationalize the strategies (e.g., continue, augment, switch, step-down) leading to individualized sequences of treatment. Often, a wide variety of critical questions must be answered when developing a high-quality adaptive intervention. Yet, there is often insufficient empirical evidence or theoretical basis to address these questions. The Sequential Multiple Assignment Randomized Trial (SMART)—a type of research design—was developed explicitly for the purpose of building optimal adaptive interventions by providing answers to such questions. Despite increasing popularity, SMARTs remain relatively new to intervention scientists. This manuscript provides an introduction to adaptive interventions and SMARTs. We discuss SMART design considerations, including common primary and secondary aims. For illustration, we discuss the development of an adaptive intervention for optimizing weight loss among adult individuals who are overweight.

KEYWORDS

Adaptive treatment strategies, Dynamic treatment regimens or regimes, Experimental design, Individualized or personalized behavioral interventions, Timing and sequencing of intervention components

The management of many health disorders, such as obesity, substance use, or depression, often entails a sequential, individualized approach whereby treatment (e.g., behavioral or medical interventions, or a combination) is adapted and readapted over time in response to the specific needs and evolving status of the individual [1, 2]. This type of sequential decision-making is necessary when there is high level of individual heterogeneity in response to treatment. This is often the case for many chronic disorders, conditions for which there is no widely effective treatment, or conditions for which there

Implications

Practice: Adaptive interventions provide clinical practitioners with a guide to the type of sequential, individualized decision-making that is necessary for the care or management of many health disorders.

Research: Behavioral interventions researchers who are interested in empirically developing high-quality adaptive interventions should consider Sequential Multiple Assignment Randomized Trials (SMART) as part of their methodological toolbox.

Policy: For more efficient use of health research resources, funding agencies should support the use of SMART for developing and discovering adaptive interventions prior to their evaluation in a randomized clinical trial.

are widely effective treatments but they are burdensome, costly, or carry side effects.

Due to the waxing and waning course of many chronic disorders such as obesity [3], for example, a treatment that demonstrates short-term weight loss for an individual may not lead to weight loss in the long-term even if the individual remains on the same treatment (within-person heterogeneity; [4]). In addition, not all individuals will respond or adhere to the weight loss treatment to the same degree, have the same side effect profile (e.g., in the case of weight loss medications), or respond within the same time-frame, to any given treatment or package of treatments (between-persons heterogeneity; [5, 6]). Indeed, what works for one individual, may not work (or may, in fact, be iatrogenic) for another. All of these considerations, which unfold over time, motivate an individualized sequence of treatments.

To give a concrete example for the need for individualized sequences of treatments, consider the treatment of individuals who are overweight (body mass index, $BMI \geq 25 \text{ kg/m}^2$) or obese ($BMI \geq$

¹214NU Institute for Social Research, University of Michigan, 426 Thompson Street, Ann Arbor, MI 48104-2321, USA

²HealthPartners Institute for Education and Research, Minneapolis, USA

³Department of Statistics and Institute for Social Research, University of Michigan, Ann Arbor, USA

Correspondence to: D Almirall dalmiral@umich.edu;

Cite this as: *TBM* 2014;4:260–274
doi: 10.1007/s13142-014-0265-0

30 kg/m²) with individual behavioral weight loss treatment (IBT) [7–9]. IBT is incorporated in most healthy dieting and weight loss interventions. It involves strategies such as goal setting, self-monitoring, and stimulus control; and it is often administered via weekly sessions and can be delivered on an individual basis or in a group-based setting. Despite the general success of interventions such as IBT (and others [7, 10–14]), a significant number of individuals do not lose a clinically significant amount of weight (for example, $\geq 7\%$ weight loss in a 26-week period (~6 months) [15]) or meet their targeted goals for weight loss. Further, research suggests that it may be possible to identify such individuals early, i.e., during treatment [16]. Therefore, developing new weight loss interventions that (i) *begin* with IBT, (ii) *identify* individuals showing early signs of nonresponse to IBT, and (iii) *adapt* subsequent treatment to these nonresponding individuals may be important for increasing the total number of people who lose weight.

Adaptive interventions (AI), defined and discussed more formally in Section “Adaptive Interventions” below, provide one way to guide or formalize individualized sequences of treatment. The following is an example adaptive intervention (Fig. 1a) for weight loss involving IBT and meal replacements [5] (MR), which involves, for example, the replacement of one to three meals per day via the provision of prepackaged food items from a planned menu: “Begin with five weekly sessions of IBT. At the

end of the 5th weekly IBT session, if the individual has lost ≥ 5 lbs (i.e., a responder), then continue on IBT. Otherwise, if the individual is a nonresponder, then augment IBT with Meal Replacements (IBT+MR) starting at the 6th weekly session.”

Often, a wide variety of critical questions must be answered when seeking to develop an adaptive intervention. Yet, in many settings, there is insufficient empirical evidence or theoretical basis to address these questions. Drawing on the example above, possible questions confronted by the investigators may include (a) “Should we begin with short duration IBT (5 weekly sessions) or long duration IBT (10 weekly sessions) before making a decision about subsequent treatment?”; or, put a slightly different way, “When is the optimal time (5 vs 10 weeks) to define nonresponse to IBT and offer an alternative or supplementary intervention strategy?”; (b) “For individuals who do not respond to IBT, should we augment IBT with MR or switch to another weight loss intervention?”; (c) “Should the decision concerning the initial duration of IBT be individualized based on factors known about the patient at baseline (e.g., individuals who have comorbidities, or a history of emotional eating and may need more time on IBT to achieve a clinically meaningful response during IBT)?”; and (d) “Could the decision to augment with MR vs switch be individualized based on other intermediate outcomes (e.g., early indicators of adherence to IBT, such as completion of self-

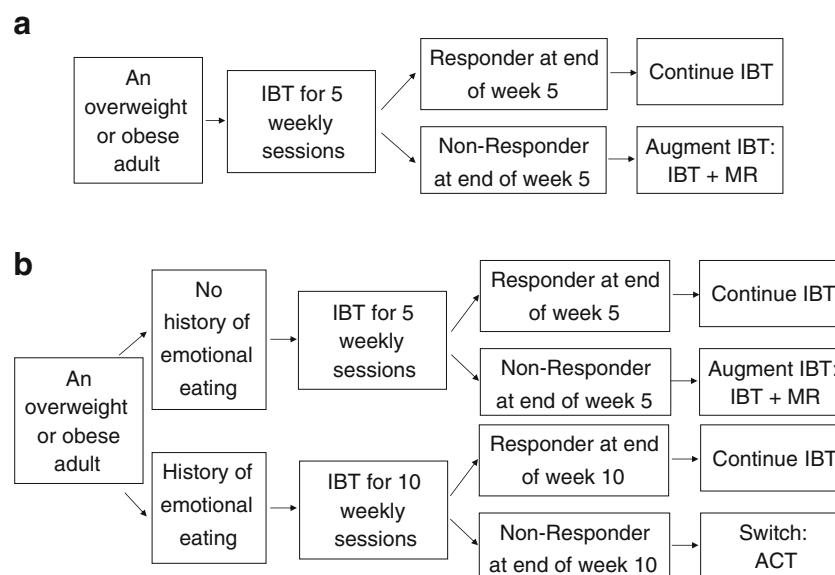


Fig 1 | a. This is one example of an adaptive intervention for weight loss. Response status is defined as losing ≥ 5 lbs by the end of the 5th weekly session on IBT. Total treatment occurs over 26 weeks. b. This is a second example of an adaptive intervention for weight loss. This example builds on the example in Fig. 1. In this example, the duration of initial IBT is individualized based on baseline information about emotional eating. *IBT* individualized behavioral treatment, *MR* meal replacements, *ACT* acceptance and commitment therapy (adapted for weight loss)

Table 1 | A description of the elements of an adaptive intervention for the two example adaptive interventions shown in Fig. 1

Figure 1a The four elements of an adaptive intervention A simple example of an adaptive intervention		Figure 1b A more deeply-tailored adaptive intervention	
Critical decision point	Type of first-stage treatment	Type of second-stage treatment	Type and duration of first-stage treatment
Treatment option(s)	IBT for 5 weekly sessions	Continued IBT or IBT+MR (augment) for 21 weekly sessions	IBT for 5 or 10 weekly sessions
Tailoring variable(s)	[None]	Change in weight from baseline to the end of 5 weeks	First-stage treatment duration (or history of emotional eating) and change in weight from baseline to the end of first-stage treatment
Decision rule	All overweight or obese adults offered IBT	If responder (change in weight ≥ 5 lbs), then continue IBT; otherwise, if nonresponder, then IBT+MR	If responder, then continue IBT; if nonresponder to IBT for 5 weeks, then IBT+MR; otherwise, if nonresponder to IBT for 10 weeks, then ACT

Responder=weight loss of 5 lbs or more and nonresponder=weight gain, no change in weight, or weight loss of less than 5 lbs
 IBT=individualized behavioral treatment, MR=meal replacements, ACT=acceptance and commitment therapy (adapted to address weight loss)

monitoring records, or failure to recommended changes in diet while on IBT)?”

The Sequential Multiple Assignment Randomized Trial (SMART)—a special type of factorial study design [17]—uses experimental design principles to obtain answers to many of the challenging critical questions around adaptive interventions, such as those raised above [18–20]. Developed explicitly for the purpose of building empirically supported adaptive interventions, they are meant to complement the use of theoretical models and expert clinical consensus for developing optimal adaptive interventions [21–23].

Despite their increasing popularity due to their real-world clinical appeal, and their fit in research aimed at developing high-quality adaptive interventions, SMARTs remain relatively new to many behavioral intervention scientists. Due to their novelty, and because SMARTs represent a significant departure from the standard two-arm randomized clinical trial (RCT), many questions remain about SMARTs and their role in the greater scientific process. For example, “What kind of questions can I answer using a SMART?” and “If I use a SMART design, does it mean I do not have to conduct an RCT?”

The purpose of this manuscript is to introduce adaptive interventions (AIs) and SMART to behavioral intervention scientists. We focus on study design concepts and how to match the study design to the types of scientific questions that might be addressed in the development and evaluation of AIs. To illustrate ideas, we discuss the development of an AI for optimizing weight loss.

ADAPTIVE INTERVENTIONS

An adaptive intervention (AI) is a sequence of decision rules that specify whether, how, when (timing), and based on which measures, to alter the dosage (duration, frequency or amount [24]), type, or delivery of treatment(s) at decision stages in the course of care. AIs are also known by a variety of different names. In the statistical literature, they are commonly referred to as dynamic treatment regimes [17, 19, 25–37] (or, more appropriately, regimens). In the mental health and substance use literatures, they are more commonly known as adaptive treatment strategies [38, 39]. Investigators often study a special type of adaptive intervention, known as a stepped-care intervention [40–44]. Stepped-care interventions begin with a low-intensity intervention that is increased if certain milestones are not achieved [45]. AIs may include stepped-up treatment (e.g., for individuals not able to lose weight) or stepped-down treatment (e.g., for individuals able to lose weight) or both. Further, in an AI, some individuals may begin with a low-intensity intervention whereas others may begin with a higher-intensity intervention, depending on their specific needs at baseline.

An AI has the following four elements: (i) decision stages, each beginning with a decision concerning treatment, and, at each stage, (ii) treatment options, (iii) tailoring variables [46], and (iv) a decision rule. Treatment options correspond to different treatment types, dosages, or delivery options, as well as various tactical treatment options (e.g., augment, switch, maintain). Tailoring variables capture information about the individual that is used in making treatment decisions. At the beginning of each decision stage, a decision rule links the tailoring variables to specific treatment options (or sets of treatment options).

There are two types of tailoring variables: baseline tailoring variables and intermediate tailoring variables. Baseline tailoring variables include information obtained prior to the first decision stage; they can be used to make treatment decisions at the first-stage or at subsequent decision stages. Intermediate tailoring variables are obtained at any time during a decision stage; they are used to make treatment decisions at subsequent stages. Intermediate tailoring variables are special types of tailoring variables in that they may lie on the causal pathway of treatment (e.g., mediators of prior treatment) or they may be early indicators for longer-term outcomes of interest (e.g., surrogates or proximal measures of the longer-term outcome).

To appreciate each element of an AI, consider the two examples shown in Fig. 1, which are also compared side-by-side in Table 1. First, we discuss the AI shown in Fig. 1a. In this example AI, short duration IBT is the only first decision stage treatment option, and no baseline tailoring variables are used. Here, the first-stage decision rule recommends offering the same treatment (i.e., short duration IBT) to all adults who are overweight or obese, regardless of their specific needs at baseline. At the second decision stage, there are two treatment options: continue IBT or augment IBT with meal replacements (IBT+MR). The second-stage decision rule recommends that responding adults (e.g., perhaps those who lose ≥ 5 lbs of their initial weight within the first 5 weeks) should remain on IBT; whereas nonresponding adults should be offered IBT+MR. Here, change in weight from baseline to the end of the 5th weekly session is the intermediate tailoring variable used to tailor the second-stage treatment. The decision rule operationalizes how to use the tailoring variable to guide the tactical decision at the second-stage. In this example, the 5 lbs cutoff is based on previous research [16], which is corroborated by clinical experience, suggesting that weight loss of less than 5 lbs during the first 5 weeks of behavioral therapy treatment is associated with longer-term insufficient weight loss (< 5 – 10 % of body weight at the end of 26 weeks), thereby indicating a need for subsequent (changes or augmentation in) treatment.

Figure 1b is an example of an AI that is more deeply-tailored than the AI in Fig. 1a (see Table 1).

By more deeply-tailored, we mean that, relative to the AI in Fig. 1a, the AI in Fig. 1b uses additional information about the individual (e.g., using additional tailoring variables) to provide more individualized treatment. First, the AI in Fig. 1b involves two first-stage treatment options: initial IBT for 5 weekly sessions vs initial IBT for 10 weekly sessions; and four second-stage treatment options: continued IBT for 16 or 21 weeks, IBT+MR for 21 weeks, or Acceptance and Commitment Therapy (ACT) for 16 weeks. ACT is a psychotherapeutic approach focusing on acceptance and mindfulness strategies for promoting health behavior change [47]. A growing body of literature suggests that ACT may be useful for managing or losing weight [48–50]. Second, the AI in Fig. 1b uses both baseline and intermediate tailoring variables. Specifically, history of binge or emotional eating at baseline is used as a baseline tailoring variable in both the first- and second-stage decision rules, and change in weight is used as an intermediate tailoring variable in the second-stage decision rule. Third, the two AIs in Fig. 1a, b differ in terms of the decision rules. Specifically, for individuals who have a history of binge or emotional eating at baseline, the first-stage decision rule recommends 10 weekly sessions on IBT; whereas, for individuals who do not have such a history, the decision rule recommends offering only 5 weekly sessions of IBT. Such baseline tailoring is based on the conjecture that to achieve a meaningful long-term reduction in weight, those who have a history of binge or emotional eating may require more time and treatment; whereas, for those who do not have such a history, it may be better to offer short duration IBT and move them more quickly to a second-stage treatment if they do not respond [51]. Concerning the second-stage decision rule, responders to 5 or 10 weekly sessions of initial IBT are continued on IBT. However, nonresponders to IBT with no history of emotional or binge eating are augmented with IBT+MR, whereas nonresponders to IBT with such a history are switched to ACT. Here, the decision to offer ACT to nonresponders with a history of emotional and binge eating is based on a conjecture that ACT may be particularly relevant for addressing this type of problem [49]. For example, ACT may be more useful than continued IBT or IBT+MR in resolving some of the problems associated with emotional or binge eating, such as feelings of ineffectiveness, perfectionist attitudes, impulsivity, low self-esteem, and poor sense of control [50].

The entire package of decision stages, treatment options, tailoring variables, and decision rules constitutes one AI; e.g., Fig. 1a constitutes a single AI and Fig. 1b constitutes a separate AI. From the perspective of an overweight individual, an AI is a sequence of treatments : e.g., an individual classified as a nonresponder at the end of the 5th weekly session is offered the sequence of 5 weekly sessions of IBT followed by 21 weekly sessions of IBT+MR.

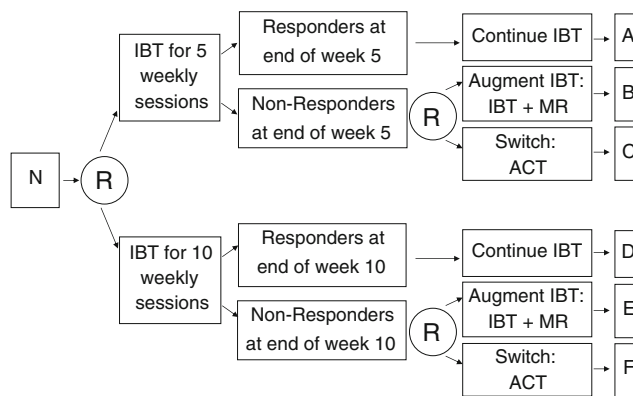


Fig 2 | This is an example sequential multiple assignment randomized trial (SMART) design for developing a weight loss intervention. Four adaptive interventions are embedded within this SMART (see Table 2). *N* sample of overweight or obese adult participants, *R* randomization, with probability 1/2, *IBT* individualized behavioral treatment, *MR* meal replacements, *ACT* acceptance and commitment therapy (adapted for weight loss)

From the perspective of the clinician/therapist, the AI is a decision rule guiding (recommending) when and how treatment should (could) be modified for both responders and nonresponders over the course of 26 weekly sessions.

SEQUENTIAL MULTIPLE ASSIGNMENT RANDOMIZED TRIALS

In many areas of research, investigators have insufficient empirical evidence or theoretical basis to assemble a high-quality adaptive intervention—that is, to choose the decision stages, treatments options, tailoring variables, and decision rules which lead to improved health outcomes. In this section, we introduce Sequential Multiple Assignment Randomized Trials (SMARTs), which are used in research to build optimal adaptive interventions. SMARTs are multistage randomized trial designs. Each participant in a SMART may move through multiple stages of treatment (i.e., first-stage treatment, second-stage treatment, etc.). Each stage corresponds to a decision stage. All SMART participants are randomized at least once, and some or all participants may be randomized more than once throughout the course of the trial. Randomizations occur at the beginning of decision stages. At each stage, randomization is used to provide data for addressing a scientific question concerning treatment options at that decision stage.

Across a wide-range of the behavioral and clinical sciences, a wide variety of SMART designs has been conducted, or is currently being conducted, to build adaptive interventions. For example, SMART has been used in oncology to develop medication algorithms to treat prostate cancer [36, 52, 53]. An example of an early precursor to the SMART is the Clinical Antipsychotics Trial of Intervention Effectiveness (CATIE) [54] in chronic schizophrenia. Lei et al. [23] describe the rationale and scientific questions addressed by four different SMART studies either recently completed or underway in

autism, attention-deficit/hyperactivity disorder (ADHD), and for various substance use disorders. The Methodology Center at Penn State University hosts a web page with the description, status, and citations to articles of various different SMART designs completed or currently underway in a variety of health settings [55].

To illustrate the key features of a SMART study design, consider the hypothetical example shown in Fig. 2. The overarching purpose of this SMART is to develop an AI beginning with IBT for optimizing weight loss among overweight adult individuals. The primary study outcome in this example SMART is longitudinal change in weight. Secondary study outcomes include physical activity, quality of life, dietary intake, and cost-effectiveness. Study outcomes are measured during monthly study (research) assessment visits from baseline/intake to the week 26 follow-up. Individuals in this example SMART participate in 26 weekly treatment sessions. Weight is also measured during each of the 26 weekly treatment sessions.

In this example SMART, all participants are randomized with equal probability to initial IBT for 5 weekly sessions vs initial IBT for 10 weekly sessions. After 5 or 10 weeks, respectively, responder (≥ 5 lbs weight loss since baseline) vs non-responder status is examined. Responders are assigned to continue on IBT, with as-needed modifications to IBT. Nonresponders are randomized again with probability 1/2 to augment IBT with MR (IBT+MR) vs switch to ACT.

As with IBT in the first-stage, there is an evidence base for IBT+MR [5, 56, 57] and ACT [58–62] as second-stage treatment options for weight loss. MR is effective for weight loss, in part, because it alleviates the need for decision-making regarding what food(s) to prepare as well as the time and effort needed to prepare the food. It provides a clear and unambiguous method for obtaining portion control and accompanying reduction in energy intake. IBT augmented with MR (IBT+MR) may, therefore,

Table 2 | This table lists the four adaptive interventions that are embedded, by design, in the example weight loss SMART shown in Fig. 2. Total time in treatment is 26 weeks for all four adaptive interventions

Embedded adaptive intervention # 1 (see figure 1a)	First-stage treatment options Short duration IBT (5 weeks)	Status at end of first-stage treatment Responder Nonresponder	Second-stage treatment options Continue IBT (21 weeks) Augment: IBT+MR (21 weeks)	Subgroups in figure 3 A+B
# 2	Short duration IBT (5 weeks)	Responder Nonresponder	Continue IBT (21 weeks) Switch: ACT (21 weeks)	A+C
# 3	Long duration IBT (10 weeks)	Responder Nonresponder	Continue IBT (16 weeks) Augment: IBT+MR (16 weeks)	D+E
# 4	Long duration IBT (10 weeks)	Responder Nonresponder	Continue IBT (16 weeks) Switch: ACT (16 weeks)	D+F

Responder=weight loss of 5 lbs or more and nonresponder=weight gain, no change in weight, or weight loss of less than 5 lbs
 IBT: individualized behavioral treatment, MR: meal replacements, ACT: acceptance and commitment therapy (adapted to address weight loss)

permit nonresponding individuals to focus their time and efforts on physical activity and experience the benefits of proper nutrition. Previous research also suggests the promise of ACT for managing or losing weight (Table 2) [58–62]. Although ACT may not produce superior weight losses on average compared to more traditional interventions, ACT may be particularly effective for certain types of individuals, specifically, those who experience higher levels of negative affect, and may, therefore, be a suitable second-stage treatment option for nonresponders.

Another rationale for considering MR and ACT as second-stage treatment options is founded in an intervention resource management perspective, with important public health significance. Due to the cost of meal replacements (MR) and the cost and availability of trained ACT psychotherapists, it would not be feasible from a public health, cost-effectiveness point of view, to provide all adult individuals who are overweight with IBT+MR or with ACT from the start. Furthermore, ACT requires a time commitment on the part of the individual, which may be burdensome. These perspectives are important since some proportion of individuals might benefit adequately in the long-term from IBT alone.

COMMON SCIENTIFIC AIMS IN A SMART

In this section, we discuss three types of scientific aims that can be addressed using data arising from a SMART (Table 3): (1) main effect aims, (2) embedded AIs aim, and (3) optimization aim. We use the example SMART in Fig. 2 to explain and give concrete examples for each type. Among these, the primary aim in a SMART is often one of the main effect aims or embedded AIs aim. The optimization aim is less confirmatory (more hypothesis generating)—it is akin to the aim of identifying moderators in a standard RCT. For this reason, the optimization aim is rarely chosen as primary; in a SMART, investigators often choose the optimization aim as a third aim. However, as with any randomized trial, ultimately the choice of primary aim is driven by scientific considerations specific to the area of study. Of course, for all of these aims, the choice of primary and secondary outcomes will differ depending on the area of application. In weight loss research, the primary outcome is often weight loss, and secondary outcomes may include quality of life or cost-effectiveness outcomes.

Main effect aims: comparison of first- or second-stage treatment options

One critical question that can be addressed with the SMART in Fig. 2 is: “What is the best initial duration of IBT for overweight adults?” Or, more specifically, “Is it better to begin adaptive interventions with a short duration IBT or long duration

Table 3 | This table lists three types of scientific aims commonly examined in a SMART, using the example SMART in Fig. 2 to illustrate

Type of aim	Example scientific questions	Contrast/analysis of interest ^b
Main effect aims	Main effect of first-stage treatment: “Is it better to begin adaptive interventions with a short duration IBT or long duration IBT?”	A+B+C versus D+E+F
	Main effect of second-stage treatment: “Among nonresponders to IBT, is it better to augment IBT with MR or to switch to ACT?”	B+E versus C+F
Embedded adaptive interventions aim ^a	Comparison of two adaptive interventions: “Is long duration IBT followed by ACT for nonresponders better than short duration IBT followed by IBT+MR for nonresponders?”	A+B versus D+F
	Identifying the best the adaptive intervention: “Which of the four embedded adaptive interventions leads to the greatest reduction in weight loss?”	Identify best among A+B, A+C, D+E and D+F
Optimization aim	To develop a more deeply-tailored adaptive intervention: “Should individuals identified as emotional/binge eaters at baseline receive longer duration IBT instead of short duration IBT?”, and “Should individuals who were nonresponders and also not adherent to initial IBT be switched to ACT?”	B+E vs C+F (for nonresponders) within levels of adherence to initial IBT, and A+B+C vs D+E+F within levels of emotional/ binge eating

Responder=Weight loss of 5 lbs or more, Nonresponder=Weight gain, no change in weight, or weight loss of less than 5 lbs

IBT individualized behavioral treatment, MR meal replacements, ACT acceptance and commitment therapy (adapted to address weight loss)

^a See Table 2 for a description of the four embedded adaptive interventions aim

^b The letters represent the different subgroups being compared in Fig. 2

IBT?” The scientific aim targeting this question concerns the comparison of the two IBT durations, averaging over the second-stage treatments for nonresponders. This comparison corresponds to the main effect of the first-stage treatment options.

A second critical question that can be addressed with the SMART in Fig. 2 is: “What is the best treatment option for nonresponders to IBT?” Or, more specifically, “Among nonresponders to IBT, should we augment IBT with meal replacements or should we switch to a psychotherapeutic approach (such as ACT)?” The scientific aim targeting this question concerns the comparison of the two treatments among nonresponders to IBT, averaging over the duration of initial IBT. This comparison corresponds to the main effect of second-stage treatments for nonresponders to IBT.

Embedded adaptive interventions aim

Both of the main effect aims above are useful for building an efficacious AI because they shed light on which treatment option (separately, at each stage) is more beneficial, on average. Here, we discuss aims that focus on how first- and second-stage treatment may work with (synergistically) or against (antagonistically) each other to impact weight loss outcomes. From this perspective, this aim involves assessing the interactive effect (as opposed to main effect) between the intervention components in the first- and second-stage.

Since the example SMART experimentally “crosses” (i.e., systematically varies) two first-stage

treatment options with two second-stage treatment options (among nonresponders), this aim involves a comparison of (2×2) four experimental groups. These groups represent the four AIs that are embedded within the SMART, by design (see Table 2). Note that the AIs themselves do not involve randomization.

The comparison of these four embedded AIs may be operationalized in various ways. One approach involves contrasting two or more of the embedded adaptive interventions. Consider a comparison of AI # 1 (initial short duration IBT; and assign IBT+MR if not responsive, continue IBT otherwise) with AI # 4 (initial long duration IBT; and assign ACT if not responsive, continue IBT otherwise) on the basis of weight loss outcomes at the follow-up at week 26. These two AIs, for example, represent the two most distinct AIs in this SMART in terms of first-stage treatment duration and second-stage treatment (i.e., an augment versus a switch): AI # 1 permits IBT a shorter amount of time to produce a response (short duration) and then recommends IBT+MR if the response is not achieved. Whereas AI # 4 permits IBT a longer period of time to achieve a response (longer duration) and then recommends ACT if the response is not achieved. That is, such a comparison permits investigators to examine whether it is better to be more stringent about achieving response during IBT and, if not responsive, augment IBT versus whether it is better to be less stringent about achieving response during IBT and, if still not responsive, give up on IBT and try an alternative.

A second way to operationalize the embedded AIs aim is to order the four AIs in terms of those with the largest average reduction in weight loss to those with the smallest reduction [63]. Since this operationalization of the aim seeks to identify the best performing AI, it differs from the first operationalization, which aims to compare two (or more) AIs using a statistical hypothesis test.

Optimization aim: the development of a more deeply-tailored adaptive intervention

This hypothesis generating aim is often considered the most interesting aim. Here, the goal is to discover and propose an AI in which the treatment is more deeply-tailored than the four AIs that are embedded in the SMART design (i.e., beyond those specified in Table 2). Practically, this aim involves identifying tailoring variables beyond those that are embedded in a SMART. This is often considered an “optimization aim” because it aims to find a more optimal sequence of treatments for each individual.

To better appreciate this, consider again the SMART study in Fig. 2. In this design, response/nonresponse status after initial IBT (i.e., based on a 5-lb change in weight from baseline to the end of first-stage treatment [16]) was used to individualize (tailor) the second-stage treatment options for all SMART study participants. Because this intermediate tailoring variable was embedded as part of the SMART study design, all AIs examined with data arising from this SMART will be a function of this variable. Indeed, the four AIs in Table 2 employ this intermediate tailoring variable as the sole tailoring variable. However, these four AIs do not (i) individualize initial IBT duration (first-stage treatment), and they (ii) do not take advantage of additional heterogeneity within the nonresponders to further individualize the choice of IBT+MR vs ACT.

The optimization aim concerns the identification of baseline tailoring variables that might be useful in making decisions about first-stage treatment (duration of first-stage IBT), as well as other intermediate tailoring variables, other than response/nonresponse based on a 5-lb change in weight, that might be useful in making decisions about second-stage treatment (IBT+MR vs ACT among nonresponders).

Concerning the investigation of baseline tailoring variables, the investigators may discover that, in the long-term, individuals who have low motivation for behavior change at baseline or a history of emotional/binge eating benefit more from receiving longer duration IBT initially, whereas those who are highly motivated or do not have a history of emotional/binge eating benefit more from shorter-duration IBT initially. Additional baseline information—such as baseline body mass index, the presence of other comorbidities, or lack of social

support at home—could also be examined. Fig. 1b, discussed earlier, provides an example AI in which baseline information is used to tailor first-stage treatment.

Concerning the investigation of other intermediate tailoring variables, the investigators might discover that different types of nonresponders to IBT would do better on augmenting with IBT+MR, whereas other types of nonresponders to IBT might do better switching to ACT. For instance, nonresponders to IBT who continue to overeat (quantity is a problem), despite eating a balanced diet and displaying the ability to make changes in physical activity (making changes in lifestyle is not a problem), may benefit more from switching to ACT rather than augmenting with MR. In other words, the individual is responding to, and complying with, the core principles IBT by initiating behavior change, yet emotional problems leading to overeating remain a barrier to weight loss. Such an individual, who now has acquired the skills useful for weight loss, may respond better to a switch to ACT. Or, as described in Fig. 1b, the choice of second-stage treatment among nonresponders may differ depending on initial treatment or baseline measures such as emotional/binge eating.

OVERVIEW OF DATA ANALYSES FOR THE COMMON SCIENTIFIC AIMS IN A SMART

In this section, we provide a brief overview of the analytic approaches associated with each of the three types of scientific aims.

Analyses associated with the two main effect aims

Standard longitudinal data analysis methods, such as linear mixed models (LMM [64]; also known as mixed-effect, random effects, hierarchical linear, or growth curve models) are used to address the two main effect aims. A key idea to understanding why standard data analysis methods are possible to address these aims is that a SMART is a form of a factorial experimental design [17]. Consistent with the analysis of factorial experiments for the development of behavioral interventions [65–67], the analyses associated with these two “main effect” aims pool together different groups of participants from the multiple subgroups A–F shown on the right margin of Fig. 2.

Specifically, the data analysis associated with the first question—“Is it better to begin with short duration IBT or long duration IBT?”—might compare change in weight loss from the beginning of the first-stage treatment to the week 26 follow-up between all individuals randomly assigned to short duration IBT (5 weeks of initial IBT; subgroups A+B+C in Fig. 2) versus all individuals randomly assigned to long duration IBT (10 weeks of initial IBT; subgroups D+E+F). This is a single factor (initial IBT duration) analysis where the factor has

two levels (short vs long duration). This analysis is identical to the analysis used to analyze data arising from a two-arm, longitudinal randomized clinical trial (RCT).

A similar data analysis approach can be used to address the second question: “Among nonresponders, is it better to augment IBT with MR or switch to ACT?” This analysis compares change in weight loss between all nonresponders randomly assigned to augmented IBT+MR (subgroups B+E) versus all nonresponders randomly assigned to switch to ACT (subgroups C+F). Again, this analysis is simply a 2-group comparison of weight trajectories.

Analysis associated with the embedded adaptive interventions aim

Again, as with the analyses of the two main effect aims above, data analyses associated with comparisons of embedded AIs pool individuals over various subgroups. For example, individuals in subgroups A+B provide data for the estimation of mean outcomes (e.g., weight loss) under AI # 1, whereas individuals in groups D+F provide data for the estimation of mean outcomes under AI # 4.

Comparing the mean outcomes between the four AIs can be done using all of the data in one regression analysis. However, unlike the standard regression analyses (e.g., LMM) associated with the main effect aims, this analysis requires a small adjustment involving weighting and replication. To appreciate why a weighted regression approach is necessary, consider that by design—i.e., nonresponders are randomized twice whereas responders are randomized only once—nonresponders would have a 1/4 chance of following the sequence of treatments they were offered, whereas responders would have a 1/2 chance of following the sequence of treatments they were offered. Therefore, in terms of second-stage treatment offerings, nonresponders are underrepresented in the data; this underrepresentation occurs by design. To account for this imbalance (i.e., to account for the underrepresentation of nonresponders), weighted regression is employed, whereby nonresponders are assigned a weight of 4 and responders are assigned a weight of 2. The weights are inversely proportional to the probability of being offered a particular treatment sequence.

Next, to appreciate why replication is necessary, consider that responders to initial short duration IBT (subgroup A; see Table 2 and Fig. 2) are consistent with AIs # 1 (A+B) and # 2 (A+C); and, similarly, responders to initial long duration IBT (subgroup D) are consistent with AIs # 3 (D+E) and # 4 (D+F). Therefore, to account for this “sharing” of responders, prior to data analysis, the outcomes (and covariates) for all responders are replicated twice.

The above provides intuition for the weighted-and-replicated regression approach for comparing the embedded AIs. This method and the adjustments necessary to provide appropriate standard errors—i.e., to account for the “double use” of responders and uncertainty in the number of individuals who are assigned a weight of 2 vs 4—are described in detail in the work by Nahum-Shani and colleagues [21]. That paper also includes example SAS code for a variety of different SMART designs. The statistical foundation for the weighting-and-replication method is found in the work by Orellana, Rotnitzky, and Robins [33, 34, 68].

Once the mean outcome and standard error is estimated for each of the embedded AIs, it is possible to conduct a hypothesis test to compare any two (or more) of the embedded AIs, or identify the AI leading to the greatest estimated weight loss [63].

Analysis associated with the optimization aim

Data analyses associated with the aim to develop a more deeply-tailored AI are akin to, but go beyond, standard moderator analysis [69–72]. In standard moderator analyses of data arising from standard randomized clinical trials (RCTs), for example, investigators often fit hypothesis generating regression models with baseline covariate-by-treatment interactions terms, such as with demographic variables (age, race/ethnicity, or sex), comorbidities (mental health or substance use), baseline weight or body-mass index, or history of treatment. These regression models often help unpack the effects of interventions because they can be used to help explain for whom treatment effects are stronger vs weaker. Such analyses are often also used with RCT data to determine whether the identified moderators are also baseline tailoring variables.

Indeed, all tailoring variables are moderator variables; whereas, not all moderators are tailoring variables. In data analysis, a tailoring variable is a special type of moderator variable in that both the magnitude *and* the direction (or presence) of the treatment effect differs at different values of the tailoring variable. (Only differences in magnitude are necessary for treatment effect moderation to be present.)

Analyses associated with the development of a more deeply-tailored AI are similar to the regressions with covariate-by-treatment interactions terms mentioned above, except with an explicit focus on exploring not only baseline but also *intermediate tailoring variables*. In the example SMART, this means identifying baseline variables that pinpoint whether shorter vs longer duration IBT is better, and baseline and intermediate variables that pinpoint whether IBT+MR vs ACT is better, among nonresponders to IBT.

“Q-Learning Regression” is a data analysis method useful for optimizing the AI while exploring candidate tailoring variables [22]. Q-Learning is

essentially a generalization of moderated regression analysis to multiple stages of treatment.

In the context of our example SMART, Q-Learning involves three steps. The first step is a regression, among early nonresponders, of weight loss from baseline to week 26 (end of study outcome) on candidate baseline tailoring variables, first-stage IBT duration, candidate intermediate tailoring variables (e.g., adherence to first-stage IBT, or within-person slope in weight change from baseline to the point of initiating second-stage treatment), a binary indicator for second-stage treatment (IBT+MR vs ACT), and covariate-by-second-stage-treatment interaction terms. The purpose of this regression is to examine candidate baseline and intermediate tailoring variables of second-stage treatment. Note that the first-stage treatment decision (5 vs 10 weeks initial treatment duration) may also serve as a tailoring variable for the decision to augment with IBT+MR vs switch to ACT.

The second step, which is based on the results of the first regression, is to assign each nonresponding individual an estimated outcome that represents their expected weight loss under the second-stage treatment (IBT+MR vs ACT) that is best for that individual. This step ensures that the evaluation of first-stage tailoring variables (the third step, below) incorporates the effect of having made the optimal future treatment decision (IBT+MR vs ACT) for each individual. Responders are assigned their observed outcome since all responders continued receiving IBT.

The third step is another regression, using all participants, of this new outcome (estimated outcome constructed in the second step above for nonresponders, observed outcome for responders) on candidate baseline tailoring variables, a binary indicator denoting whether the subject was randomized to 5 vs 10 weeks initial duration IBT, and covariate-by-first-stage-treatment interaction terms. The purpose of this final regression is to examine candidate baseline tailoring variables (e.g., emotional/binge eating or level of severity at baseline) of first-stage treatment while taking into account the fact that a future optimal decision has been made.

The end of result (e.g., the product) of this aim is a proposal for an optimal AI that may more deeply tailor treatment over time. For instance, in the context of the SMART in Fig. 2, the result of this aim may suggest the AI in Fig. 1b, or to give a second example, it may yield an AI such as the following: “Begin with weekly sessions of IBT. Individuals who have a history of emotional/binge eating should receive initial IBT for 10 weeks (long duration IBT); whereas all other individuals should receive IBT for 5 weeks initially (short duration IBT). Adherent nonresponders to short duration IBT should receive IBT+MR in the second-stage. All other nonresponders (i.e., those who are not adherent to initial IBT (regardless of duration) and those who are adherent to long duration IBT) should switch to ACT.”

As with the AI in Fig. 1b, observe how this second example AI goes beyond the four design-embedded AIs (Table 2). In this second example AI, baseline diagnosis for emotional/binge eating was identified as a tailoring variable for determining initial IBT duration. Nonresponse status, duration of initial treatment, and initial IBT adherence status are used as the three tailoring variables determining second-stage treatment (IBT+MR vs ACT). A full discussion of Q-Learning is outside of the scope of this manuscript; for details, including a worked example, see Nahum-Shani et al. [22] and associated tutorial materials on the first author’s website. In addition, an add-on procedure for SAS that implements Q-Learning [73] has been developed with an associated user’s guide [74].

SAMPLE SIZE CONSIDERATIONS

A common misconception is that SMARTs require prohibitively large sample sizes. This often stems from a belief that a large number of participants must end up in each of the final subgroups of a SMART design (i.e., subgroups A–F in Fig. 2). This belief stems from a misunderstanding of how the data is commonly analyzed. For example, it is common to think that the data arising from a SMART such as the one shown in Fig. 2 is analyzed as a 6-way analysis of covariance comparing subgroups A–F. Such an analysis may, indeed, require large sample sizes; however, as we have discussed, such an analysis does not correspond to any of the primary aims described above.

As is well known, the minimum sample size for any experimental trial is dictated by the primary aim for that trial [75]. Above, we have identified two types of aims that often serve as primary aims in a SMART: main effect aims and the embedded AIs aim. To illustrate what the sample size requirements may look like for a SMART, consider sample size calculations for the two main effect aims and the embedded AIs aim using the example SMART in Fig. 2. For simplicity, we focus on detecting medium effect sizes [76] of 0.5 (Cohen’s d) for each of these aims.

First, we consider sample size requirements assuming the main effect of first-stage treatment (the effect of initial IBT duration) is the primary aim. Since the associated analysis for this aim is a longitudinal comparison of two groups (A+B+C vs D+E+F), the sample size requirements for this effect is identical to the sample size requirement for a two-group longitudinal randomized clinical trial. Therefore, assuming a within-person correlation of 0.50 in longitudinal weight, and a study drop-out rate of 10 %, a total sample size of at least $N=122$ (i.e., 61 in group A+B+C vs 61 in group D+E+F) would be needed in order to detect a medium effect size [58] (Cohen’s $d=0.5$) with at least 85 % power and type-I error of 5 %.

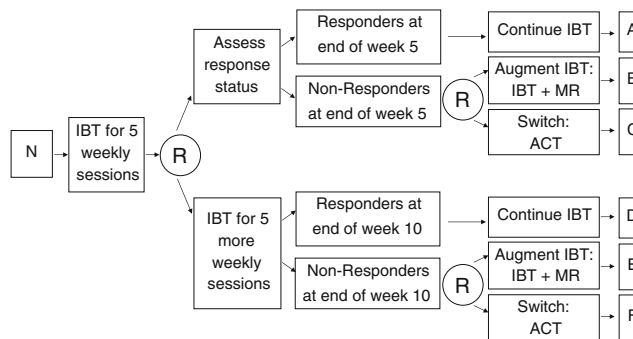


Fig 3 | This is a variation of the example SMART in Fig. 3 in which the decision to provide 5 vs 10 weeks of initial IBT occurs at the end of 5 weeks on IBT instead of at baseline. N sample of overweight or obese adult participants, R randomization, with probability 1/2, IBT individualized behavioral treatment, MR meal replacements, ACT acceptance and commitment therapy (adapted for weight loss)

Second, we consider the sample size requirement assuming the main effect of second-stage treatment (the effect of augment IBT with MR vs switch to ACT among nonresponders to IBT) is the primary aim. This analysis is also a two-group comparison (subgroups B+E vs subgroups C+F). Yet, because this analysis is among nonresponders, the sample size calculations need to take into account the rate of nonresponse to IBT as follows. For simplicity, here the rates are assumed to be equal between short vs long duration initial IBT; however, in applications where this is not an appropriate assumption, we recommend the conservative approach of choosing the smaller nonresponse rate. Assuming a nonresponse rate of 0.60 to initial IBT, and keeping all other assumptions the same as above, would require a total sample size of at least $N=204=122/0.6$; that is, $204/2=102$ in group A+B+C vs 102 in group D+E+F, which implies $102 \times 0.6=61$ in group B+E vs 61 in group C+F. Note that a nonresponse rate lower than 0.6 would result in a larger total N .

Third, we consider the sample size requirement assuming the embedded AIs aim is the primary aim, operationalized as identifying which of the four embedded adaptive interventions leads to the greatest amount of weight loss (as opposed to operationalizing it as a hypothesis test). Following Oetting et al. [63], a total sample size of at least $N=154$ is required in order to correctly identify the best embedded AI with at least 95 % probability. This calculation assumes that the best and second-best AIs differ by no smaller than a medium effect size (Cohen's $d=0.5$) and the study drop-out rate is 10 %. (A web applet for this calculation exists at the Methodology Center website [77]). An alternative, potentially less conservative approach, is to calculate the sample size to contrast two (or more) embedded adaptive interventions by simulation experiment [78].

A SMART DESIGN VARIATION

Different SMART designs are suitable for addressing different types of scientific questions [21, 23].

An interesting variation of the SMART shown in Fig. 2 centers on whether the decision to provide 5 vs 10 weeks of initial IBT occurs at some point after initial IBT (but prior to 10 weeks). The design in Fig. 2 supposes that the decision of 5 vs 10 weeks occurs at baseline; and consistent with this consideration, all participants are randomized at baseline to 5 vs 10 weeks duration of IBT. However, suppose that the decision to provide 5 vs 10 weeks of initial IBT is one that occurs (or investigators believe should occur) at the end of week 5, rather than at baseline. In this case, a variation of the SMART design shown in Fig. 2 (see Fig. 3) would provide 5 weeks of initial IBT to all participants. Then, at the end of week 5, all participants are randomized to either (1) immediately assess response/nonresponse status for purposes of deciding whether to continue IBT, augment IBT with MR, or switch to ACT or (2) provide an additional 5 weeks of initial IBT before assessing response/nonresponse status (i.e., postpone the assessment for another 5 weeks). Nonresponders (at the end of week 5 or week 10) would be re-randomized to IBT+MR vs ACT, as before.

Apart from the important practical matter of which type of duration decision makes the most sense in clinical practice, the investigator's decision to design the SMART as in Fig. 2 vs the variation in Fig. 3 also has scientific implications. Consider the following trade-off: Since, in the SMART in Fig. 2, individuals are initially aware of the assignment to 5 vs 10 weeks of initial IBT, this design permits investigators to examine interesting "expectancy effects of initial IBT duration". That is, even though the two arms are nominally identical up to week 5 in terms of the intervention being provided (IBT), it is possible that during the initial 5 weeks, participants in the two arms may differ in weight loss trajectories due to knowing how long they will remain on initial IBT. One conjecture is that participants in shorter-duration IBT, for instance, may become more motivated (than participants in longer duration) to achieve quicker weight loss knowing that their progress is going to be assessed relatively soon.

The latter design is less suitable for shedding light on such interesting expectancy effects of initial IBT duration. On the other hand, since in the latter design the decision concerning 5 vs 10 weeks of initial IBT occurs at week 5, this permits greater opportunity for developing a more deeply-tailored AI. For instance, the latter design would permit investigators to understand whether information about the progress of individuals during the first 5 weeks (in addition to baseline information) can be used to tailor the decision to assess response/nonresponse immediately (at week 5) versus postpone the assessment for another 5 weeks. This examination would be part of the optimization aim analysis. Thus, the latter design offers additional opportunities for using data to inform the decision to provide 5 vs 10 weeks of initial IBT.

DISCUSSION

For simplicity, this manuscript focused on the development of an adaptive weight loss intervention in the acute phase of treatment. However, in clinical practice, adaptive interventions may also span across the acute and maintenance phase continuum. For example, individuals who meet their target weight loss goals at the end of week 26 of acute phase treatment may be offered a personal contact intervention via telephone [79] to sustain their weight loss in the maintenance phase, whereas the intervention intensity may be increased for others who do not meet their target weight loss goals. Indeed, the ideas in this manuscript extend readily to this type of application, and within or between any phases/sequences of care.

SMART designs differ significantly from standard randomized clinical trials (RCT) in terms of their overarching aim. Whereas the overarching aim of a SMART is to *construct* a high-quality adaptive intervention based on data, the overarching aim of an RCT is to *evaluate* an already-developed intervention versus a suitable control. Examples of suitable controls may be usual care, enhanced care, a fixed (nonadaptive) intervention, etc. That is, an RCT is highly useful when investigators already have the theoretical basis and empirical evidence necessary for constructing the best/optimal adaptive intervention, and they wish to test the effectiveness of this already-developed AI by comparing it to a suitable control. For instance, Jakicic and colleagues [44] developed an 18-month stepped-care AI for weight loss, in which the “contact frequency, contact type and other strategies were modified over time depending on the achievement of weight loss goals at 3-month intervals”. In developing this AI prior to study, decisions were made (i) about the total number of decision stages, including when and how often they would occur (i.e., there was a step every 3 months over 18 months), (ii) to use one tailoring variable (i.e., weight loss), (iii) about the treatment options (i.e., group sessions, behavioral

lessons, 1 vs 2 sessions of telephone counseling, individual sessions, and meal replacements), and (iv) about the decision rules linking weight loss and treatment options at each time point (i.e., different weight loss cutoffs were decided on at different steps). After these decisions were made, the investigators evaluated this adaptive intervention versus a standard behavioral weight loss intervention using a 2-arm RCT. The aim of this study was not to construct an adaptive intervention; rather, the aim was to evaluate the above-described intervention.

However, as noted earlier, in many other cases investigators have insufficient empirical evidence or theoretical basis to form a high-quality adaptive intervention. That is, investigators often confront important open questions such as “What is the best first-stage treatment?”, “What is the best subsequent treatment?”, “What is the optimal intensity and scope for the first- or subsequent-stage interventions?”, “What is the optimal timing of a change in treatment?”, “How often should the intensity of the intervention be stepped-up or stepped-down?”, “Should adherence to initial treatment be used in addition to achieving certain weight loss goals to decide how to modify the intervention?”, or “What other measures can be used to adapt the intervention over time so as to effectively address the specific and changing needs of the individual?” SMART designs can help address these critical questions empirically, using experimental design principles, prior to evaluation.

The end result of a SMART is a proposal for an optimal AI, e.g., such as the more deeply-tailored AI in Fig. 1b. Following the development of an optimized AI using data arising from a SMART, an investigator may choose to evaluate the optimized AI versus a suitable control using a subsequent RCT.

SMARTs are part of the Multiphase Optimization Strategy (MOST; [66]), a framework for constructing effective multicomponent behavioral interventions. To appreciate this, consider that: (a) AIs, which SMARTs aim to construct, are multicomponent interventions (e.g., different components may be provided at different time points to different individuals); (b) factorial designs (along with other approaches) are often used in the optimization stage of MOST to develop more potent interventions [65, 67] and, as stated earlier, SMART is a type of factorial design [17]; and (c) using a SMART to construct the best AI prior to evaluating it in an RCT is consistent with the *resource management principle* of MOST, which states that “available research resources must be managed strategically so as to gain the most information [to] move science forward fastest” [66]. For example, concerning (c), it would be less cost-effective—both in terms of actual dollars spent and, importantly, the value of the scientific information gained (see Section “[Overview of data analyses for the common scientific aims in a SMART](#)” of [1])—to conduct one RCT to evaluate the

effect of short versus long duration initial IBT, followed by a second RCT (among nonresponders to the IBT duration which appears best on average based on the results of the first RCT) to evaluate the effect ACT vs IBT+MR, followed by a third RCT to evaluate the effect of the adaptive intervention as a whole (for both responders and nonresponders) versus a suitable control.

In terms of the actual conduct of the trial, a SMART differs from other experimental designs (such as RCTs and factorial designs) in that randomizations occur repeatedly over time. The core methodological rationale for the randomizations, however, remains the same. That is, just as randomization is aimed at permitting an unbiased comparison (i.e., free of alternative explanations due to confounding or treatment selection bias [75]) between the experimental treatment and a control conditions in an RCT or between different levels of treatment components in a factorial design, the randomizations in a SMART are aimed at permitting unbiased comparisons between treatment components (or their levels) *at each decision stage* in the development of an AI. SMARTs also differ from other experimental designs in that randomizations may be restricted based on an intermediate response/nonresponse to earlier treatment; e.g., second-stage randomization to ACT vs IBT+MR in Fig. 2 occurs only for nonresponders to IBT.

A practical issue concerns the randomized allocation of participants in a SMART. Investigators may choose to randomize participants up-front (at baseline). That is, for example, study participants might be randomized at baseline, to one of the embedded four AIs listed in Table 2. Or investigators may generate allocations in “real time” as each participant reaches a point of randomization. Both approaches permit stratified random allocation; this is used to control potential bias due to chance imbalances in treatment groups on key prognostic factors. However, the former approach only allows stratified random allocation based on baseline prognostic factors, whereas in the latter approach randomizations can make use of a wider variety of prognostic factors. For example, in Fig. 2, first-stage randomizations may be stratified based on baseline measures (including those considered as candidate baseline tailoring variables), such as baseline weight or BMI; and second-stage randomizations, among nonresponders, may be stratified on intermediate outcomes, such as adherence to IBT, or changes in weight from baseline observed prior to the second-stage. Stratifying the second-stage randomization on intermediate outcomes is not possible with up-front randomization to the embedded AIs; yet such stratification is particularly attractive when nonadherence in the first-stage (to IBT) may be prognostic of outcomes to the second-stage treatments. Other practical considerations and challenges in the design and conduct of SMART are discussed elsewhere [80].

In addition to the misconception that SMARTs require large sample sizes, another common concern about SMARTs has to do with blinded assessment of outcomes. For example, there may be concern that staff's knowledge of both initial treatment assignment and the value of the tailoring variable may lead to differential assessment (e.g., information bias) in the collection of study outcomes. As with any randomized trial, the key to avoiding this type of bias in a SMART is to make a clear distinction between the *measures used for treatment* (e.g., tailoring variables) versus the *outcomes used for research* [80]. For example, in the SMART in Fig. 2, the weight loss measures used to determine response/nonresponse at the end of the 5th and 10th weekly session is part of the definition of the embedded AIs. The therapist providing IBT, for example, may collect these measures. To avoid information bias, these measures would not be used to address the research aims. Rather, a separate set of research outcomes collected by an independent evaluator (IE; i.e., an assessor who is blind to treatment assignment) could be used to address the research aims. Some investigators may choose to conduct a SMART where blinded research outcomes are used for both treatment and research; however, since IE's are not a part of clinical practice, careful consideration would need to be given to the impact of this on the applicability of the embedded AIs in clinical practice.

Misconceptions about SMART also arise when the term “adaptive design” is used to describe SMARTs because the phrase is not sufficiently precise [81]. SMART studies are used to develop “adaptive interventions”. However, SMART studies are not necessarily “adaptive trial designs” in the sense that in adaptive trial designs, the clinical trial can be altered during the course of the trial (e.g., minimum required sample size is recalculated, or a treatment arm is abandoned) as data is gathered during the conduct of the trial [82].

Acknowledgments: The development of this article was funded by the following grants from the National Institutes of Health: P50DA010075 (Murphy, Almirall), R03MH09795401 (Almirall), RC4MH092722 (Almirall), P30DK092924 (Sherwood), and P30DK050456 (Sherwood).

1. Murphy SA, Lynch KG, Olsin D, et al. Developing adaptive treatment strategies in substance abuse research. *Drug Alcohol Depend.* 2007; 88(2): S24-S30.
2. Lavori PW, Dawson R. Adaptive treatment strategies in chronic disease. *Annu Rev Med.* 2008; 59: 443p-453p.
3. Middleton KM, Patidar SM, Perri MG. The impact of extended care on the long-term maintenance of weight loss: a systematic review and meta-analysis. *Obes Rev.* 2012; 13(6): 509-517.
4. Butryn ML, Webb V, Wadden TA. Behavioral treatment of obesity. *Psychiatric Clin North Am.* 2011; 34(4): 841-859.
5. Rock CL, Flatt SW, Sherwood NE, et al. Effect of a free prepared meal and incentivized weight loss program on weight loss and weight loss maintenance in obese and overweight women: a randomized controlled trial. *JAMA.* 2010; 304(16): 1803-1810.
6. Jeffrey RW, Drewnowski A, Epstein LH, et al. Long-term maintenance of weight loss: current status. *Health Psychol.* 2000; 19(1): 5-16.

7. Tate DF, Jeffery RW, Sherwood NE, et al. Long-term weight losses associated with prescription of higher physical activity goals. Are higher levels of physical activity protective against weight regain? *Am J Clin Nutr.* 2007; 85(4): 954-959.
8. Waleekhachonloet OA, Limwattananon C, Limwattananon S, et al. Group behavior therapy versus individual behavior therapy for healthy dieting and weight control management in overweight and obese women living in rural community. *Obes Res Clin Pract.* 2007; 1(4): 223-232.
9. Makris A, Foster GD. Dietary approaches to the treatment of obesity. *Psychiatr Clin North Am.* 2011; 34(4): 813-827.
10. Wing RR, Jeffery RW. Food provision as a strategy to promote weight loss. *Obes Res.* 2001; 9(4): 271S-275S.
11. The Diabetes Prevention Program (DPP): description of lifestyle intervention. *Diabetes Care.* 2002; 25(12): 2165-2171.
12. Ryan DH, Espeland MA, Foster GD, et al. Look AHEAD (Action for Health in Diabetes): design and methods for a clinical trial of weight loss for the prevention of cardiovascular disease in type 2 diabetes. *Control Clin Trials.* 2003; 24(5): 610-628.
13. Jeffery RW, Wing RR, Sherwood NE, et al. Physical activity and weight loss: does prescribing higher physical activity goals improve outcome? *Am J Clin Nutr.* 2003; 78(4): 669-670.
14. Levy RL, Jeffery RW, Langer SL, et al. Maintenance-tailored therapy vs. standard behavior therapy for 30-month maintenance of weight loss. *Prev Med.* 2010; 51(6): 457-459.
15. Knowler WC, Barrett-Connor E, Fowler SE, et al. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N Engl J Med.* 2002; 346: 393-403.
16. Nackers LM, Ross KM, Perri MG. The association between rate of initial weight loss and long-term success in obesity treatment: does slow and steady win the race? *Int J Behav Med.* 2010; 17(3): 161-167.
17. Murphy SA, Bingham D. Screening experiments for developing dynamic treatment regimes. *J Am Stat Assoc.* 2009; 184: 391-408.
18. Lavori PW, Dawson R. A design for testing clinical strategies: biased adaptive within-subject randomization. *J R Statist Soc A.* 2000; 163: 29-38.
19. Lavori PW, Dawson R. Dynamic treatment regimes: practical design considerations. *Clin Trials.* 2004; 1: 9-20.
20. Murphy SA. An experimental design for the development of adaptive treatment strategies. *Stat Med.* 2005; 24: 1455-1481.
21. Nahum-Shani I, Qian M, Almirall D, et al. Experimental design and primary data analysis methods for comparing adaptive interventions. *Psychol Methods.* 2012; 17(4): 457-477.
22. Nahum-Shani I, Qian M, Almirall D, et al. Q-Learning: a data analysis method for constructing adaptive interventions. *Psychol Methods.* 2013; 2012; 17(4): 478-94.
23. Lei H, Nahum-Shani I, Lynch K, et al. A "SMART" design for building individualized treatment sequences. *Annu Rev Clin Psychol.* 2012; 8: 21-48.
24. Voils CI, Chang Y, Crandell J, et al. Informing the dosing of interventions in randomized trials. *Contemp Clin Trials.* 2012; 33: 1225-1230.
25. Murphy SA. Optimal dynamic treatment regimes. *J R Stat Soc Series B.* 2003; 65: 331-355.
26. van der Laan MJ, Petersen ML. History-adjusted marginal structural models and statically-optimal dynamic treatment regimes. UC Berkeley Division of Biostatistics Working Paper Series. 2004; 158. <http://www.bepress.com/ucbbiostat/paper158>.
27. 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.
28. Hernán MA, Lanoy E, Costagliola D, et al. Comparison of dynamic treatment regimes via inverse probability weighting. *Basic Clin Pharmacol Toxicol.* 2006; 98: 237-242.
29. Moodie EEM, Richardson TS, Stephens DA. Demystifying optimal dynamic treatment regimes. *Biometrics.* 2007; 63: 447-455.
30. Bembon O, van der Laan MJ. Analyzing sequentially randomized trials based on causal effect models for realistic individualized treatment rules. *Stat Med.* 2008; 27(19): 3689-3716.
31. Murphy SA, Almirall D. Dynamic treatment regimes. In: Kattan MW, ed. *Encyclopedia of Medical Decision Making.* Thousand Oaks: Sage Publications; 2009.
32. Chakraborty B, Murphy SA, Strecher V. Inference for non-regular parameters in optimal dynamic treatment regimes. *Stat Methods Med Res.* 2010; 19(3): 317-343.
33. Orellana L, Rotnitzky A, Robins JM. Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, Part I: Main Content. *Int J Biostat.* 2010; 6(2): Article No. 8.
34. Orellana L, Rotnitzky A, Robins JM. Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, part ii: proofs of results. *Int J Biostat.* 2010; 6(2): Article No. 9.
35. 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. *Int J Biostat.* 2010; 6(2): Article No. 18.
36. Wang L, Rotnitzky A, Lin X, et al. Evaluation of viable dynamic treatment regimes in a sequentially randomized trial of advanced prostate cancer. *J Am Stat Assoc.* 2012; 107(498): 493-508.
37. Laber E, Lizotte D, Qian M, Pelham W, and Murphy SA. Statistical inference in dynamic treatment regimes. Unpublished.
38. Almirall D, Compton SN, Rynn MA, et al. SMARTer discontinuation trials: with application to the treatment of anxious youth. *J Child Adolesc Psychopharmacol.* 2012; 22(5): 364-374.
39. Dawson R, Lavori PW. Placebo-free designs for evaluating new mental health treatments: the use of adaptive treatment strategies. *Stat Med.* 2004; 23(21): 3249-3262.
40. Wilson GT, Vitousek KM, Loeb KL. Stepped care treatment for eating disorders. *J Consult Clin Psychol.* 2000; 68(4): 564-572.
41. Carels RA, Darby L, Cacciapaglia HM, et al. Applying a stepped-care approach to the treatment of obesity. *J Psychosom Res.* 2005; 59(6): 375-383.
42. Carels RA, Young KM, Coit CB, et al. The failure of therapist assistance and stepped-care to improve weight loss outcomes. *Obesity.* 2008; 16(6): 1460-1462.
43. Carels RA, Wott CB, Young KM, et al. Successful weight loss with self-help. *J Behav Med.* 2009; 32(6): 503-509.
44. Jakicic JM, Tate DF, Lang W, et al. Effect of a stepped-care intervention approach on weight loss in adults. *JAMA.* 2012; 307(24): 2617-2626.
45. McKellar J, Austin J, Moos R. Building the first step: a review of low-intensity interventions for stepped care. *Add Sci Clin Pract.* 2012; 7(1): 26.
46. Strecher VJ, Shiffman S, West R. Randomized controlled trial of a web-based computer-tailored smoking cessation program as a supplement to nicotine patch therapy. *Addiction.* 2005; 100(5): 682-688.
47. Hayes SC, Luoma J, Bond F, et al. Acceptance and commitment therapy: model, processes, and outcomes. *Behav Res Ther.* 2006; 44(1): 1-25.
48. Lillis J, Hayes SC, Bunting K, et al. Teaching acceptance and mindfulness to improve the lives of the obese: a preliminary test of a theoretical model. *Ann Behav Med.* 2009; 37: 58-69.
49. Lillis J, Hayes SC, Levin ME. Binge eating and weight control: the role of experiential avoidance. *Behav Modif.* 2011; 35(3): 252-264.
50. De Zwaan M, Mitchell J, Seim HC, et al. Eating related and general psychopathology in obese females with binge-eating disorder. *Int J Eat Disord.* 1994; 15: 43-52.
51. Forman EM, Butryn ML, Juarascio AS, et al. The mind your health project: a randomized controlled trial of an innovative behavioral treatment for obesity. *Obesity.* 2013; 21(6): 1119-1126.
52. Thall P, Logothetis C, Pagliaro L, et al. Adaptive therapy for androgen-independent prostate cancer: a randomized selection trial of four regimens. *J Natl Cancer Inst.* 2007; 99(21): 1613-1622.
53. Almirall D, Lizotte D, Murphy SA. SMART design issues and the consideration of opposing outcomes. *J Am Stat Assoc.* 2012; 107(498): 509-512.
54. Lieberman JA, Stroup TS, McEvoy JP, et al. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *N Engl J Med.* 2005; 353: 1209-1223.
55. The Methodology Center. Projects using SMART. Available at <http://methodology.psu.edu/ra/adap-treat-strat/projects>. Accessibility verified December 31, 2013.
56. Wing RR, Jeffery RW, Burton LR, et al. Food provision vs structured meal plans in the behavioral treatment of obesity. *Int J Obes Relat Metab Disord.* 1996; 20(1): 56-62.
57. Heymsfield SB, van Mierlo CA, van der Knaap HC, et al. Weight management using a meal replacement strategy: meta and pooling analysis from six studies. *Int J Obes Relat Metab Disord.* 2003; 27(5): 537-549.
58. Forman EM, Hoffman KL, McGrath KB, et al. A comparison of acceptance- and control-based strategies for coping with food cravings: an analog study. *Behav Res Ther.* 2007; 45(10): 2372-2386.
59. Lillis J, Hayes SC, Bunting K, et al. Teaching acceptance and mindfulness to improve the lives of the obese: a preliminary test of a theoretical model. *Ann Behav Med.* 2009; 37(1): 58-69.
60. Forman EM, Butryn M, Hoffman KL, et al. An open trial of an acceptance-based behavioral treatment for weight loss. *Cogn Behav Pract.* 2009; 16: 223-235.
61. Juarascio A, Shaw J, Forman E, et al. Acceptance and commitment therapy as a novel treatment for eating disorders: an initial test of efficacy and mediation. *Behav Modif.* 2013; 37(4): 459-489.

62. Forman EM, Hoffman KL, Juarascio AS, et al. Comparison of acceptance-based and standard cognitive-based coping strategies for craving sweets in overweight and obese women. *Eat Behav.* 2013; 14(1): 64-68.
63. Oetting AI, Levy JA, Weiss RD, et al. Statistical Methodology for a SMART Design in the development of adaptive treatment strategies. In: Shrout PE, Keyes KM, Ornstein K, eds. *Causality and Psychopathology: Finding the Determinants of Disorders and their Cures*. Arlington: American Psychiatric Publishing, Inc; 2011: 179-205.
64. Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal Data*. New York: Springer; 2001.
65. Chakraborty B, Collins LM, Strecher V, et al. Developing multicomponent interventions using fractional factorial designs. *Stat Med.* 2009; 28: 2687-2708.
66. Collins LM, Baker TB, Mermelstein RJ, et al. The multiphase optimization strategy for engineering effective tobacco use interventions. *Ann Behav Med.* 2011; 41: 208-226.
67. Collins LM, Chakraborty B, Murphy SA, et al. Comparison of a phased experimental approach and a single randomized clinical trial for developing multicomponent behavioral interventions. *Clin Trials.* 2009; 6(1): 5-15.
68. Robins JM, Orellana L, Rotnitzky A. Estimation and extrapolation in optimal treatment and testing strategies. *Stat Med.* 2008; 27: 4678-4721.
69. Kraemer H, Stice E, Kazdin A, et al. How do risk factors work together? Mediators, moderators, and independent, overlapping, and proxy risk factors. *Am J Psychiatry.* 2001; 158(6): 848-856.
70. Kraemer H, Wilson G, Fairburn C. Mediators and moderators of treatment effects in randomized clinical trials. *Arch Gen Psychiatry.* 2002; 59: 877-883.
71. Kraemer H, Frank E, Kupfer D. Moderators of treatment outcomes. *JAMA.* 2006; 296(10): 1286-1288.
72. MacKinnon DP, Luecken LJ. How and for whom? Mediation and moderation in health psychology. *Health Psychol.* 2008; 27(2): S99.
73. PROC QLEARN [Software] (2012). The Methodology Center, Penn State University. Available at <http://methodology.psu.edu/downloads/procqlearn>. Accessibility verified December 31, 2013.
74. Ertefaie A, Almirall D, Huang L, Dziak JJ, Wagner AT, Murphy SA. SAS PROC QLEARN user's guide. University Park: The Methodology Center, Penn State University.
75. Piantadosi S. *Clinical Trials: A Methodologic Perspective*. 2nd ed. Hoboken: Wiley; 2005.
76. Cohen J. *Statistical Power Analysis For The Behavioral Sciences*. 2nd ed. Hillsdale: Erlbaum; 1988.
77. The Methodology Center. Projects using SMART sample size calculator for continuous outcomes. Available at <http://methodology.psu.edu/downloads>. Accessibility verified December 31, 2013.
78. Buyze J, Rompaye BV, Goetghebeur E. Designing a sequentially randomized study with adherence enhancing interventions for diabetic patients. *Stat Med.* 2010; 29: 1114-1126.
79. Svetkey LP, Stevens VJ, Brantley PJ, et al. Comparison of strategies for sustaining weight loss: the weight loss maintenance randomized controlled trial. *JAMA.* 2008; 299: 1139-1148.
80. Almirall D, Compton SN, Gunlicks-Stoessel M, et al. Preparing for a sequential multiple assignment randomized trial for developing an adaptive treatment strategy: designing a SMART pilot study. *Stat Med.* 2012; 31(17): 1887-1902.
81. The Methodology Center. are adaptive interventions Bayesian? Available at <http://methodology.psu.edu/eresources/ask/sp11>. Accessibility verified December 31, 2013.
82. Berry SM, Carlin BP, Lee JJ, et al. *Bayesian adaptive methods for clinical trials*. Boca Raton: CRC Press; 2011.