

The Multiphase Optimization Strategy for Engineering Effective Tobacco Use Interventions

Linda M. Collins, Ph.D. · Timothy B. Baker, Ph.D. · Robin J. Mermelstein, Ph.D. · Megan E. Piper, Ph.D. · Douglas E. Jorenby, Ph.D. · Stevens S. Smith, Ph.D. · Bruce A. Christiansen, Ph.D. · Tanya R. Schlam, Ph.D. · Jessica W. Cook, Ph.D. · Michael C. Fiore, MD, MPH, MBA

Published online: 4 December 2010
© The Society of Behavioral Medicine 2010

Abstract The multiphase optimization strategy (MOST) is a new methodological approach for building, optimizing, and evaluating multicomponent interventions. Conceptually rooted in engineering, MOST emphasizes efficiency and careful management of resources to move intervention science forward steadily and incrementally. MOST can be used to guide the evaluation of research evidence, develop an optimal intervention (the best set of intervention components), and enhance the translation of research findings, particularly type II translation. This article uses an ongoing study to illustrate the application of MOST in the evaluation of diverse intervention components derived from the phase-based framework reviewed in the companion article by Baker et al. (Ann Behav Med, in press, 2011). The article also discusses considerations, challenges, and potential benefits associated with using MOST and similar principled approaches to improving intervention efficacy, effectiveness, and cost-effectiveness. The applicability of this methodology may extend beyond smoking cessation to the development of behavioral interventions for other chronic health challenges.

Keywords Smoking cessation · Behavioral interventions · Phase-based framework · Multiphase optimization strategy · MOST

L. M. Collins
The Pennsylvania State University,
University Park, PA, USA

T. B. Baker · M. E. Piper (✉) · D. E. Jorenby · S. S. Smith ·
B. A. Christiansen · T. R. Schlam · J. W. Cook · M. C. Fiore
University of Wisconsin,
Madison, WI, USA
e-mail: mep@ctri.medicine.wisc.edu

R. J. Mermelstein
University of Illinois, Chicago,
Chicago, IL, USA

Introduction

Despite the need for tobacco cessation approaches that work, the development of novel cessation interventions of demonstrated effectiveness and cost-effectiveness has been relatively slow. If all tobacco users could be provided with evidence-based treatments that substantially increased their chances of becoming tobacco-free, not only would morbidity and mortality be reduced but also millions of dollars in health-related costs could be saved. This article, in conjunction with Baker et al. [1], presents a new approach for improving smoking cessation treatments. Baker et al. [1] describe a phase-based framework for tobacco use cessation research. The goal of the present article is to present a methodology that will enable researchers to be efficient and programmatic in their development of novel, effective, and cost-effective smoking cessation interventions and to describe an ongoing application of this methodology to a smoking cessation intervention based on the phase-based framework. The applicability of this methodology may extend beyond smoking cessation to the development of behavioral interventions for other chronic health challenges, such as prevention and treatment of addiction to alcohol and other drugs, prevention of obesity, promotion and maintenance of weight loss, and management of disorders such as diabetes, asthma, HIV/AIDS, and cancer.

The Phase-Based Framework and Multicomponent Interventions for Smoking Cessation

Briefly, the Baker et al. framework partitions the cessation process into four discrete phases based on current theories of cessation and empirical data: (a) motivation, (b) precessation, (c) cessation, and (d) maintenance. The phase-based framework can be integrated with models that support the use of

particular interventions to address phase-specific challenges to cessation success. For instance, associative learning models, pharmacologic and motivational models of dependence, or social learning models can be incorporated with data on dependence and the process of smoking cessation to identify challenges, opportunities, or intervention targets for the four phases. Then these same models and data can be integrated with prior treatment evaluation research to generate intervention components suitable for the four different phases of cessation. Thus, the phase-based framework provides an organizational structure that helps integrate data and more molecular theories to guide intervention strategies during each of the four phases of the cessation process.

Implicit in the phase-based framework is the idea that it may be necessary to include multiple treatment components both within and across the four phases of cessation. Multicomponent interventions that combine various behavioral and pharmacotherapy treatments are common in smoking cessation (e.g., [2, 3]). Behavioral interventions include techniques such as skill training, problem solving, support provision, motivational interventions, and education. Multicomponent interventions may also include pharmaceutical components such as nicotine replacement, bupropion, or varenicline.

Multicomponent cessation interventions are typically subjected as a package to a program evaluation [3] that seeks to address the question: “Does this intervention have a statistically significant effect?” In this article, we propose that although this is a critically important question, there is another, equally important question: “Is this intervention achieving its maximum public health impact?” Depending on the situation, the term “public health impact” may be assessed in terms of efficacy, effectiveness, cost-effectiveness, or some other relevant criterion. Statistical significance and maximum public health impact are distinct and independent concepts, as Table 1 shows. If an intervention does not achieve a statistically significant effect and has not achieved its maximum public health impact, it is possible that improvements could result in an effect that is large enough to reach statistical significance. However, if the intervention is already yielding its maximum public health impact and this effect is not large enough to achieve statistical significance, a different

approach is warranted. Similarly, a statistically significant behavioral intervention may or may not have achieved its maximum public health impact. In our view, it is likely that the vast majority of behavioral interventions that achieve statistically significant program effects could readily be modified to increase their public health impact appreciably. We suggest that behavioral intervention development can and should be aimed at both statistically significant program effects and maximum public health impact.

The Multiphase Optimization Strategy

Achievement of both of these outcomes is the goal of the methodological framework described in this article, the multiphase optimization strategy (MOST) [4, 5]. Based on ideas drawn from engineering and related fields and adapted to behavioral intervention science, MOST is a principled and comprehensive framework for optimizing and evaluating behavioral interventions. In engineering, optimization is a goal that sometimes can be unambiguously attained; in intervention science, optimization is more likely to be a conceptual standard to work toward. Thus, in intervention science, optimization can be considered a process rather than an end goal per se.

Optimization of behavioral interventions is relevant to both type I and type II translational research. Type I translational research involves moving basic science (e.g., laboratory) findings to the clinical context and is commonly referred to as “bench-to-bedside” research. Type II translational research involves moving from more controlled clinical trials to real-world community settings. We are particularly interested in type II translation because of its importance in tobacco control. Translating cessation intervention research into healthcare settings and other real-world health care contexts (e.g., quit lines) is crucial because such settings offer unparalleled opportunities to deliver cessation treatments to large numbers of smokers. Despite its importance, such translation research is relatively rare [6–8], perhaps because using current approaches translation can be arduous and slow, sometimes taking 15 to 20 years [9, 10]. We see MOST as a means of making type II translational research more efficient and, we hope, more

Table 1 Two critical questions about behavioral interventions

Is this intervention achieving its maximum public health impact?	Does this intervention have a statistically significant effect?	
	No	Yes
No	Improvements to improve public health impact possible, may lead to statistically significant effect	Improvements to improve public health impact possible
Yes	Different intervention approach called for	Objective of multiphase optimization strategy

common. This will ultimately facilitate the broad implementation of promising cessation strategies.

The Present Article

This article has several objectives. The first is to present an updated description of MOST that is more detailed than those in earlier articles [4, 5]. The second is to illustrate MOST by describing an application from our work. The third is to show how MOST can be used in translational research, particularly type II translational research. The fourth is to review some considerations, challenges, and potential benefits associated with using MOST and similar principled approaches to improving intervention efficacy, effectiveness, and cost-effectiveness.

In this article, we describe research we are currently conducting using MOST to develop a multicomponent smoking cessation intervention. The research is aimed at optimization of intervention effectiveness and takes place in healthcare settings. The intervention components under consideration stem directly from the phase-based framework reviewed in the companion article by Baker et al. [1] and constitute a diverse set representing both counseling and pharmacotherapy.

An Approach to Intervention Development and Optimization Inspired by Engineering: the Multiphase Optimization Strategy

Why Should Intervention Science Adopt Ideas from Engineering?

Because the methods used in intervention science today have been developed for purposes of intervention evaluation rather than optimization, smoking cessation intervention scientists currently do not have clear methodological guidelines about how to conduct optimization research. We suggest turning to research methods developed for fields such as engineering, manufacturing, and product development (e.g., [11, 12]) to provide a blueprint for efficient and programmatic smoking cessation research. At first glance it may seem that engineering has little in common with clinical intervention development, but reflection reveals several parallels. In manufacturing and product development, theory often suggests many different component parts or materials that could be used, which could then be combined to form many alternative “draft” versions of the product, or prototypes, that could be built and tested. Because building and testing each of these alternative prototypes would be expensive in terms of personnel and materials and could slow progress, engineering methodologists have developed a more efficient strategy. The

strategy is based on the premise that out of the large array of potential prototypes, there is a much smaller number that are the best and most promising. For this reason, rather than building and testing all possible prototypes, engineers typically first conduct experiments to search efficiently and systematically through the various components (and versions of those components) that are candidates to be included in the product, with the objective of identifying the most promising set without having to assemble and test each prototype. After the search for promising components has been completed in an initial phase and fine-tuning has taken place if necessary, in subsequent phases a prototype is built out of the components that have been identified as promising. This prototype is then subjected to a full test. Taking a programmatic and sequenced experimental approach enables product development to make faster progress while husbanding research resources.

In much the same way, the phase-based framework of tobacco cessation, along with various tobacco dependence theories, suggests many different intervention components. For example, different intervention components may be intended to motivate a smoker to attempt to quit, increase initial abstinence, increase cessation program participation, increase patient adherence, prolong abstinence, or promote re-quitting. The many intervention components suggested by theory could be combined to form hundreds or even thousands of plausible alternative smoking cessation intervention packages. We propose that by adopting a programmatic and sequenced experimental approach similar to that used in engineering, it is possible to search efficiently and systematically to identify the most promising smoking cessation intervention components and levels of components, assemble these components and levels into an optimal treatment package, and then subject only the optimal treatment package to a full efficacy or effectiveness trial. We believe that just as it has hastened product development in other fields, a phased experimental approach will hasten progress in building better smoking cessation interventions, while making the most of research resources.

Two Basic Principles from Engineering

Based on our examination of engineering methods, we have identified two basic principles that translate readily to methodology for intervention science. According to the first, the *resource management* principle, available research resources must be managed strategically so as to gain the most information and the most reliable information and thereby move science forward fastest. The resource management principle has implications for how research should be conducted. In engineering, research designs are chosen by prioritizing which information is most important to gain and then targeting resources accordingly [12]. There is an

emphasis on randomized experimentation because such designs produce the most reliable scientific information, usually most efficiently (e.g., [13]). According to the resource management principle, an investigator using MOST should seek an experimental design that addresses the highest-priority research questions in the most efficient way. The choice of a specific experimental design depends upon the research questions to be investigated.

The second principle, the *continuous optimization* principle, states that a new cycle of research should begin as soon as the previous round of development research is concluded, in order to build on previous work and make further incremental improvements. In the manufacturing and product development fields, there is rarely an expectation that once a product has been developed and is ready to be marketed, the job is done. Instead, research generally begins anew, devoted to adding or refining features so that the new product does more than its predecessor; to developing a product that does what its predecessor did but more effectively, efficiently, conveniently, or cheaply; or to responding to changes in the environment or in the needs, desires, and preferences of the customer base. The new cycle is informed both by the research conducted as part of previous cycles and by hypothesis-generating secondary data analyses.

Building on Engineering Principles: an Overview of MOST

MOST [4, 5], which has some conceptual roots in the phased approach to intervention development and evaluation proposed by the UK's Medical Research Council [14, 15] applies the resource management and continuous optimization principles to improve the efficiency of smoking cessation research and the effectiveness and cost-effectiveness of smoking cessation interventions. MOST is a framework for building and optimizing multicomponent behavioral interventions. We wish to stress that by “framework” we mean a general approach rather than an off-the-shelf procedure. How MOST is conducted in any specific application may vary greatly depending upon the motivating research questions, public health area, available resources, and other aspects of the situation. In this article, we offer an expansion, clarification, and elaboration of MOST.

Figure 1 is a flowchart outlining MOST. As the figure shows, MOST consists of a sequence of steps aimed at the systematic optimization of a multicomponent intervention. The sequence of steps begins with a theoretical model, which informs the identification and selection of intervention components to be examined. Next, the intervention components are examined via randomized experimentation. This experimentation is aimed at gathering information that will be useful in making decisions about each of the intervention components. For example, a decision may concern whether to include a particular component in an intervention or which

intensity level of the component should be used (e.g., one vs. five sessions of counseling). In the next step, additional experimental or nonexperimental work may be done for the purpose of refinement and fine-tuning. As is consistent with the resource management principle, the designs of any randomized experiments are carefully chosen with a high priority on efficiency and economy. The purpose is to gather information that will be used to make decisions about which intervention components and component levels will be included in the intervention. Once the information is gathered, it is used to assemble a beta (i.e., draft) version of the intervention. The efficacy or effectiveness of this draft intervention is then confirmed in an evaluation by means of a standard randomized controlled trial (RCT). If efficacy/effectiveness is confirmed, the intervention is released. Consistent with the continuous optimization principle, a new cycle of MOST, aimed at further improvements, would begin immediately.

In the next section, we walk through the steps of MOST illustrated in Fig. 1, using as an example our research applying the MOST framework to optimize the effectiveness of a clinic-based smoking cessation intervention. We hope that this example may motivate some readers to consider how MOST could be applied in their intervention research.

An Illustration of the Application of MOST to Optimize a Smoking Cessation Intervention

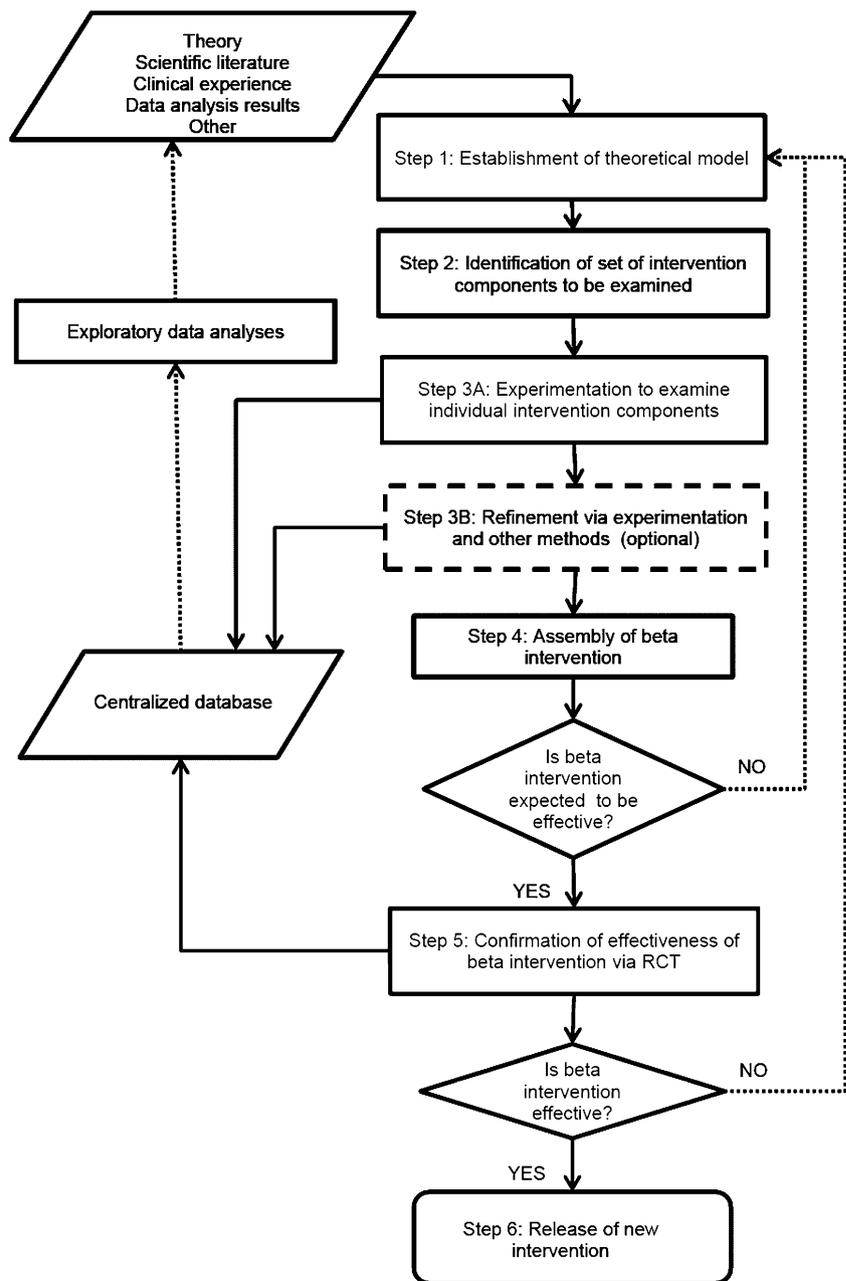
Step 1: Establishment of Theoretical Model

A clearly articulated theoretical model is the starting point for MOST. As Fig. 1 shows, this theoretical model may be derived from theory, scientific literature, clinical experience, exploratory data analyses, or any other relevant information. In our work, this theoretical model is based on the phase-based framework described in detail in the companion article by Baker et al. [1], as well as various tobacco dependence theories.

Step 2: Identification of the Set of Intervention Components to Be Examined

Our theoretical model suggests multiple intervention components and levels that could be included in a smoking cessation treatment package. We selected six intervention components to examine experimentally. Three of these components pertain to the precessation phase, two pertain to the cessation phase, and one pertains to the maintenance phase. For each of these components, we ultimately need to make a decision either about whether to include the component in the intervention or about what level of the component is best. It is important

Fig. 1 Flowchart of the multi-phase optimization strategy. A *rectangle* represents action. A *rhombus* represents information that may be input and/or output. A *diamond* represents a decision. The *round-cornered rectangle* represents the product, in this case the new intervention to be released



to recognize that although these intervention components have been analyzed previously in clinical trials, the research described here is intended to produce unique information on these components, i.e., estimates of individual component effects in a real-world effectiveness setting. Prior research does not directly supply this information.

The specific decisions to be made, research questions, intervention components, and levels of the intervention components we selected for examination are:

1. Decision: Use of the nicotine patch is common during the cessation phase, but research [16] indicates that the

nicotine patch may also be helpful during precessation. We need to decide whether our intervention should include use of the nicotine patch during precessation. Research question: *Does precessation use of the nicotine patch improve initial cessation outcomes relative to no precessation use of the nicotine patch?* Intervention component: precessation nicotine patch. Levels: patch, no patch.

2. Decision: Research also indicates that use of oral nicotine replacement therapy (NRT) self-administered ad lib (as needed) may be helpful during precessation [17]. We need to decide whether our intervention should include use of oral NRT during the precessation

- phase. Research question: *Does precessation use of oral NRT improve initial cessation outcomes relative to no precessation use of oral NRT?* Intervention component: precessation ad lib oral NRT. Levels: oral NRT, no oral NRT.
3. Decision: Research indicates that counseling addressing issues such as smoking reduction, withdrawal coping, environmental restrictions on smoking, intra-treatment social support, autonomous motivation, and practice quit attempts may be helpful during the precessation phase [3, 18]. We need to decide whether to include this counseling in our intervention. Research question: *Does precessation counseling improve initial cessation outcomes relative to no precessation counseling?* Intervention component: precessation counseling. Levels: intensive, none.
 4. Decision: Counseling during the cessation phase, emphasizing elements such as intra-treatment social support and skill-based competence, is efficacious [3], but it is not known how much counseling is cost-effective. Given the expense of counseling, this is an important question. We need to decide whether it is justified to include more intensive counseling, operationally defined as three 20-min sessions, or whether a minimal level of counseling, consisting of one 3-min session, is sufficient. The minimal in-person counseling level is based on the 2008 Public Health Service Guideline recommendations for brief clinician counseling [3] and resembles what happens in many real-world clinic settings. Research question: *Does intensive in-person counseling during the cessation phase improve initial cessation outcomes relative to minimal in-person counseling?* Intervention component: cessation in-person counseling. Levels: intensive, minimal.
 5. Decision: It is very efficient to deliver counseling over the telephone. These phone sessions are similar to the cessation interventions typically offered via a tobacco cessation quitline, such as those offered by Free and Clear [19–22]. We need to decide whether it is justified to include intensive phone counseling during the cessation phase, operationally defined as three 15-min sessions, or whether a minimal amount of phone counseling, consisting of one 10-min session, is sufficient. Research question: *Does intensive counseling during the cessation phase delivered over the telephone improve initial cessation outcomes relative to minimal counseling delivered over the telephone?* Intervention component: cessation phone counseling. Levels: intensive, minimal.
 6. Decision: There is some evidence that long-term medication use during the maintenance phase may be more effective than brief use [23–27], but the evidence is mixed [28]. We anticipate that extending medication to 16 weeks

will be particularly effective at enhancing long-term cessation outcomes because extended medication directly addresses major threats to long-term success (e.g., progression from lapse to relapse, anhedonia [29–32]) identified via the phase-based framework. We need to decide whether our intervention should include extended medication duration. Research question: *Does an extended duration of 16 weeks of medication use during the maintenance phase improve long-term cessation outcomes more than the standard 8-week duration of medication use?* Intervention component: maintenance medication duration. Levels: 16 weeks, 8 weeks.

Step 3A: Experimentation to Examine Individual Intervention Components

Overview

The traditional approach of choosing intervention components and levels is based solely on information such as clinical experience, informal hypotheses, and post hoc nonexperimental analyses. This information directly informs assembly of a treatment package. In contrast, MOST calls for using this kind of information to inform randomized experimentation designed to provide definitive answers about the performance of each component. The experimental results then directly inform assembly of a treatment package. The purpose of this step of MOST is to conduct the experimentation.

In our research, in this step we are conducting randomized experimentation to gather evidence about the effectiveness of the six individual intervention components described above. Ultimately this evidence will form the basis for decisions about which components and component levels will be included when the beta version of the smoking cessation treatment package is assembled in step 4.

Because selection of an experimental design for use in step 3A is critically important, we will describe the considerations that went into our choice. In general, the experimental design chosen should possess two signature characteristics. First, it must separate component effects, enabling estimation of the contribution of each component. Second, it must husband time and research resources—in other words, it must be highly efficient.

We estimated the sample size that would be needed to achieve sufficient power for a comparison of the two levels of each of the components listed in Table 1. The smallest effect size could be detected with power ≥ 0.9 with $N=512$ (this allows for 5% attrition), using number of days abstinent during the 2-week post-quit period as a short-term outcome (the companion article by Baker et al. [1] discusses the need to choose phase-appropriate outcome

variables. For this research, these are short-term outcome variables that focus on cessation phase events). We then considered several design alternatives that would enable us to maintain this level of power, computing the sample size requirements using the formulas that appear in Collins et al. [33]. The alternatives considered were individual experiments, a comparative treatment design, a complete factorial design, and a fractional factorial design. We review these alternatives below; some key similarities and differences among the designs considered are summarized in Table 2.

Individual Experiments and Comparative Treatment Designs

The individual experiment approach involves conducting separate two-condition experiments. In this example, there would be six experiments, one corresponding to each of the intervention components. In each experiment, the treatment would consist of one intervention component set to *on*, with the remaining five intervention components set to *off*. To maintain power, each experiment would require $N=512$, for a total required sample size of $N=3,072$ across the six experiments. In the comparative treatment design, there would be a single control group and six treatment groups. In each of the treatment groups, one of the intervention components would be set to *on*, and the remaining five would be set to *off*. This design is essentially the same as conducting six separate experiments, except that six treatment groups share a single control group. Thus, the comparative treatment approach requires only seven experimental conditions, as opposed to a total of 12 conditions required by the individual experiments approach. To maintain the same level of power as conducting six separate experiments, the comparative treatment experiment would require $N=1,792$ subjects. A key limitation to both the individual experiments and comparative treatment approaches is that interactions among treatment components cannot be estimated, even though these interactions remain present [33] and potentially have an impact on the results.

Factorial Designs

In a complete factorial experiment [34], the levels of two or more independent variables are systematically varied, or “crossed”, so that all possible combinations of levels of the factors are implemented. The classic and most familiar example is the 2×2 factorial design, in which there are two factors, each with two levels, resulting in four experimental conditions (more about the use of factorial experiments in intervention science may be found in Collins et al. [33] and Chakraborty [35]). Although currently factorial experiments with more than two factors are rarely conducted in intervention science, they have some attractive properties and merit consideration alongside other alternatives. One attractive property is their efficient use of experimental subjects. Table 2 shows that a factorial experiment examining the six intervention components of interest here would require only $N=512$ subjects, which is less than 20% of what is required by the individual experiments approach and less than 30% of what is required by the comparative treatment approach. Another attractive property of factorial experiments is that they enable estimation of interactions between factors.

These attractive properties led us to consider conducting a factorial experiment to examine the effects of the six intervention components. However, the efficiency of factorial experiments is a two-edged sword: on the one hand, factorial experiments permit tests of multiple factors; on the other hand, as the number of factors increases, the number of experimental conditions required increases rapidly. Table 2 shows that a factorial experiment with six two-level factors requires 64 experimental conditions (i.e., unique combinations of factors for a group of subjects). This large number of different experimental conditions presents a significant implementation challenge. At this point, we wondered whether we could find an experimental design that would enable us to examine all six intervention components using $N=512$ but would require implementation of fewer than 64 experimental conditions. This led us to consider a fractional factorial design.

Table 2 Comparison of features of design alternatives

Design	Number of subjects needed to maintain power ≥ 0.9	Number of experimental conditions	Interactions
Individual experiments	3,072	12	None can be estimated
Comparative treatment	1,792	7	None can be estimated
Complete factorial	512	64	All can be estimated
Fractional factorial	512	8–32 depending on design chosen	Selected subset can be estimated

Fractional Factorial Designs

In fractional factorial designs, only a carefully selected fraction of experimental conditions are implemented. These designs have the same statistical power as a complete factorial experiment and therefore require the same number of subjects. However, they require fewer experimental conditions and thus are more economical and easier to implement. Table 2 shows that depending on the design chosen, a fractional factorial experiment with six factors requires anywhere from eight to 32 experimental conditions. The trade-off presented by fractional factorial designs is that certain effects cannot be disentangled from other effects. This phenomenon, known as aliasing [12], is an inevitable result of removing conditions from a complete factorial experiment (aliasing is also present in individual experiments and comparative treatment designs; see Collins et al. [33]). Only complete factorial experiments are free of aliasing). In general, the more economy that is afforded by a fractional factorial design, the more aliasing of effects is required. In other words, when six factors are to be investigated, a fractional factorial design with eight experimental conditions will involve more aliasing than a corresponding fractional factorial design with 32 experimental conditions.

Because any fractional factorial design's aliasing can be readily determined, the investigator can choose strategically from among the wide variety of available fractional factorial designs. The first step in considering a fractional factorial design is to sort the effects in the ANOVA into three categories: effects that are the primary targets of the experiment (category A), effects that are not the primary targets of the experiment and are likely to be negligible in size (category B), and effects that are not the primary target of the experiment but are potentially sizeable (category C). It is given in every fractional factorial experiment that all the main effects are primary targets of the experiment and are therefore in category A. Usually, selected two-factor interactions and, for some studies, a three-factor interaction are placed in category A as well. In general, using fractional factorial designs requires making the assumption that most higher-order interactions (i.e., interactions that involve three or more factors) are in category B.

If a target effect is aliased with one or more other effects that are negligible, the resulting estimate will be primarily attributable to the target effect. It follows that the general strategy in selecting a fractional factorial design is to find a design in which the effects that are the primary targets of the experiment (category A) are aliased only with effects that are not targets of the experiment and are likely to be negligible (category B) and are *not* aliased with any possibly sizeable effects (from either category A or category C).

Note that, as was discussed above, the primary objective of this experiment is to acquire information for use in making decisions about which components and levels to include in an intervention. In other words, the primary objective is not to obtain pristine estimates of effects per se. From this perspective and based on the resource management principle, aliasing is acceptable as long as it appreciably increases the efficiency of the design and is unlikely to lead to a poor decision.

Table 3 shows how we categorized the effects. In addition to the main effects, we identified five two-way interactions as targets of the experiment and therefore placed them in category A. These are the two-way interactions between pairs of precessation components and the two-way interactions between precessation counseling and each of the cessation counseling modalities. We also identified one three-way interaction as a target effect: the three-way interaction among the precessation components. We selected these target effects because we believe that they are likely to have the largest impact on our decisions about which components and levels of components to include in the intervention. Although we expect the remaining two-way interactions to be negligible, we placed them in category C to ensure that they would not be aliased with any category A effects. We did this to be conservative, in case we were wrong about the size of these effects. The remaining three-way interactions and all interactions involving four or more factors were placed in category B.

The Design We Chose

Once we had thought through the categorization of effects, selecting a design was accomplished readily using Proc FACTEX in SAS [36, 37] (a brief tutorial about how to select a fractional factorial design may be found in Collins et al. [33]). We identified a fractional factorial design that cut the number of required conditions in half to 32. In this design, each effect is aliased with one other effect. Each main effect is aliased with a five-way interaction, and each two-way interaction is aliased with a four-way interaction. The three-way interaction in category A is aliased with another three-way interaction. In every case, the target effects are aliased with effects that are expected to be negligible. The design is depicted in Table 4.

With $N=512$, there will be 16 subjects assigned to each of the 32 experimental conditions in this factorial experiment. The reader may wonder how statistical power ≥ 0.9 can be maintained with only 16 subjects per condition, when, in the individual experiments, comparative treatment, or RCT approaches, 16 subjects per condition would nearly always be far from adequate. The answer lies in an important difference between how effects are estimated based on factorial experiments and how effects are

Table 3 Categorization of effects in preparation for selecting a fractional factorial design

Category	Effects
A: Targets of the experiment	All six main effects Two-way interactions Precessation patch × Precessation oral NRT Precessation patch × Precessation counseling Precessation oral NRT × Precessation counseling Precessation counseling × Cessation in-person counseling Precessation counseling × Cessation phone counseling Three-way interaction Precessation patch × Precessation oral NRT × Cessation counseling
B: Not targets of the experiment; negligible in size	All interactions involving three or more factors except the three-way interaction in category A
C: Not targets of the experiment; possibly sizeable	All two-way interactions except those in category A

NRT nicotine replacement therapy

estimated in these other approaches. In factorial experiments, main effects and interactions are estimated based on combinations of experimental conditions. Different effects are based on different combinations of experimental conditions, but each effect is based on all of the conditions and therefore on all of the subjects. For example, the main effect of precessation nicotine patch will be based on a comparison of the mean of the conditions in which the patch is given (conditions 1–16 in Table 4) to the mean of the conditions in which the patch is not given (conditions 17–32). Similarly, the main effect of precessation counseling will be based on a comparison of the mean of the conditions in which precessation counseling is provided (conditions 1–4, 9–12, 17–20, and 25–28 in Table 4) to the mean of the conditions in which precessation counseling is not provided (conditions 5–8, 13–16, 21–24, and 29–32). Each of these comparisons involves all 512 subjects. Thus, in factorial experiments, adequate power can readily be achieved even when the per-condition sample size is very small, as long as the overall N is sufficiently large.

By contrast, suppose we had used a comparative treatment design with $N=512$. This design would have a control group plus six experimental conditions, one corresponding to each intervention component; there would be roughly 73 subjects assigned to each condition. Here the effect of precessation patch would be estimated by comparing the mean for those in the one condition that was assigned the precessation patch to the mean for the control group (this would not be exactly the same effect as the one based on the factorial experiment described above). Because this two-group comparison would omit the subjects who were in the other treatment conditions in the design, it would be based on only 146 subjects. Thus, it would be associated with much less statistical power than

the corresponding main effect based on the factorial experiment.

The fractional factorial design depicted in Table 4 has some interesting features. Like many fractional factorial designs, it does not contain a traditional control condition in which each factor is set to the lower level. It also does not contain a “full” experimental condition in which each treatment factor is set to the higher level. Thus, the two conditions that would make up a traditional RCT are absent. Although it may seem counterintuitive, these conditions are expendable in this design because the purpose is to examine the effect of each of the six components, not to evaluate a treatment package as a whole. It is also worth noting that every condition in this design involves giving the participant at least one component of the intervention set to the higher level, and over 80% of participants will receive at least two components set to the higher level. By contrast, in a two-arm RCT, half the participants are assigned to the control group.

Outcome Variables

Depending on the intervention component, different sets of outcome variables will be used for decision making. As discussed in the Baker et al. [1] companion paper, the phase-based framework, in concert with dependence theories, not only suggests intervention components but it also suggests the use of phase-specific outcomes. The three precessation phase components will be evaluated using the following phase-specific outcome variables: ability to establish initial cessation, number of days abstinent in the 2-week period following the target quit day, a measure of post-quit self-efficacy, and a measure of withdrawal and craving. The two cessation phase components will be

Table 4 Experimental conditions in fractional factorial design used to investigate six intervention components

Condition	Precessation nicotine patch	Precessation ad lib oral NRT	Precessation counseling	Cessation in-person counseling	Cessation phone counseling	Maintenance medication duration (weeks)
1	Patch	Oral NRT	Intensive	Minimal	Minimal	8
2	Patch	Oral NRT	Intensive	Minimal	Intensive	16
3	Patch	Oral NRT	Intensive	Intensive	Minimal	16
4	Patch	Oral NRT	Intensive	Intensive	Intensive	8
5	Patch	Oral NRT	None	Minimal	Minimal	16
6	Patch	Oral NRT	None	Minimal	Intensive	8
7	Patch	Oral NRT	None	Intensive	Minimal	8
8	Patch	Oral NRT	None	Intensive	Intensive	16
9	Patch	No oral NRT	Intensive	Minimal	Minimal	16
10	Patch	No oral NRT	Intensive	Minimal	Intensive	8
11	Patch	No oral NRT	Intensive	Intensive	Minimal	8
12	Patch	No oral NRT	Intensive	Intensive	Intensive	16
13	Patch	No oral NRT	None	Minimal	Minimal	8
14	Patch	No oral NRT	None	Minimal	Intensive	16
15	Patch	No oral NRT	None	Intensive	Minimal	16
16	Patch	No oral NRT	None	Intensive	Intensive	8
17	No patch	Oral NRT	Intensive	Minimal	Minimal	16
18	No patch	Oral NRT	Intensive	Minimal	Intensive	8
19	No patch	Oral NRT	Intensive	Intensive	Minimal	8
20	No patch	Oral NRT	Intensive	Intensive	Intensive	16
21	No patch	Oral NRT	None	Minimal	Minimal	8
22	No patch	Oral NRT	None	Minimal	Intensive	16
23	No patch	Oral NRT	None	Intensive	Minimal	16
24	No patch	Oral NRT	None	Intensive	Intensive	8
25	No patch	No oral NRT	Intensive	Minimal	Minimal	8
26	No patch	No oral NRT	Intensive	Minimal	Intensive	16
27	No patch	No oral NRT	Intensive	Intensive	Minimal	16
28	No patch	No oral NRT	Intensive	Intensive	Intensive	8
29	No patch	No oral NRT	None	Minimal	Minimal	16
30	No patch	No oral NRT	None	Minimal	Intensive	8
31	No patch	No oral NRT	None	Intensive	Minimal	8
32	No patch	No oral NRT	None	Intensive	Intensive	16

Bold font indicates high level of factor.

NRT nicotine replacement therapy

evaluated using number of days abstinent in the 2-week period following the target quit day and the measure of withdrawal and craving. Maintenance medication duration will be evaluated using two outcomes: latency to first cigarette after the target quit day and latency to seven consecutive days of smoking after the target quit day. We are also collecting data on costs associated with each intervention component.

Making Decisions Based on Experimental Results

Once the experiment has been conducted, the results will be used to address the six research questions listed above. For

the precessation components, the decisions to be made all concern whether a component should be included in the intervention. For the two cessation components and the maintenance component, the decisions concern which level of each component should be included in the intervention. The decisions will be primarily based on the size and direction of the main effect for a component.

Evidence of interactions with other components, particularly from the list of interactions in category A (Table 3), will also be considered. An interaction occurs when the effect of one factor varies significantly depending on the level of at least one other factor. In particular, we will look

carefully for evidence of interactions in which the effect of two components together is much smaller than the sum of the corresponding main effects. An example would be if precessation nicotine patch and precessation counseling each have a positive main effect, but when the two occur together, the effect is much smaller than the sum of the main effects. This might suggest that only one of the intervention components is needed, even though both have positive main effects. Other important considerations will be tolerability/acceptability of the intervention, patient adherence, and the cost of each component. For instance, if a particular component is relatively expensive to be considered for inclusion, it will need to demonstrate a correspondingly larger effect size than will a relatively inexpensive component.

Step 3B: Refinement via Experimentation and Other Methods

The rectangle representing this step in Fig. 1 is dashed to indicate that the refinement step is optional. In this step, additional research may be conducted as needed to gather more fine-grained information that will be useful in decision making. Our plans do not call for a refining step in the present cycle of MOST, but we can provide some hypothetical examples of the kind of work that might be conducted in this refinement step.

One activity in this step might be to determine the optimal level or dose of one or more intervention components. This would pertain to intervention components that can take on a range of levels, for example, the maintenance medication duration factor in the present study. The fractional factorial experiment described above will be used to determine whether there is a difference between 8 and 16 weeks of medication. It is necessary to perform this factorial experiment first because if the results indicate that there is no difference, then no refinement is warranted. However, if there is a difference between 8 and 16 weeks, it may be desirable to collect information about what would be the optimal number of weeks. For example, the benefit of a longer duration may begin to level off after, say, 14 weeks. This possibility can be investigated via additional experimentation that randomly assigns selected medication durations between 8 and 16 weeks. Another example of a refining activity might be to fine-tune the decision rules that govern the assignment of intervention components and doses in an adaptive intervention, possibly by using principles from control engineering [38, 39].

Step 4: Assembly of Beta Intervention

The objective of this step is assembly of the best beta (i.e., draft) version of the intervention. In this step, the components to be included in the beta version and the levels to which these components will be set are selected

based on the information gathered in the two steps discussed above.

Is the beta Intervention Expected to be Effective?

After the beta version of the intervention has been assembled, we will have a rough idea of the potential effectiveness and cost-effectiveness of the beta intervention, based on the estimates of the sizes of main effects and selected interactions obtained in the research conducted as part of the screening and refining steps and the estimates of cost. It is possible that the results will suggest that the beta intervention—which, based on the experimental results, is the best combination of intervention components and levels—is not likely to be more effective or cost-effective than a currently available standard of care intervention. If this happens, we will return to the beginning of the MOST framework, reconsider the theoretical model and data, and choose a new set of intervention components for examination (this is represented by the dotted arrow in Fig. 1). It is important to note that this would avoid devoting resources to a full-scale RCT evaluation of a smoking cessation intervention that is unlikely to be successful.

This return to the beginning represents a continuation of incremental progress toward a successful intervention. The previously completed work is not wasted; on the contrary, it provides a firm foundation for further progress. Based on the experimental results, the investigator knows precisely which components performed satisfactorily and which did not. Any components that performed satisfactorily can be retained for future inclusion in the intervention and in general do not need to be re-examined in the new experimentation and refinement steps. If the next MOST cycle proceeds to step 5, these components can be included in the treatment package to be evaluated in an RCT. Exploratory and secondary data analyses performed on data from the centralized database may be used to generate new hypotheses, to point the way toward improving poorly performing components, and to revise the theoretical model underlying the intervention.

Step 5: Confirmation of Effectiveness of beta Intervention via RCT

If the beta version of the smoking cessation intervention appears likely to have a sufficiently large effect, we will proceed to this step and conduct a standard RCT. The purpose of the RCT will be to confirm that the beta intervention has a statistically significant effect. The RCT will have two conditions: an experimental condition in which participants are assigned the beta version of the smoking cessation intervention developed using MOST and a comparison condition in which participants are assigned a

representative example of “usual care”. In other studies, depending on the key research questions, other suitable comparison conditions could be used, such as a wait-list control or pre-optimization version of the intervention. As discussed above, based on the previous experimentation, the investigator will have a rough sense of the expected effect size. This knowledge can be useful in powering the RCT. The primary outcome variables for the RCT will be phase-based outcomes [1] and the more traditional long-term outcomes (time to lapse/relapse and point-prevalence abstinence at 6 months post-quit).

Is the beta Version of the Intervention Effective?

The question of whether the beta version of the smoking cessation intervention is effective will be addressed via a standard hypothesis test at the usual alpha level, based on the RCT. If the results indicate that the beta version of the intervention is more effective than the current version of the intervention, we will proceed to the next step in MOST. If the results indicate that the beta version of the intervention is not more effective than the current version, we will return to the beginning of the MOST framework and reconsider the theoretical model and data (this is represented by the dotted arrow in Fig. 1). However, we are optimistic that by using the MOST framework we will engineer a significantly improved smoking cessation intervention.

Step 6: Release of New Intervention

In this step, the new intervention is assigned to a unique version number, documented and released to the scientific and clinical communities. The previous steps of MOST have been aimed at selection of intervention components that are feasible, effective, and cost-effective. This will facilitate the release of this intervention package and its incorporation into regular clinical practice. The use of MOST to facilitate translation is discussed further below.

Beginning a New Cycle of MOST

Motivated by the continuous optimization principle, at the time of the release of the new intervention package, we will have already begun gathering scientific literature, discussing clinical observations, and conducting secondary data analyses in preparation for beginning a new cycle of MOST to identify further improvements to the intervention. The secondary data analyses will be conducted on data from the experiments that were conducted in the course of investigating individual intervention components, refinement, and confirmation of the beta intervention and placed in a centralized database (see Fig. 1). Data in the centralized database are available for exploratory and secondary data

analyses that may generate hypotheses for future cycles of the MOST procedure aimed at additional improvements to the intervention. These analyses may also inform the theory underlying the intervention. This database is cumulative, so that as the MOST procedure is repeated (as the continuous optimization principle suggests), more data are acquired, archived, and made available for analysis.

An Integrated Translation Strategy: Using MOST to Further Type II Translation

Several authors (e.g., [3, 40–43]) have suggested that the process of translating research findings into real-world interventions could be hastened through the use of a strategy that builds translation potential into initial treatment evaluation. According to these authors, successful translation of smoking cessation interventions depends upon three criteria: (a) ease of delivery, (b) effectiveness in real-world settings, and (c) compatibility and integration with real-world delivery systems. We propose that using MOST to select intervention components and levels based on these three criteria can be an effective way to develop intervention packages that can readily be implemented and sustained in real-world systems. This is described in more detail for each of the three criteria below.

Ease of Delivery

Any intervention components to be examined using MOST should be selected based on their high translation potential and their appropriateness for dissemination/translation into real-world use contexts [40, 44]. Instead of the typical situation where intensive, expensive, and burdensome interventions are initially tested in the efficacy context (e.g., often with up to a dozen lengthy in-person counseling visits), with relatively little concern for ultimate translation, initial evaluation of intervention components could in many cases have translation potential “built-in”, i.e., the intervention components are developed and initially evaluated in the effectiveness context (see [10]). This approach would favor intervention components that are relatively brief and designed to produce minimal burden on both program/clinic staff and patients/participants [6]. For instance, in our research, as described above, even the higher levels of phone and in-person cessation phase counseling examined are relatively brief and involve only a few contacts.

Investigators using MOST to optimize the effectiveness of interventions may wish to include intervention components of varying resource intensity because burden and complexity is a leading cause of poor translation [6]. This approach will permit the development of interventions that reflect the optimal combination of benefit and intensity/

costs. For instance, consistent with chronic care models [45, 46], cessation intervention components might shift much responsibility for intervening from the primary care physician to appropriate support staff (e.g., case managers, health educators, quitline counselors, or pharmacists). In addition, the intervention components and levels examined should differ in terms of the burden they impose on patients. Intervention components and levels that impose a greater burden on patients should be incorporated into an optimal comprehensive intervention only if justified by empirical data on effectiveness, cost-effectiveness, patient utilization, and reach. In other words, we recommend selecting intervention components and levels for inclusion in a comprehensive intervention so as to optimize a burden/outcome ratio established in real-world settings and populations. This is consistent with recommendations that treatment evaluation research include a broad array of outcomes relevant to real-world adoption [10, 44].

Effectiveness in Real-World Settings

Because intervention effectiveness is an important driver of intervention adoption and support, it is important that smoking cessation intervention optimization using MOST identifies intervention components and levels that deliver significant benefit in real-world healthcare settings with diverse real-world populations [10, 44]. The best strategy to ensure this is to test all candidate intervention components and levels in real-world use contexts so that their performance is based upon effectiveness, rather than efficacy, data. The blurring of efficacy and effectiveness research may have some costs. For instance, very intensive and complex assessments and intervention components would not be tested under ideal circumstances, and thus, their true optimal effects would be missed. This could have negative theoretical as well as clinical consequences. Thus, there remains a role for efficacy research in those cases where the theoretical value of an experiment is high, or where the intervention is not intended for widespread dissemination. However, in many cases, it will be desirable to conduct MOST in real-world contexts and populations.

Of course, efficacy and effectiveness research are not mutually exclusive. Research goals may foster a mix of real-world and research features and mechanisms, e.g., recruitment may be conducted in real-world settings via clinic personnel, but actual assessment and intervention delivery may be carried out by research personnel serving as chronic care case managers [45–47]. Also, many of the goals of efficacy research could perhaps be attained even in the effectiveness context via the use of enhanced technology and communications systems. For instance, eHealth interventions, smart phones/cell phones, and the electronic medical record [48] could permit provision of interventions

and collection of time-stamped data in ways that do not impose extraordinary subject or staff burdens. In the experiment depicted in Table 4, data will be collected from the health care systems' electronic medical records and directly from participants via telephone.

Compatibility and Integration with Real-World Delivery Systems

If research to optimize smoking cessation interventions in real-world contexts is to be successful and yield broadly generalizable effects, then the relation between the intervention elements and the real-world context (e.g., a health care program) must be broadly applicable, feasible, and effective. That is, functions such as smoker identification, tracking, intervention delivery, appraisals of outcomes, and so on, must be translatable (e.g., [44]). In other words, the system features in which the research is embedded are major determinants of the likely translation potential of the research. Therefore, researchers must ensure that these system features will not only satisfy their research goals (e.g., result in adequate recruitment) but also be feasible for a broad range of targeted real-world contexts. This feasibility can be assessed by gathering input from targeted participants, consumers, and providers during intervention and research program development [44]. For instance, the role and function of interveners, even if they are research staff, must be ones that are potentially feasible for real-world personnel to deliver. That is why, in our research, medical assistants on staff at the participating primary care clinics complete the initial recruitment for the experiment. Interventions will be delivered by research staff functioning as clinical care managers. This model could easily be transitioned to one in which medical assistants continue to recruit smokers but then refer them to health counselors or chronic disease management staff who then provide the specific cessation intervention.

The Gordian Knot of clinical tobacco control is an inability to achieve more consistent delivery of evidence-based tobacco cessation interventions by real-world healthcare personnel [6, 49–52]. The investigator must consider how system resources can be engineered to ensure the systematic delivery of effective interventions [2, 3, 42, 43, 53] and also must consider the appropriate professional role of the interveners in a real-world context. The possibilities are legion with viable models ranging from clinic staff merely referring to a state quitline or an eHealth intervention [54, 55] to a chronic care model approach to disease management in which staff take primary responsibility for assessment, intervening, and patient tracking [7]. The vital point is that to a great extent, any intervention research conducted in real-world settings simultaneously tests the interventions and the

real-world contexts in which the interventions are embedded. Thus, in designing research using MOST, it is important to attend to contextual features and systems that will determine the effectiveness and dissemination potential of the intervention components and levels under consideration.

Additional Considerations in Type II Translational Research

The three features described above certainly do not exhaust all the steps needed to effect type II translation. For instance, as noted by Glasgow [10, 44], the investigator needs to sample an array of participant, provider, and system features to determine if the delivery and effects of intervention components are moderated by such factors. In addition, analyses should target a broad array of outcomes including effectiveness, cost-effectiveness, reach, and adherence so that interventions are assessed on criteria or outcomes that go beyond efficacy and have clear relevance to dissemination [44].

Some Considerations, Challenges, and Potential Benefits Associated with Using MOST

Considerations Associated with Using MOST

What Constitutes a Component?

Up to this point, we have purposefully avoided providing a definition of the term *intervention component* because this term may have different meanings in different situations. An intervention component can be any aspect of an intervention that is of interest and can be separated out for study. Components may (a) be part of the intervention per se, (b) pertain to how a particular component or the entire intervention is administered or who administers it, (c) be some aspect of the environment in which the intervention takes place, (d) pertain to adherence to the intervention, or (e) represent any aspect of the intervention pertaining to type I translation, efficacy, effectiveness, cost, or type II translation. Within the research reported here, a variety of intervention components were examined, including pharmaceuticals and different types of counseling. All of these components were candidates to be included as parts of the intervention. The research reported here also included a component pertaining to optimal duration of nicotine replacement. Each application of MOST must start with a theoretical model (step 1 in Fig. 1) to determine the relevant intervention components and identify the ones that are the highest priority for examination.

How Many Components?

A closely related consideration is the number of intervention components to examine in a single cycle of MOST. Given that a component may pertain to virtually any aspect of an intervention, the list of components that are interesting and worthwhile for examination may be nearly endless. We suggest taking the resource management principle into account when deciding how many components to examine. This involves prioritizing the list of components in terms of importance and then selecting the experimental design that enables examination of as many of the components at the top of the list as available resources will support. The choice of experimental design is critical because it is a major determinant of what can be accomplished with a given level of resources.

In most cases, it will be impossible to examine every potential component in a single cycle of MOST. We see this as a realistic perspective on conducting scientific work rather than a limitation of MOST. An investigator who subscribes to the continuous optimization principle will expect to conduct numerous cycles of MOST, in which there will be opportunities to examine important intervention components that could not be examined in previous cycles.

Evaluating Design Alternatives

Different experimental designs address subtly different research questions. Moreover, a design that is efficient for addressing one research question may be inefficient for addressing a different research question. Thus, when planning a cycle of MOST, it is essential to consider all the design options available in light of the resource management principle (Collins et al. [33] discuss how to compare the efficiency and cost of various designs).

To illustrate the use of the resource management principle, let us compare and contrast the information provided by the factorial experiment and the RCT when they are used in MOST. In the experimentation step of MOST (step 3), the research questions are concerned with the effects of individual intervention components. Therefore, the resource management principle suggests that the investigator will want to seek the experimental design that enables valid estimation of the effects of individual intervention components in the most efficient manner. As we described above, in our research, a fractional factorial design was the clear choice for experimentation on individual intervention components. Like a complete factorial design, the fractional factorial design enabled us to estimate the main effect of each intervention component as well as selected interactions between components. The fractional factorial design we

selected has the same statistical power as the corresponding complete factorial design, while requiring only half the experimental conditions.

By contrast, in the confirmation step of MOST (step 5), the primary research question concerns on evaluating the effectiveness of the beta version of the intervention as compared to a control group. Here the RCT is usually the most efficient and appropriate design. A factorial experiment would in most cases be much less efficient and may not even contain the experimental conditions necessary for addressing the primary research question.

In short, intervention scientists have at their disposal a variety of experimental designs, including factorial experimental designs, the RCT, and others. Each of these designs is at its best when applied to a particular type of research question. When used in concert, as they are in MOST, the various types of experiments can address a comprehensive set of research questions and produce information that bears directly on increasing the public health impact of behavioral interventions.

Choosing Outcome Variables

One important consideration is whether the primary outcome variables for the purpose of making decisions about components and levels should, or even can, be the same as the primary outcome variables for the purpose of evaluating the resulting treatment package. In the smoking cessation study described here, decisions about which intervention components and levels to select for inclusion in the intervention will be based on phase-specific outcomes such as the number of days abstinent during the 2 weeks following the target quit day. By contrast, when the assembled intervention package is evaluated by means of an RCT, the primary outcome variables will be long-term outcomes. We decided to use short-term outcomes to select components and component levels because we believe short-term outcomes will be more sensitive to the performance of the intervention components. In addition, using short-term outcomes will help to keep the overall study on a reasonable timeline. Short-term outcomes have been demonstrated to be highly correlated with 6-month point-prevalence abstinence (e.g., [56]).

Behavioral interventions in some areas of tobacco control may have a primary outcome variable that is much more temporally distal than 6-month point-prevalence abstinence. For example, a school-based smoking prevention program may be administered when children are 12 years old and have as its key outcome variable point prevalence abstinence at age 16. If the intervention is designed to act directly on mediators, which in turn affect the outcome, measures of the mediators may provide the best short-term outcomes for purposes of making decisions about components and component levels.

In addition to health behavior outcomes, there are a number of other outcome variables that can be considered both in making decisions about components and component levels and in evaluating the intervention package. Examples include cost-effectiveness, participant burden, attractiveness of the intervention to participants, compliance, and attrition. These outcomes can all be considered to develop an effective package that uses resources prudently and will be able to translate to real-world clinical settings.

Drawing Scientific Conclusions vs. Making Decisions

When a randomized experiment is conducted, the probabilities associated with a type I error (mistakenly rejecting the null hypothesis) and a type II error (mistakenly failing to reject the null hypothesis) are known, at least approximately. When a statistical analysis is performed to draw scientific conclusions, there is a widely accepted scientific standard that statistical significance will be concluded if the probability of a type I error does not exceed 0.05. We recommend conforming to this standard in MOST when using the results of the RCT to confirm the efficacy or effectiveness of the beta version of the intervention because the RCT is aimed at drawing a scientific conclusion about the effect of the intervention.

However, the analyses that lead up to assembling the beta version of the intervention are primarily used for making decisions about which components and levels to include in the beta version of the intervention. Under these circumstances, the resource management principle suggests weighing the relative cost of type I and type II errors and possibly using more flexible guidelines. If the cost of overlooking a potentially effective component is considered greater than the cost of mistakenly including it in the intervention, the type I error rate can be increased to be larger than 0.05 in order to decrease the type II error rate. An investigator may determine that a very expensive component under consideration will have to demonstrate an effect size greater than some pre-specified value in order to make its inclusion worthwhile, even if a lower effect size would achieve statistical significance. In general, for the purpose of making decisions when building a multicomponent intervention, traditional hypothesis testing per se is less important than using sensible and appropriate criteria for choosing intervention components and levels. In this context, statistical hypothesis testing is best treated as the servant rather than the master.

Challenges

Implementation of Factorial Experiments in the Field

For applications of MOST that involve a large factorial experiment, one challenge may be implementing this

experiment in the field. To maintain fidelity to the design, it is critical to ensure that each participant receives the intervention components included in his or her assigned experimental condition and only those components. To implement the 32-condition factorial experiment depicted in Table 4, we are assigning a case manager to each participant. The case managers will use a database and a computerized tracking system for ensuring and documenting that each participant receives the appropriate components and levels at the appropriate times.

Careful planning, well-trained staff, and creative use of computers and the Internet can greatly increase the feasibility of implementing a large number of experimental conditions in field settings. However, depending on available resources, it may not be possible to implement 32 experimental conditions. We wish to stress that this does not imply that MOST cannot be used. MOST does not require that any particular design approach be used, only that the resource management principle be followed in selecting a design. Following the resource management principle entails considering all available research designs, even those that may be a bit unfamiliar, to find the one that makes the best use of the resources at hand to address the highest-priority research questions and move science forward the fastest. In some cases, it may be necessary to consider taking a calculated risk, such as making the assumptions that will permit use of a highly efficient fractional factorial design. As long as the resource management principle is followed, MOST can be used to make incremental progress in intervention optimization, even with limited resources.

Decision Making with Multiple Outcome Variables

Often decisions concerning components and levels must be made based on multiple outcome variables. For example, in the study described here, we will be making decisions about the precessation components based on smoking behavior, self-efficacy, withdrawal/craving, cost, and feasibility. Sometimes these decisions may require trade-offs between outcomes (for example, if a component looks relatively strong in terms of one outcome measure but not in another, then the investigator will be forced to choose which outcome is more important). We are actively investigating ways to incorporate formal decision making models when such trade-offs are required during intervention optimization.

Optimizing for Whom?

In the research described here, we are looking for intervention components and levels of components that work best overall. A future step will be to investigate

adaptive intervention strategies (e.g., [57]) that are tailored to specific characteristics of individual participants. A starting point for this work will be secondary analyses performed on the data in the centralized database (Fig. 1). In particular, we will look for individual characteristics that interact with intervention components. For example, if an intervention component has a larger effect for women than for men, this would suggest that a different version or different intensity of the component might be more successful with men. Recent methodological work on building and optimizing adaptive interventions has included development of the Sequential Multiple Assignment Randomized Trial [58] and exploration of the idea of applying control algorithms from engineering to adaptive behavioral interventions to improve their performance [38, 39]. These approaches fit well within the MOST framework.

The Realities of Funding

As our project and a few others (e.g., [55]) have demonstrated, it is possible to obtain funding for MOST and similar approaches from the National Institutes of Health (NIH). Still, obtaining research funding to implement MOST can be challenging. First, MOST rests on the conviction that useful information can be gained by examining the effects of individual intervention components, provided that experimental designs are chosen prudently. However, there are some pervasive and enduring misconceptions about experimental design (discussion of these may be found in [33] and [35]) that may lead some reviewers of grant proposals to make the unwarranted assumption that examination of individual intervention components with acceptable statistical power is not feasible. Of course, it is up to the applicant to offer a compelling argument for feasibility in a grant proposal; we hope that when presented with such arguments reviewers will keep an open mind.

Second, because MOST often requires sequential decision making, the details of the approach to be used in one step may be dependent on the outcomes of previous steps. This may make it impossible for an applicant to be specific about downstream elements of the proposed research. For example, it would not be possible to determine the best design for the refinement step (step 3B) without first having the results from the experimentation step (step 3A). Various contingencies may be listed in an application up to a point, but there may be too many alternatives to include them all within a reasonable page limit. In a case like this, a fundable priority score is within reach only if reviewers are comfortable with a presentation that offers a well-reasoned strategy for some aspects of the research rather than a detailed account of tactics.

A third challenge is whether a full cycle of MOST can be completed within the 5-year duration of the typical NIH funding cycle. This depends on many factors, such as how experienced the investigators are with the MOST approach, how quickly study participants can be recruited and how well they can be retained, and how long the experimentation is expected to take. One important limiting factor is how long it will take to obtain results in the confirmation step (step 5). As discussed above, in some fields, it may be possible to measure outcomes of ultimate interest soon after the intervention, whereas in others it may be necessary to wait months or even years. In the project described in this article, we plan to conduct the confirmation step after our current 5-year funding period is over. This will require applying for additional funding for the confirmation step.

We would like to comment on the perceived necessity of producing an implementation-ready behavioral intervention within a 5-year funding cycle. The 5-year funding cycle is merely an administrative necessity. It has no intrinsic scientific meaning or merit and was never meant to be a scientific imperative. We believe that rather than trying to grab the brass ring of a statistically significant treatment package in each 5-year cycle, it is better to adopt a more programmatic approach like MOST and improve the intervention iteratively, in cycles of whatever length makes the most sense for the problem at hand. In some areas, an RCT will take a long time. Here, repeated applications of the experimentation and (where appropriate) refinement steps may be advisable before an RCT is undertaken.

The NIH can promote the use of intervention optimization approaches like MOST. One avenue is helping reviewers of grant proposals increase their awareness of emerging viable alternatives to traditional intervention development approaches. Funding announcements can stress both evaluation and optimization. Perhaps most importantly, the National Institutes of Health can help foster a culture that subscribes to a long-range view on intervention science, one that stresses programmatic and incremental progress toward the goal of maximum public health impact.

Potential Benefits of Using MOST

We believe that the MOST framework offers several benefits over other approaches in common use. First, because this is a principled and systematic approach to improving smoking cessation interventions, it is likely to be the fastest and most direct way to more effective smoking cessation intervention in the long run. A key feature of MOST is that each new intervention produced will have been engineered and empirically demonstrated to be an improvement over the previous version. Thus, adoption of this approach can be expected to result in demonstrable incremental and cumulative progress.

Second, this approach provides a straightforward method to test hypotheses about new intervention components or the revision of existing components as they are suggested by emerging new scientific findings, technologies, counseling approaches, or pharmaceuticals. MOST can also be used to update interventions periodically to incorporate new approaches or innovations while ensuring that the changes made do not erode effectiveness.

Third, MOST provides a natural method of including cost information in decision making. The efficacy or effectiveness of each intervention component can be viewed in relation to its cost in terms of money, length of time required, implementation logistics, amount of training required, or other resources. Weighing efficacy or effectiveness against cost can play an important role in selecting intervention components and levels of components. For example, it is possible to aim decision making at achieving cost-effectiveness, determining the most efficacious or effective intervention that can be delivered within a certain cost range, controlling implementation complexity by choosing a maximum number of components to include and choosing the ones that maximize effectiveness, or achieving the most effective intervention that takes a certain specified amount of time to deliver.

Fourth, MOST will contribute to the building of a coherent cumulative base of scientific knowledge about smoking cessation. The resource management principle, when properly and consistently applied, can increase the yield of scientific information from research and thereby hasten the progress of knowledge acquisition. In addition, when designs are used that tease out the effects of several individual intervention components simultaneously, it becomes possible to perform analyses to investigate whether a specific intervention component is mediated by a particular variable. In contrast, mediation analyses based on standard RCTs are limited to investigating mediation of the treatment effect as a whole. Thus, the approach proposed in this paper will not only improve clinical smoking cessation treatment but will further the understanding of its underlying mechanisms.

Conclusions

In this article and the companion article by Baker et al. [1], we propose a comprehensive strategy for achieving more effective tobacco interventions, based on [1] use of a phase-based framework to guide the choice of intervention components and outcome measures and [2] use of the MOST framework for efficient development, optimization, and evaluation of interventions. The strategy requires taking a long view that emphasizes a steady pace of programmatic and incremental intervention development.

We believe that the strategy we propose will speed the identification of effective tobacco interventions and their translation into real-world use. This strategy may have wide applicability to the development of behavioral interventions for other chronic health challenges.

Acknowledgment This research was supported by a grant 9P50CA143188-11 from the National Cancer Institute. Dr. Collins was supported via grant P50DA10075, from the National Institute on Drug Abuse. Dr. Baker was supported via NCI 1K05CA139871. Dr. Piper was supported by grant 1UL1RR025011 from the Clinical and Translational Science Award (CTSA) program of the National Center for Research Resources (NCRR), National Institutes of Health (NIH). Dr. Cook was supported by K08DA021311.

Conflicts of Interest Linda M. Collins, Timothy B. Baker, Robin Mermelstein, Megan E. Piper, Stevens S. Smith, Bruce A. Christiansen, Tanya R. Schlam, and Jessica W. Cook have no potential conflicts of interest to disclose. Douglas E. Jorenby has received research support from the National Institute on Drug Abuse, the National Cancer Institute, Pfizer, Inc., Sanofi-Synthelabo, and Nabi Biopharmaceuticals. He has received support for educational activities from the National Institute on Drug Abuse and the Veterans Administration and consulting fees from Nabi Biopharmaceuticals. Over the last 3 years, Michael C. Fiore served as an investigator on research studies at the University of Wisconsin that were funded by Nabi Biopharmaceuticals.

References

- Baker TB, Mermelstein R, Collins LM, et al. New methods for tobacco dependence treatment research. *Ann Behav Med.* 2011; in press.
- Fiore MC, Bailey WC, Cohen SJ. *Treating tobacco use and dependence: Clinical practice guideline.* Rockville: U.S. Department of Health and Human Services, U.S. Public Health Service; 2000.
- Fiore MC, Jaen CR, Baker TB, et al. *Treating tobacco use and dependence: 2008 update.* Rockville: U.S. Department of Health and Human Services, U.S. Public Health Service; 2008.
- Collins LM, Murphy SA, Nair VN, Strecher VJ. A strategy for optimizing and evaluating behavioral interventions. *Ann Behav Med.* 2005;30(1):65–73.
- Collins LM, Murphy SA, Strecher V. The Multiphase Optimization Strategy (MOST) and the Sequential Multiple Assignment Randomized Trial (SMART): New methods for more potent eHealth interventions. *Am J Prev Med.* 2007;32(5 Suppl):S112–8.
- Hollis JF, Bills R, Whitlock E, Stevens VJ, Mullooly J, Lichtenstein E. Implementing tobacco interventions in the real world of managed care. *Tob Control.* 2000;9(Suppl 1):I18–24.
- Hung DY, Shelley DR. Multilevel analysis of the Chronic Care Model and 5A services for treating tobacco use in urban primary care clinics. *Health Serv Res.* 2008;44:103–27.
- Kottke TE, Solberg LI, Brekke ML, Cabrera A, Marquez MA. Delivery rates for preventive services in 44 Midwestern clinics. *Mayo Clin Proc.* 1997;72(6):515–23.
- Glasgow RE, Emmons KM. How can we increase translation of research into practice? Types of evidence needed. *Annu Rev Public Health.* 2007;28:413–33.
- Glasgow RE, Lichtenstein E, Marcus AC. Why don't we see more translation of health promotion research to practice? Rethinking the efficacy-to-effectiveness transition. *Am J Public Health.* 2003;93(8):1261–7.
- Box GEP, Hunter, W.G., & Hunter, J.S. *Statistics for experimenters: An introduction to design, data analysis, and model building.* New York: Wiley; 1978.
- Wu CF, Hamada M. *Experiments: Planning, analysis, and parameter design optimization.* New York: Wiley; 2000.
- Shadish WR, Cook TD, Campbell DT. *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton-Mifflin; 2002.
- Campbell MK, Mollison J, Steen N, Grimshaw JM, Eccles M. Analysis of cluster randomized trials in primary care: A practical approach. *Fam Pract.* 2000;17(2):192–6.
- Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: The new Medical Research Council guidance. *BMJ.* 2008;337:a1655.
- Rose JE, Behm FM, Westman EC, Kukovich P. Precessation treatment with nicotine skin patch facilitates smoking cessation. *Nicotine Tob Res.* 2006;8(1):89–101.
- Moore D, Aveyard P, Connock M, Wang D, Fry-Smith A, Barton P. Effectiveness and safety of nicotine replacement therapy assisted reduction to stop smoking: Systematic review and meta-analysis. *BMJ.* 2009;338:b1024.
- Hughes JR, Carpenter MJ. The feasibility of smoking reduction: An update. *Addiction.* 2005;100(8):1074–89.
- Hollis JF, McAfee TA, Fellows JL, Zbikowski SM, Stark M, Riedlinger K. The effectiveness and cost effectiveness of telephone counselling and the nicotine patch in a state tobacco quitline. *Tob Control.* 2007;16 Suppl 1:i53–9.
- Zhu SH, Stretch V, Balabanis M, Rosbrook B, Sadler G, Pierce JP. Telephone counseling for smoking cessation: Effects of single-session and multiple-session interventions. *J Consult Clin Psychol.* 1996;64(1):202–11.
- Zhu SH, Tedeschi G, Anderson CM, et al. Telephone counseling as adjuvant treatment for nicotine replacement therapy in a “real-world” setting. *Prev Med.* 2000;31(4):357–63.
- Zhu SH, Anderson CM, Tedeschi GJ, et al. Evidence of real-world effectiveness of a telephone quitline for smokers. *New Engl J Med.* 2002;347(14):1087–93.
- Tonstad S, Tonnesen P, Hajek P, Williams KE, Billing CB, Reeves KR. Effect of maintenance therapy with varenicline on smoking cessation: A randomized controlled trial. *JAMA.* 2006;296(1):64–71.
- Covey LS, Glassman AH, Jiang H, et al. A randomized trial of bupropion and/or nicotine gum as maintenance treatment for preventing smoking relapse. *Addiction.* 2007;102(8):1292–302.
- Hays JT, Hurt RD, Rigotti NA, et al. Sustained-release bupropion for pharmacologic relapse prevention after smoking cessation. A randomized, controlled trial. *Ann Intern Med.* 2001;135(6):423–33.
- Dale Horst W, Klein MW, Williams D, Werder SF. Extended use of nicotine replacement therapy to maintain smoking cessation in persons with schizophrenia. *Neuropsychiatr Dis Treat.* 2005;1(4):349–55.
- Helgason AR, Tomson T, Lund KE, Galanti R, Ahnve S, Gilljam H. Factors related to abstinence in a telephone helpline for smoking cessation. *Eur J Public Health.* 2004;14(3):306–10.
- Fiore MC, Smith SS, Jorenby DE, Baker TB. The effectiveness of the nicotine patch for smoking cessation. A meta-analysis. *JAMA.* 1994;271(24):1940–7.
- Piper ME, Federmen EB, McCarthy DE, et al. Using mediational models to explore the nature of tobacco motivation and tobacco treatment effects. *J Abnorm Psychol.* 2008;117(1):94–105.
- Jorenby DE, Hays JT, Rigotti NA, et al. Efficacy of varenicline, an alpha4beta2 nicotinic acetylcholine receptor partial agonist, vs placebo or sustained-release bupropion for smoking cessation: A randomized controlled trial. *JAMA.* 2006;296(1):56–63.

31. Ferguson SG, Shiffman S, Gwaltney CJ. Does reducing withdrawal severity mediate nicotine patch efficacy? A randomized clinical trial. *J Consult Clin Psychol*. 2006;74(6):1153–61.
32. Shiffman S, Scharf DM, Shadel WG, et al. Analyzing milestones in smoking cessation: Illustration in a nicotine patch trial in adult smokers. *J Consult Clin Psychol*. 2006;74(2):276–85.
33. Collins LM, Dziak JJ, Li R. Design of experiments with multiple independent variables: A resource management perspective on complete and reduced factorial designs. *Psychol Methods*. 2009;14(3):202–24.
34. Fisher RA. The arrangement of field experiments. *Journal of the Ministry of Agriculture of Great Britain*. 1926;33:505–13.
35. Chakraborty H. The design and analysis aspects of cluster randomized trials. In: Biswas A, Datta S, Fine JP, Segal MR, eds. *Statistical Advances in the Biomedical Sciences*. Hoboken: Wiley; 2008:67–80.
36. SAS Institute. *SAS/STAT 9.1 User's Guide*. Cary: SAS Institute; 2004.
37. SAS Institute. *SAS/QC9.1 User's Guide*. Cary: SAS Institute; 2004.
38. Rivera DE, Pew MD, Collins LM. Using engineering control principles to inform the design of adaptive interventions: A conceptual introduction. *Drug Alcohol Depend*. 2007;88 Suppl 2:S31–40.
39. Zafra-Cabeza A, Rivera DE, Collins LM, Riado MA, Camacho EF. A risk-based model predictive control approach to adaptive interventions in behavioral health. *Proceedings of the 45th IEEE Conference on Decision and Control*. 2006:673–8.
40. Carroll KM, Rounsaville BJ. Bridging the gap: A hybrid model to link efficacy and effectiveness research in substance abuse treatment. *Psychiatr Serv*. 2003;54(3):333–9.
41. Carroll KM, Rounsaville BJ. A vision of the next generation of behavioral therapies research in the addictions. *Addiction*. 2007;102(6):850–62; discussion 63–9.
42. Hopkins DP, Briss PA, Ricard CJ, et al. Reviews of evidence regarding interventions to reduce tobacco use and exposure to environmental tobacco smoke. *Am J Prev Med*. 2001;20(2 Suppl):16–66.
43. Hopkins DP, Husten CG, Fielding JE, Rosenquist JN, Westphal LL. Evidence reviews and recommendations on interventions to reduce tobacco use and exposure to environmental tobacco smoke: A summary of selected guidelines. *Am J Prev Med*. 2001;20(2 Suppl):67–87.
44. Glasgow RE, Klesges LM, Dziewaltowski DA, Bull SS, Estabrook P. The future of health behavior change research: What is needed to improve translation of research into health promotion practice? *Ann Behav Med*. 2004;27(1):3–12.
45. Glasgow RE, Orleans CT, Wagner EH. Does the chronic care model serve also as a template for improving prevention? *Milbank Q*. 2001;79(4):579–612, iv–v.
46. McLellan A, Lewis, DC, O'Brein, CP, Kleber, HD. Drug dependence, a chronic medical illness: Implications for treatment, insurance, and outcomes evaluation. *JAMA*. 2000;284(13):1689–95.
47. Hung DY, Rundall TG, Tallia AF, Cohen DJ, Halpin HA, Crabtree BF. Rethinking prevention in primary care: Applying the chronic care model to address health risk behaviors. *Milbank Q*. 2007;85(1):69–91.
48. Kreps GL, Neuhauser L. New directions in eHealth communication: Opportunities and challenges. *Patient Educ Couns*. 2010;78(3):329–36.
49. Friend K, Levy DT. Smoking treatment interventions and policies to promote their use: A critical review. *Nicotine Tob Res*. 2001;3(4):299–310.
50. Evaluation resources. Available at: www.dhs.ca.gov/tobacco. 2000.
51. Centers for Disease Control and Prevention. Use of FDA-approved pharmacological treatments for tobacco dependence—United States, 1984–1998. *MMWR*. 2000;49:665–8.
52. Rothenich SF, Woolf SH, Johnson RE, et al. Effect on cessation counseling of documenting smoking status as a routine vital sign: An ACORN study. *Ann Fam Med*. 2008;6(1):60–8.
53. Fiore MC, Bailey WC, Cohen SJ. *Smoking cessation: Clinical practice guideline no. 18*. Rockville: U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research; 1996.
54. Japuntich SJ, Zehner ME, Smith SS, et al. Smoking cessation via the internet: A randomized clinical trial of an internet intervention as adjuvant treatment in a smoking cessation intervention. *Nicotine Tob Res*. 2006;8 Suppl 1:S59–67.
55. Strecher VJ, McClure J, Alexander G, et al. Web-based smoking cessation components and tailoring depth: Results of a randomized trial. *Am J Prev Med*. 2008;34:373–81.
56. Kenford SL, Fiore MC, Jorenby DE, Smith SS, Wetter D, Baker TB. Predicting smoking cessation. Who will quit with and without the nicotine patch. *JAMA*. 1994;271(8):589–94.
57. Collins LM, Murphy SA, Bierman KL. A conceptual framework for adaptive preventive interventions. *Prev Sci*. 2004;5(3):185–96.
58. Murphy SA, Lynch KG, Oslin D, McKay JR, TenHave T. Developing adaptive treatment strategies in substance abuse research. *Drug Alcohol Depend*. 2007;88 Suppl 2:S24–30.