Moving beyond the treatment package approach to developing behavioral interventions: addressing questions that arose during an application of the Multiphase Optimization Strategy (MOST)

By: David L. Wyrick, Kelly L. Rulison, Melodie Fearnow-Kenney, Jeffrey J. Milroy, Linda M. Collins

Abstract:

Given current pressures to increase the public health contributions of behavioral interventions, intervention scientists may wish to consider moving beyond the classical treatment package approach that focuses primarily on achieving statistical significance. They may wish also to focus on goals directly related to optimizing public health impact. The Multiphase Optimization Strategy (MOST) is an innovative methodological framework that draws on engineering principles to achieve more potent behavioral interventions. MOST is increasingly being adopted by intervention scientists seeking a systematic framework to engineer an optimized intervention. As with any innovation, there are challenges that arise with early adoption. This article describes the solutions to several critical questions that we addressed during the first-ever iterative application of MOST. Specifically, we describe how we have applied MOST to optimize an online program (*myPlaybook*) for the prevention of substance use among college student-athletes. Our application of MOST can serve as a blueprint for other intervention scientists who wish to design optimized behavioral interventions. We believe using MOST is feasible and has the potential to dramatically improve program effectiveness thereby advancing the public health impact of behavioral interventions.

Keywords: Multiphase Optimization Strategy | Intervention development | Substance use prevention | Experimental design

Article:

Implications

Research: Researchers should move beyond the "treatment package" approach to intervention development and evaluation and explore ways to optimize behavioral interventions to maximize public health impact.

Practice: Intervention scientists should investigate iterative approaches to intervention development that incorporate systematic and resource efficient principles of product engineering.

Policy: Resources should be committed to support behavioral interventions that have been optimized to a specific and meaningful criterion.

Due to growing expectations that behavioral interventions should have a clinically meaningful public health impact, intervention scientists have started considering how to optimize their programs. A program has been optimized when it meets a priori criteria, expressed in terms of attributes such as efficacy, effectiveness, efficiency, or cost-effectiveness. One approach to achieving this goal is to use the Multiphase Optimization Strategy (MOST) developed by Collins et al. [*1–3*]. MOST is a framework for program development and evaluation that is inspired by approaches used in engineering research and product development [*1, 4*]. Using MOST allows intervention scientists to optimize behavioral interventions by specifying a desired criterion and systematically engineering the program to meet this criterion.

MOST consists of three phases: Preparation, Optimization, and Evaluation. In the Preparation phase, intervention scientists draw on one or more theories to develop a conceptual model that will form the basis for the behavioral intervention. As part of this phase, they also identify which components, or parts of the intervention, to evaluate and develop and pilot those components. Finally, they select the desired optimization criterion, such as achieving a clinically meaningful public health impact (e.g., a specific effect size). In the Optimization phase, intervention scientists conduct one or more experiments to obtain information about the performance of each component. This information is used to decide which components should be retained to form the optimized intervention, that is, the intervention that meets the specified criterion. In the Evaluation phase, intervention scientists evaluate the performance of the optimized intervention using a randomized controlled trial (RCT). Because, as noted above, MOST is a framework rather than an off-the shelf procedure, its implementation can differ considerably across applications. Examples of applications of MOST can be found in Strecher et al. [*5*], Collins et al. [*3*], Caldwell et al. [*6*], and McClure et al. [*7*].

Because many intervention scientists are relatively unfamiliar with MOST, they may be concerned about perceived challenges associated with its implementation. In this article, we review several questions that intervention scientists may ask as they decide whether to adopt MOST. We describe how we addressed these questions as we implemented MOST to improve the effectiveness of *myPlaybook*, an alcohol and other drug prevention program aimed at NCAA student-athletes. Our goal is to alleviate concerns for those who see the value of MOST but have concerns about the feasibility of implementing it in the field. Before turning

to the questions, we begin with a brief overview of how we have applied MOST to optimize *myPlaybook*.

APPLYING MOST TO OPTIMIZE *myPlaybook*

myPlaybook: an online intervention targeting substance use among college student-athletes

College student-athletes are at increased risk of heavy alcohol use, smokeless tobacco use, and the use of performance-enhancing substances compared to non-athletes [8–11], yet to date there are no effective behavioral interventions that successfully target the full range of substance use that is most problematic among this subpopulation of college students. In response to this void, our team developed *myPlaybook,* an online substance use prevention program that is specifically tailored to college student-athletes. The reach, appeal, and economy of online programs make them an ideal option for college athletic departments; however, these programs typically have not demonstrated the same levels of effectiveness as facilitator-led interventions. We conducted an initial pilot study of *myPlaybook* using the classical "treatment package" approach, in which the entire program was delivered and evaluated as a single "package." The results of this study did not provide us with clear guidance about how we could improve *myPlaybook*'s effectiveness. Thus, we decided to apply MOST with the objective of improving *myPlaybook* so that its effectiveness would approach that of facilitator-led interventions.

The preparation phase of MOST

The first step within the preparation phase is to develop a conceptual model. Substance use prevention programs often are based on multiple behavioral theories so as to target an array of modifiable risk and protective factors [12]. Before developing *myPlaybook*, we drew on Social Norms Theory [13, 14], the Health Belief Model [15, 16], and the Theory of Reasone Action [17, 18] to identify several modifiable risk and protective factors that are associated with alcohol and other drug use among college student-athletes. According to our model (Fig. 1), social norms about peer substance use, positive and negative expectancies about the effects and consequences of substance use, and self-efficacy to prevent alcohol-related harm influence behavioral intentions to resist use and prevent harm. Behavioral intentions lead to engaging in (or avoiding) substance use; engaging in substance use can in turn lead to negative consequences (e.g., impaired performance). Behavioral intentions to prevent harm can also impact consequences directly (e.g., a student does not change alcohol use habits, but uses strategies such as not driving after drinking to prevent harm).
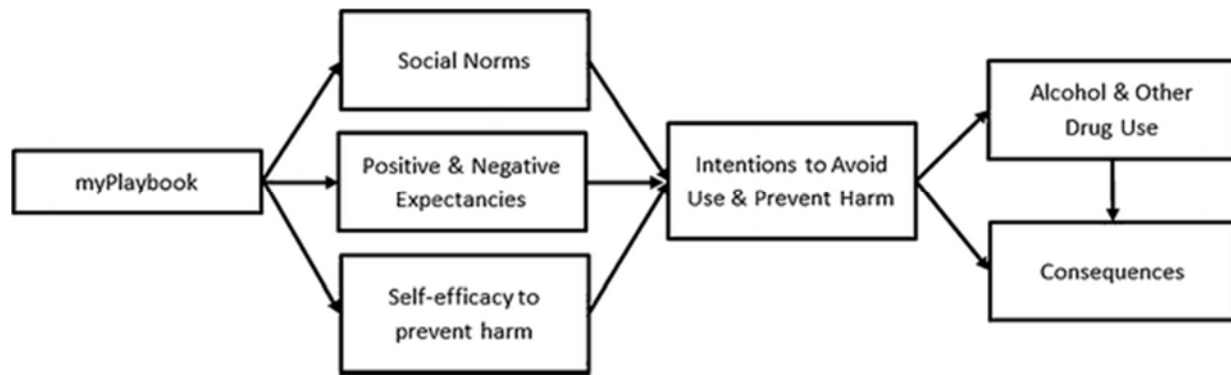
Fig 1 The *myPlaybook* conceptual model. This model guided various decisions that we made during the development of *myPlaybook* and our application of MOST

Based on our conceptual model, we developed *myPlaybook* to target each of the risk and protective factors depicted in Fig. 1. The original version of *myPlaybook* consisted of six lessons. Lesson (1) provided informational content about the NCAA rules around banned substances and drug testing. The remaining five lessons—(2) Alcohol, (3) Marijuana, (4) Tobacco, (5) Performance-Enhancing Drugs and Nutritional and Dietary Supplements, and (6) Prescription and Over-the-Counter Drugs—each targeted the three risk and protective factors as they pertained to a specific substance. We hypothesized that changes in these risk and protective factors would mediate the association between *myPlaybook* and substance use outcomes. Given their role in the conceptual model for *myPlaybook*, below we refer to each of these risk and protective factors as mediating variables.

The next step within the Preparation phase is to identify the specific components that will be evaluated. There were many components that we could have selected, but we decided to treat each of the six existing*myPlaybook* lessons as independent components. Before making this decision, we considered reorganizing the *myPlaybook* lessons so that each lesson targeted one of the three mediating variables (i.e., social norms, expectancies, and self-efficacy) across multiple substances and using these reorganized lessons as the components. We also considered using other potential components, such as different instructional design tools/strategies (e.g., quizzes vs. interactive flash animations) and features of the online delivery system (e.g., self-enrollment in *myPlaybook* vs. enrollment by the athletics department). In the end, we decided to use the lessons as the components because we believed that the most critical feature of *myPlaybook* was delivering lessons capable of changing the targeted mediating variables. We also decided to use the existing lessons as the components, rather than revise the lessons, because we needed some guidance as to which lessons were already effective at changing behavior and which mediating variables were already changed by content within the lessons. We developed a series of questions (described in the next section) to provide this guidance.

The third step within the Preparation phase is to select the desired optimization criterion. Different optimization criteria can be used for different interventions. In the current study, we

operationally defined optimization as the largest effect size that can be achieved after two cycles of testing and revision. In other words, our goal is to maximize the public health impact of *myPlaybook* given the limited time and financial resources that we have available to conduct the component selection experiments. We considered streamlining the intervention by dropping ineffective lessons, but decided against this optimization criterion because each *myPlaybook* lesson contains critical information that is required by the NCAA. Other optimization criteria that we considered included the most effective intervention that can be delivered without exceeding a prespecified per person cost and the amount of time to deliver the program. We did not select these criteria because *myPlaybook* is delivered online (thus keeping the relative cost per person very low) and it is already a relatively brief intervention (the full program takes students less than 2 h to complete).

The optimization phase of MOST

Because our objective is to improve the effectiveness of *myPlaybook*, we decided to take a systematic and incremental approach by using a novel iterative procedure in the Optimization phase of MOST. The procedure involves alternating highly efficient experiments to evaluate all of the components with revisions of any components that do not perform as expected. Below, we describe the process that we used in the first component selection experiment. Our goal was to determine whether each component (i.e., lesson) was achieving a specific effect size, and if not, to determine whether some or all aspects of the lesson (i.e., content within the lesson targeting specific mediating variables) needed to be revised. During this first component selection experiment, we collected data on behavioral and mediating variables at baseline (pretest), immediately after students completed *myPlaybook* (immediate posttest), and 30 days after they completed *myPlaybook* (30-day follow-up). We decided which components required revision by answering the following series of questions about each component:

Question 1

Did the component achieve an effect size of $d \geq 0.3$ for the targeted behavioral outcome at 30-day follow-up? (e.g., did the alcohol lesson impact 30-day alcohol use?) We selected an effect size of $d \geq 0.3$ as the standard for component effectiveness because it represents a clinically meaningful reduction in substance use and is comparable to effects observed in evaluations of similar online behavioral interventions [19].

- If yes: We will not revise this component. Although we still could try to improve this component, such revisions could potentially *weaken* rather than improve the component. Furthermore, any resources that we spent revising an effective component would detract from resources we could use to improve ineffective components. Therefore, we will not revise any effective components, but we will include them as is in subsequent component selection experiments to try to replicate our initial results.

- If no: Move on to Question 2.

Question 2

Does the component achieve an effect size of $d > 0.4$ for each hypothesized mediating variable within that lesson (i.e., social norms, positive and negative expectancies, and harm prevention strategies) at the 30-day follow-up? We set a higher bar for concluding that a component has effectively changed the mediating variables, because we expected that stronger effects on the mediating variables would be needed to translate these effects into behavioral change.

- If yes: If there is an effect on one or more of the mediating variables without a corresponding effect on the behavioral outcome, this suggests that either (1) it takes longer than 30 days for a change in the mediating variable(s) to translate into a change in behavior or (2) the proposed "mediating variable(s)" are not causally related to the outcome (i.e., the conceptual model is wrong). We cannot disentangle these alternatives within the first component selection experiment, which only included a 30-day follow-up. Therefore, if a component achieves an effect size of $d \geq 0.4$ for a specific mediating variable, we will extend the timing of follow-up in the next component selection experiment and revisit the conceptual model to determine if revision of the content specific to that mediating variable is necessary. For example, if the alcohol lesson has an effect size of $d = 0.5$ on social norms, $d = 0.2$ on expectancies, and $d = 0.1$ on harm prevention, we will extend the follow-up in the next component selection experiment to longer than 30 days and revise the content within the alcohol lesson that addresses expectancies and harm prevention (the mediating variables that did not achieve $d \geq 0.4$).

- If no: Move on to Question 3.

Question 3

Does the component achieve an effect size of $d > 0.4$ for each hypothesized mediating variable at the immediate posttest?

- If yes: If there is an effect on the mediating variable at the immediate posttest (but no effect at the 30-day follow-up), this suggests that the initial effect on that mediating variable decays over time. Therefore, we will develop a booster session during the revision process to help sustain the effect on that particular mediating variable.

- If no: For any mediating variables that do not achieve the targeted effect size at the immediate posttest, we will revise the content of the component targeting that mediating variable.

After revisions are completed, all components will be evaluated in a second randomized component selection experiment. We will then revise any remaining components that still did not achieve the specified effect sizes, and conduct a final experiment to evaluate component effects. Then, we will select the best version of each component to comprise the optimized intervention.

The evaluation phase of MOST

We will evaluate the optimized intervention, *myPlaybook Beta*, using a RCT. Instead of a no-treatment or wait-list control, we plan to use the original version of *myPlaybook* as our comparison group in the RCT. Comparing a new treatment against the current "standard of excellence" is common in clinical trials. This approach not only addresses any ethnical concerns about withholding a potentially effective intervention from the control group participants, it also provides a rigorous test of whether the optimized version of *myPlaybook* is a significant improvement over the original version of *myPlaybook*.

IMPLEMENTATION QUESTIONS

In this section, we identify several questions that intervention scientists often ask when they are considering whether to implement MOST. We then describe how we addressed each of these questions as we used MOST to optimize *myPlaybook*.

Can MOST be used if the intervention has already been developed and/or tested?

MOST can be used either to develop an intervention "from scratch" or to improve an existing intervention. We initially developed and delivered *myPlaybook* as a single intervention "package" outside of the MOST framework. Thus we already had developed the conceptual model described above when we first began considering the use of MOST. The primary challenge of starting MOST at this stage is that we had to decide how to break an existing intervention package into several components that could be separated and tested, instead of designing *myPlaybook* with this goal in mind from the beginning.

What if the intervention has essential content that all participants must receive?

Because MOST involves estimating the effect of separate intervention components, intervention scientists who use MOST must consider whether and how the components can be studied independently. Identifying independent components may be fairly straightforward when the components pertain to, say, mode of delivery rather than the content of the intervention itself. The process may be more challenging when the components include program content and some of the content is essential for all participants. For example, one of the six *myPlaybook* lessons introduces information that is required of all NCAA student-athletes (i.e., information about banned substances and drug testing procedures). We decided to provide this foundational lesson to *all* study participants regardless of the condition they were assigned. Therefore, we only manipulated five of the six lessons in our component selection experiments.

Does the optimization phase of MOST require prohibitively large numbers of participants?

If an efficient experimental design is used, the Optimization Phase does not require unusually large sample sizes. To make economical use of research participants, Collins et al. [20] recommended using a factorial experimental design during the component selection experiment phase of MOST, with each intervention component treated as a factor. These designs require many fewer participants than, for example, conducting individual experiments on each component [20]. In addition, factorial experiments enable intervention scientists to test for interactions among components (e.g., whether one component works better when combined with another component) which could not be done if each of the components were tested in separate experiments [20]. The iterative approach in the current study involves three experiments, each of which requires its own sample, so the total across the three experiments does amount to a larger sample than would be required by most single RCTs.

Is the number of experimental conditions required by a factorial experiment prohibitive?

Using a factorial design allows us to efficiently use participants, but it can lead to another challenge: requiring a large number of experimental conditions. For example, in our first component selection experiment, we needed to test five components. A complete factorial design would have required $2^5 = 32$ different experimental conditions. Because *myPlaybook* is delivered online, implementing 32 conditions would have been a manageable computer programming task. However, we wanted to stratify schools by division (i.e., levels of intercollegiate athletics within the NCAA), so that every condition had at least one school from each of the three divisions. With 32 conditions, we would have needed 96 schools, whereas our power analysis (described in more detail below) indicated that we only needed 56 schools. Therefore, we decided to use a fractional factorial design [21–23]. This design requires fewer experimental conditions than a complete factorial design, without changing the number of participants required. The fractional factorial design we selected cuts the number of conditions in half, from 32 to 16. Table 1 lists the conditions in our experimental design. In each condition, each component was either included ("On") or excluded ("Off"). Although implementing 16 conditions was feasible for our study of an online intervention, in other situations, resource limitations may mean that intervention scientists need to either use a fractional factorial design that requires even fewer conditions or evaluate fewer than five components.

Table 1 Conditions in fractional factorial experimental design used in component selection experiment

Condition number	Introduction	Manipulated intervention components				
		Alcohol	Tobacco	Marijuana	Performance-enhancing substances	Prescription and OTC drugs

1	On	Off	Off	Off	Off	On
2	On	Off	Off	Off	On	Off
3	On	Off	Off	On	Off	Off
4	On	Off	Off	On	On	On
5	On	Off	On	Off	Off	Off
6	On	Off	On	Off	On	On
7	On	Off	On	On	Off	On
8	On	Off	On	On	On	Off
9	On	On	Off	Off	Off	Off
10	On	On	Off	Off	On	On
11	On	On	Off	On	Off	On
12	On	On	Off	On	On	Off
13	On	On	On	Off	Off	On
14	On	On	On	Off	On	Off
15	On	On	On	On	Off	Off
16	On	On	On	On	On	On

One drawback of using a fractional factorial design is that whenever conditions are removed, certain effects cannot be disentangled from each other. It is possible to strategically select a fractional factorial design so that effects of key scientific interest are combined with effects that are expected to be negligible in size [21]. For example, we were comfortable assuming that any interactions involving three or more components would be small in size. Therefore, we selected a fractional factorial design that combines main effects with four-way interactions, and combines two-way interactions with three-way interactions. Because our decisions about component effectiveness will be based primarily on main effects, we believe that this is an acceptable trade-off for the economy afforded by eliminating 16 conditions per experiment.

What if cluster randomization is necessary?

Like many substance use prevention programs, *myPlaybook* targets the individual. However, in our study, student-athletes were clustered within colleges and universities, so we had to decide whether to randomize at the student-athlete level or at the school level. Given that athletes may spend considerable time with each other, we decided to randomize at the school level to reduce the risk of contagion between student-athletes in different conditions. We initially had several concerns about the feasibility of using a cluster-randomized design in the Optimization phase of MOST.

Power analysis

First, we were concerned about the prospect of conducting a power analysis for a cluster-randomized fractional factorial design. Fortunately, Dziak et al. [24] have developed a macro [25] that can be used to carry out the calculations for a power analysis exactly for this situation. In our case, we were limited in the number of freshmen student-athletes who would be available at any given school (approximately 75–150 student-athletes at the start of the school year), so we needed to determine how many schools we needed to recruit for each component selection experiment. Research conducted on another online college alcohol prevention program reported survey response rates at 1-month follow-up to be around 80 % [26]. Assuming an average of 100 student-athletes per school and a response rate of 80 % at 30-day follow-up and using the formulas provided by Dziak et al. [24], we determined that we needed 56 schools to have 90 % power with a two-tailed $\alpha = 0.05$ to detect a behavioral effect size of $d = 0.3$. Note that as with any power analysis, we selected values based on a combination of existing research and our goals for the current project. Therefore, other researchers may select different values for their analyses (e.g., we conducted our power analysis using 90 % power to reduce the probability of a Type II error, but others could use a lower value, such as the more standard value of 80 % power, for their calculations).

Recruitment

Once we knew that we needed 56 schools, our second concern was the challenge of recruiting that many schools. We recruited nationally by sending email blasts through the NCAA and various listservs to promote informational webinars. The webinars allowed us to promote the study and share expectations for participating campuses. Not only was this approach to recruitment efficient in terms of time and money, but it also creates a sustainable avenue for future project specific training as well as training outside of the research context. Clearly, feasibility of implementing *myPlaybook* at so many schools was facilitated by the fact that it is delivered online, however even when an intervention is not delivered online, with careful planning and supervision it is possible to conduct a factorial experiment in the field. For example, Collins et al. [3] implemented 32 experimental conditions in health care clinics. In addition, intervention scientists could consider randomizing the intervention at an intermediate level (in classrooms or to specific teams) rather than at the school level.

Analyses

Our third concern was how to analyze data from a cluster-randomized factorial experiment in which the unit of inference is the individual. However, it was relatively straightforward to analyze our data using a multi-level modeling framework within PROC MIXED in SAS.

What if the primary outcome must be measured months or years after the intervention is delivered?

Some interventions target outcomes that may take time to develop or change, such as substance use, or outcomes that may be hard to capture within a shorter time frame, such as getting a sexually transmitted infection. In our study, we needed to fit an entire experiment—from data collection through component revision—within the span of a single year, to be ready for the next experiment. Thus, we could not follow students long enough to measure substance use for an extended period of time. Because *myPlaybook* is designed to operate by affecting the mediating variables shown in Fig. 1, we decided to use information about the effect of each component on both the behavioral outcomes and the hypothesized mediating variables to determine whether a component was effective. This approach enables us to optimize the intervention within the time frame of a grant funding cycle. Questions about whether it takes time for the immediate effects of*myPlaybook* to translate into behavioral effects or whether the effects of *myPlaybook* are sustained over longer periods of time will be addressed in the Evaluation Phase of MOST, when an RCT of the optimized intervention is conducted.

What does MOST add that cannot be obtained via a process evaluation?

Intervention scientists are very accustomed to using multiple sources of data to improve their programs. For example, process evaluation can answer questions about how the program was implemented, the number of participants served, dropout rates, and how participants experienced the program. Although process measures are an essential data source for understanding how various pieces of an intervention may work, judgments based on post hoc non-experimental analyses are not sufficient to answer questions about the individual contributions of each component (e.g., whether each component is effective, whether all of the components are even needed, or whether each component is as effective as it could be). By contrast, MOST uses fully randomized and powered experiments to answer questions about the effects of each component.

However, process evaluation can be an important complement to MOST. In the study described here, we are conducting process evaluations to help us interpret the results from our component selection experiments and to guide revisions of *myPlaybook*. For example, we will use electronic data recorded as participants complete their assigned *myPlaybook* components to determine how long it takes student-athletes to finish each online component and whether they are learning the information presented. Problems in these domains may suggest that the content and strategies do not engage the student or that the instructions are unclear. We will assess students' impressions gathered from all participants as part of the follow-up survey and from a random selection of

participants by means of focus groups. We also have established an Expert Advisory Panel to review each component that "needs revision" and provide feedback about how consistent each component is with the latest science.

Is it possible to secure external funds to implement MOST?

The optimization of *myPlaybook* is being funded with a Small Business Innovative Research (SBIR) grant (R44DA023735) through the National Institute on Drug Abuse (NIDA). At this writing, other applications of MOST have been funded by NIDA (R01DA029084), the National Cancer Institute (P50CA143188; P50CA101451; R01CA138598), the National Institute of Diabetes and Digestive and Kidney Diseases(R01DK097364), and the National Heart, Lung, and Blood Institute (R01HL113272). A related question is whether it is feasible to complete all three phases of MOST during a single 5-year funding cycle. This depends on considerations such as how much of the Preparation Phase has been completed by the start of the funding cycle; the time frame for completion of intervention delivery; how rapidly the required number of experimental participants can be recruited; and the time lag between intervention participation and the measurement of outcomes. Other intervention scientists may consider requesting funding to complete the Optimization Phase, followed by a separate application to complete the Evaluation Phase.

DISCUSSION

Slow adoption of innovations is not uncommon [27], as people consider questions about both the value of the innovation and challenges associated with implementation. Left unaddressed, such questions can pose a threat to scientific progress by slowing the adoption of potentially paradigm-shifting innovations. In this article, we described how we addressed these questions to demonstrate that it was feasible to use MOST to optimize *myPlaybook*. We believe that if intervention scientists give careful consideration to addressing questions such as those described and find ways to overcome any challenges, it is likely that they will conclude that it is feasible to use MOST.

Our experience thus far has led us to identify several benefits of applying MOST to optimize *myPlaybook*. First, MOST efficiently uses scarce resources such as money and participants. Second, MOST enables us to identify which specific components need to be refined based on experimental evidence rather than extracted from non-experimental post hoc analyses. Finally, using MOST allows us to strive for a greater public health impact rather than settling on statistical significance alone.

In this article, we described an iterative approach to improving *myPlaybook*. The alternative would be to conduct a treatment package RCT followed by post hoc analyses, revising the treatment package, conducting another RCT, and then repeating the cycle. Although this process has the advantage of being familiar, it takes longer than the approach described here. Moreover, post hoc analyses based on data from an RCT provide less definitive evidence about the

effectiveness of individual components than the factorial experiments used in MOST. Therefore, these analyses give intervention scientists less information to guide their decisions about revising the intervention.

Ideally, an iterative process of intervention improvement like the one we are using in the Optimization Phase would be repeated as many times as necessary to achieve a complete set of highly effective components. Realistically, resource and time limitations will always constrain the investigation to some number of cycles. It would be helpful if, when applying for funding, it were acceptable to build some flexibility into the research plan that would empower the investigator to change the direction of the research plan based on the results of the work to date. For example, a research plan could call for making a decision after two cycles of experimentation/revision/experimentation whether (a) another cycle of revision is needed, or (b) it is time to move to the Evaluation Phase of MOST and evaluate he treatment package by means of an RCT. Although NIH study sections generally do not support the inclusion of such flexibility in research proposals, in our view this flexibility would improve the use of research resources and allow the field of intervention science to move forward faster.

CONCLUSIONS

We believe that the time is right to move beyond the treatment package approach and begin a paradigm shift in methods for the development, optimization, and evaluation of behavioral interventions. MOST provides a step-by-step blueprint for intervention scientists to rethink "business as usual" and achieve this goal. The engineering principles that are at the core of MOST provide much-needed guidance on how to move from early stage development and testing to evaluation of an optimized behavioral intervention. Such a systematic and principled approach is needed to ensure the efficient use of limited resourced and to maximize the public health impact of behavioral interventions.

References

1. Collins LM, Murphy SA, Nair VN, Strecher VJ. A strategy for optimizing and evaluating behavioral interventions. *Ann Behav Med*. 2005; 30: 65-73.

2. Collins LM, Murphy SA, Strecher VJ. 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-S118.

3. Collins LM, Baker TB, Mermelstein RJ, et al. The multiphase optimization strategy for engineering effective tobacco use interventions. *Ann Beh Med*. 2011; 41(2): 208-226.

4. Collins LM, Chakraborty B, Murphy SA, Strecher V. Comparison of a phased experimental approach and a single randomized clinical. *Clin Trials*. 2009; 6(1): 5-15.

5. Strecher VJ, McClure JB, Alexander GL, et al. Web-based smoking-cessation programs: results of a randomized trial. *Am J Prev Med*. 2008; 34(5): 373-381.

6. Caldwell LL, Smith EA, Collins LM, et al. Translational research in South Africa: evaluating implementation quality using a factorial design. *Child Youth Care Forum*. 2012; 41: 119-136.

7. McClure J, Derry H, Riggs K, et al. Questions about quitting (Q2): design and methods of a multiphase optimization strategy (MOST) randomized screening experiment for an online, motivational smoking cessation intervention. *Contemp Clin Trials*. 2012; 33(5): 1094-1102.

8. Doumas DM, Haustveit T, Coll KM. Reducing heavy drinking among first year intercollegiate athletes: a randomized controlled trial of web-based normative feedback. *J Appl Sport Psych*. 2010; 22(3): 247-261.

9. Turrisi R, Larimer M, Mallett K, et al. A randomized clinical trial evaluating a combined alcohol intervention for high-risk college students. *J Stud Alcohol Drugs*. 2009; 70(4): 555-567.

10. Yusko DA, Buckman JF, White HR, Pandina RJ. Alcohol, tobacco, illicit drugs, and performance enhancers: a comparison of use by college student athletes and nonathletes. *J Am Coll Health*. 2008; 57(3): 281-290.

11. Yusko DA, Buckman JF, White HR, Pandina RJ. Risk for excessive alcohol use and drinking-related problems in college student athletes. *Addict Beh*. 2008; 33(12): 1546-1556.

12. Hansen W, Dusenbury L, Bishop D, Derzon J. Substance abuse prevention program content: systematizing the classification of what programs target for change. *Health Educ Res*. 2007; 22(3): 351-360.

13. Berkowitz AD. An overview of the social norms approach. In: Lederman LC, Stewar LP, eds. *Changing the culture of college drinking*. Cresskill: Hampton; 2004: 193-214.

14. Perkins HW. The emergence and evolution of the social norms approach to substance abuse prevention. In: Perkins HW, ed. *The social norms approach to preventing school and college age substance abuse: a handbook for educators, counselors, and clinicians*. San Francisco: Jossey-Bass; 2003: 3-18.

15. Becker MH, Radius SM, Rosenstock IM, Drachman RH, Schuberth KC, Teets KC. Compliance with a medical regimen for asthma: a test of the health belief model. *Public Health Rep*. 1978; 93(3): 268-277.

16. Rosenstock IM. Historical origins of the health belief model. *Health Educ Behav*. 1974; 2(4): 328-335.

17. Ajzen I, Fishbein M. *Understanding attitudes and predicting social behavior*, vol. 278. Englewood Cliffs, New Jersey: Prentice-Hall; 1980.

18. Fishbein M, Ajzen I. *Belief, attitude, intention, and behavior: an introduction to theory and research*. Reading: Addison-Wesley; 1975.

19. Neighbors C, Lewis MA, Atkins DC, et al. Efficacy of web-based personalized normative feedback: a two-year randomized controlled trial. *J Consult Clin Psych*. 2010; 78(6): 898.

20. Collins LM, Dziak JJ, Li R. Design of experiments with multiple independent variables: a resource management perspective on complete and reduced factorial designs. *Psych Methods*. 2009; 14(3): 202-224.

21. Wu C-F, Hamada M. *Experiments: planning, analysis, and parameter design optimization*. New York: Wiley; 2000.

22. Chakraborty B, Collins LM, Strecher VJ, Murphy SA. Developing multicomponent interventions using fractional factorial designs. *Stat Med*. 2009; 28(21): 2687-2708.

23. Nair V, Strecher V, Fagerlin A, et al. Screening experiments and the use of fractional factorial designs in behavioral intervention research. *Am J Pub Health*. 2008; 98(8): 1354-1359.

24. Dziak JJ, Nahum-Shani I, Collins LM. Multilevel factorial experiments for developing behavioral interventions: power, sample size, and resource considerations. *Psych Methods*. 2012; 17(2): 153-175.

25. FactorialPowerPlan [computer program]. Version 1.0. University Park: The Methodology Center, The Pennsylvania State University; 2013.

26. Chiauzzi E, Green TC, Lord S, Thum C, Goldstein M. My student body: a high-risk drinking prevention web site for college students. *J Am Coll Health*. 2005; 53(6): 263-274.

27. Rogers EM. *Diffusion of innovations*. New York: Free Press; 2003.