

Optimization of Multicomponent Behavioral and Biobehavioral Interventions for the Prevention and Treatment of HIV/AIDS

Linda M. Collins, Kari C. Kugler & Marya Viorst Gwadz

AIDS and Behavior

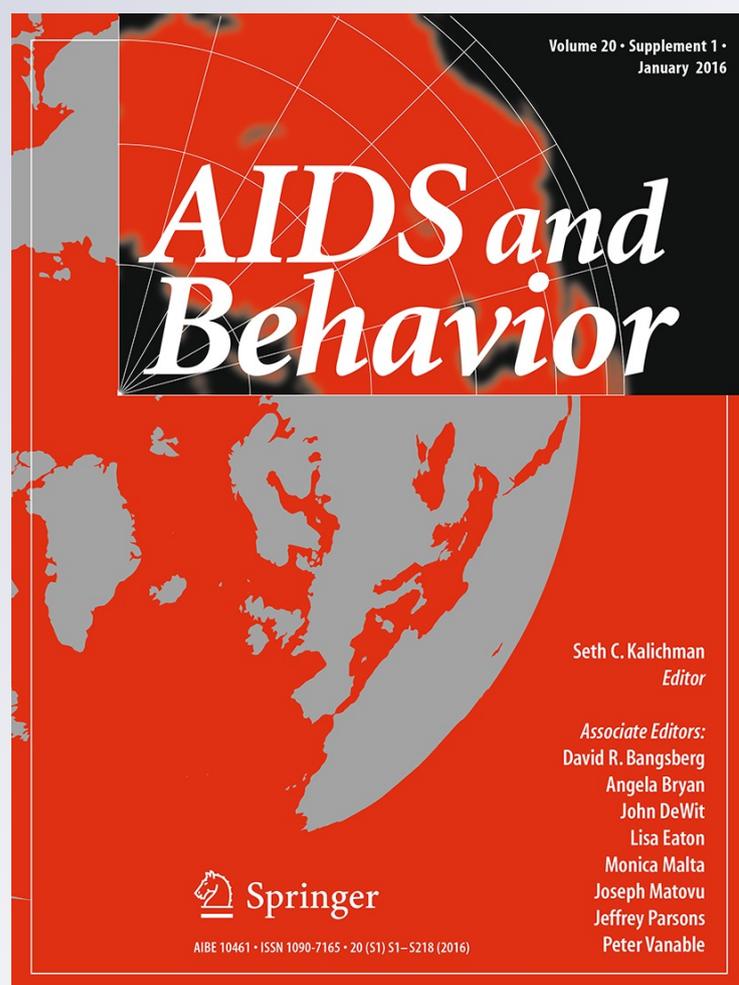
ISSN 1090-7165

Volume 20

Supplement 1

AIDS Behav (2016) 20:197-214

DOI 10.1007/s10461-015-1145-4



Your article is protected by copyright and all rights are held exclusively by Springer Science +Business Media New York. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".

Optimization of Multicomponent Behavioral and Biobehavioral Interventions for the Prevention and Treatment of HIV/AIDS

Linda M. Collins¹ · Kari C. Kugler² · Marya Viorst Gwadz³

Published online: 4 August 2015
© Springer Science+Business Media New York 2015

Abstract To move society toward an AIDS-free generation, behavioral interventions for prevention and treatment of HIV/AIDS must be not only effective, but also cost-effective, efficient, and readily scalable. The purpose of this article is to introduce to the HIV/AIDS research community the multiphase optimization strategy (MOST), a new methodological framework inspired by engineering principles and designed to develop behavioral interventions that have these important characteristics. Many behavioral interventions comprise multiple components. In MOST, randomized experimentation is conducted to assess the individual performance of each intervention component, and whether its presence/absence/setting has an impact on the performance of other components. This information is used to engineer an intervention that meets a specific optimization criterion, defined a priori in terms of effectiveness, cost, cost-effectiveness, and/or scalability. MOST will enable intervention science to develop a coherent knowledge base about what works and does not work. Ultimately this will improve behavioral interventions systematically and incrementally.

Keywords Behavioral intervention · Biobehavioral intervention · Multiphase optimization strategy · Factorial design · Fractional factorial design

Introduction

Overview

The purpose of this article is to introduce the multiphase optimization strategy (MOST) to the HIV/AIDS research community, first by describing the approach and then by providing an illustration of its application to intervention development [1–4]. Inspired by engineering principles, MOST provides a framework for developing, optimizing, and evaluating behavioral and biobehavioral interventions for the prevention and treatment of HIV/AIDS. The objective of MOST is to produce behavioral interventions that are effective, economical, efficient, and scalable—critical factors in an era of both serious public health concerns and constrained resources. MOST can be used to build new interventions, as we describe in the hypothetical example below, or to improve upon existing evidence-based programs. In this article, we will use the term “behavioral intervention” to refer broadly to any program designed to change individual behavior with the objective of preventing, treating, and/or adapting to HIV/AIDS. Thus our use of this term includes biobehavioral interventions that involve both biomedical and behavioral components.

The Role of Behavior and Behavioral Interventions in HIV/AIDS Prevention and Treatment

The epidemiology of HIV/AIDS is complex, and risk for HIV varies substantially by economic, political, and

✉ Linda M. Collins
lmcollins@psu.edu

¹ Department of Human Development and Family Studies and The Methodology Center, The Pennsylvania State University, 404 Health & Human Development Building, University Park, PA 16802, USA

² The Methodology Center, The Pennsylvania State University, University Park, PA, USA

³ Center for Drug Use and HIV Research (CDUHR), College of Nursing, New York University, New York, NY, USA

environmental structural factors such as housing, unemployment or incarceration rates, and/or service quality and availability [5]; social factors such as social networks, norms, and stigma [6, 7]; and individual risk behavior [24], with the relative importance of each factor varying by population and local context [8–10]. This multi-faceted risk profile is reflected in the national portfolio of evidence-based interventions that address structural, social, or individual risk factors or risk environments, sometimes taking a multi-level or combination approach [11, 12].

To provide a succinct illustration of the application of MOST, the present article focuses on behavioral interventions to reduce individual-level risk, although we wish to emphasize that MOST can be used to develop and evaluate structural interventions. There is no question that structural interventions have great potential to be potent and cost-effective solutions to HIV-related public health problems [11, 12]. At the same time, there is continued recognition of the need for efficient behavioral interventions to address individual factors in many contexts and at multiple levels [13]. Indeed, even in the context of structural and social drivers of risk, each case of HIV in a population results from individual behavior, mainly unprotected sex and/or the sharing of injection drug use paraphernalia with someone who is infected with HIV [14, 15]. Individuals in both high- and low-risk contexts can reduce their sexual risk of contracting HIV by being abstinent, selecting HIV-uninfected partners, being monogamous with an HIV-uninfected partner, engaging in safer sex practices (e.g., always using a condom), and/or adhering to a pharmaceutical pre-exposure prophylaxis (PrEP) regimen. Injection-related risk can be eliminated by abstaining from injection drug use or avoiding shared injection paraphernalia [14]. People living with HIV/AIDS (PLWHA) can reduce the possibility of transmitting HIV to others through awareness of their own serostatus, adhering to antiretroviral therapy (ART) to bring HIV viral load to undetectable levels, and avoiding sexual and drug injection behaviors that may place others at risk [14, 16, 17].

The goal of some behavioral interventions is modification of behaviors that directly affect the risk of HIV acquisition. For example, an intervention might aim to decrease the possibility of HIV transmission through earlier diagnosis of HIV infection by using a peer-driven approach to seek out and test individuals at high risk for undiagnosed HIV infection and linking HIV-infected individuals to HIV primary care [18]. The goal of other interventions is modification of risk behaviors that can lead to HIV acquisition and transmission. Alcohol use is one example of a risk behavior that is particularly salient in the prevention and treatment of HIV. Alcohol use is highly prevalent among populations at risk for and living with HIV [19] and is associated with incident HIV infection [19,

20]. Although the relationship between alcohol and HIV risk behavior varies by partner type and context [21, 22], it is well established that alcohol use can contribute to sexual risk behavior [23]. This effect begins with even small amounts of alcohol and increases with the quantity of alcohol consumed [20]. Further, it has been shown that the use of alcohol and other substances is related to inadequate adherence to both PrEP and ART [24–26], which is alarming because strict adherence to these pharmaceutical protocols is required in order for them to be effective [27, 28]. At the same time, treatment as prevention (TasP) requires very high levels of sustained adherence to maintain low viral loads and reduce the probability of forward transmission of HIV [17, 29]. Finally, even moderate alcohol use is contraindicated for PLWHA on ART [30], because in addition to its negative effects on ART adherence, alcohol use is associated with increased viral load [31] and poor health outcomes [24]. Alcohol-focused interventions aimed at reducing HIV risk have, for example, targeted heavy episodic drinking among college students [32] and used motivational interviewing and cognitive-behavioral skills training to improve ART adherence among hazardous drinkers [33].

The Need for a New Approach to Development of Behavioral Interventions

Behavioral interventions must be highly effective if they are to move society toward an AIDS-free generation [34]. However, effectiveness alone, although critical, is not sufficient. To be most useful to society, interventions must also be economical, efficient, and readily scalable. By scalable, we mean interventions must be implementable with fidelity in real-world settings without the need to reduce the intervention's length, complexity, or expense by making ad hoc modifications that will have an unknown impact on effectiveness. Moreover, maximal progress will be made toward ending the global HIV/AIDS pandemic only when research continually builds on prior results to produce incrementally, materially, and demonstrably better and better behavioral interventions over time; that is, when standards for effectiveness, economy, efficiency, and scalability are continually raised.

However, for several reasons it is difficult to produce increasingly effective, economical, efficient, and scalable interventions by conducting research that relies *solely* on the randomized controlled trial (RCT). First, the vast majority of behavioral interventions comprise multiple components, but an RCT that yields significant results in the desired direction indicates only that the intervention package *as a whole* has had a positive effect. The RCT cannot reveal which specific individual components of the intervention are having a positive effect on the outcome or which are not contributing

much and could be removed. Similarly, a non-significant RCT result (or a significant iatrogenic result) indicates that the intervention package as a whole does not have a positive effect, but this result cannot reveal whether any of the individual components may be worth retaining. It is even possible that one or more iatrogenic components could be offsetting the positive effects of others.

Second, a significant RCT result does not enable estimation of the size of the effects of individual components, so it is impossible to evaluate whether a particular component's effect is large enough to justify its cost. For example, many of today's behavioral interventions include a motivational interviewing component [35], which can be costly because it requires trained personnel and is generally time-consuming to implement. Thus an investigator may wonder whether the motivational interviewing component contributes enough to this intervention to make it worth the cost. This question can be answered only if an estimate of the individual effect of the motivational interviewing component can be obtained.

Third, an RCT does not enable examination of whether the presence of one component enhances or weakens the effect of another component. For example, the motivational interviewing component may enhance the effect of another component, such as cognitive-behavioral skills training, if it helps motivate participants to apply the skills they learn. The impact of one component on the effect of another is represented statistically by an interaction. Interactions between intervention components cannot be estimated with a two-arm RCT of a multi-component intervention.

We are emphatically not suggesting abandonment of the RCT, which we believe will always play a key role in evaluation of behavioral interventions. Instead, we advocate the use of an approach that capitalizes on cutting-edge intervention science methodology and includes additional experimental designs along with the RCT.

The Multiphase Optimization Strategy (MOST)

MOST consists of three phases: preparation, optimization, and evaluation. The preparation and evaluation phases involve many activities that are also a part of the classical process in which an intervention is developed and then immediately evaluated with an RCT; we refer to this as the "treatment package approach." However, MOST introduces an additional phase into the intervention development process: the optimization phase. In this phase, efficient randomized experimentation is conducted to gather information about the individual performance of each intervention component and whether the presence or absence of a component has an impact on the performance of other components. This information is then used to engineer the intervention to meet a specific optimization

criterion, defined a priori in terms of effectiveness, economy, efficiency, and/or scalability. For example, the optimization criterion might call for selecting the subset of components that produces the most cost-effective intervention, or the most effective intervention that does not exceed a certain cost (e.g., \$400 per person) or duration (e.g., an intervention under 5 h). Whatever the optimization criterion, using MOST it is always possible to aim for an intervention made up solely of components with empirically demonstrated effectiveness,¹ with no inactive or counterproductive elements. In the evaluation phase of MOST, the intervention's overall effectiveness is assessed against an appropriate control via an RCT.

Hypothetical Example: ART Adherence Among Hazardous Drinkers

For illustrative purposes, throughout this article we will use a hypothetical example of a multi-component behavioral intervention aimed at improving ART adherence among PWLHA who use alcohol at hazardous levels. The ultimate objective of this hypothetical intervention is HIV viral load suppression. Figure 1 depicts a simple (again, hypothetical) conceptual model of the proximal predictors of alcohol use and/or adherence to ART among those with problem drinking. It also shows the individual components to be potentially included in the intervention and which proximal predictor each component is intended to influence. In other words, Fig. 1 depicts the primary mediation pathway for each component, showing how it is hypothesized to contribute to reduced HIV viral load, the intervention's primary outcome variable.

Conceptual Model

The hypothetical conceptual model is grounded in the theory of planned behavior [36], self-determination theory [37], and general social-cognitive theory [38, 39]. Social-cognitive theory conceptualizes the behavior change process as an ongoing and dynamic interaction among individual/cognitive (e.g., normative health beliefs, mental health status, and behavioral skills), social (e.g., positive role models, social support), and environmental factors (which may moderate intervention effects). Next, the theory of planned behavior links health beliefs and social factors with behavior through the critical pathway of intentions [36]. Last, self-determination theory highlights the importance of durable, intrinsic motivation for behavior

¹ To save space, we will use the word "effectiveness" to refer to either efficacy or effectiveness where it is not necessary to distinguish between the two.

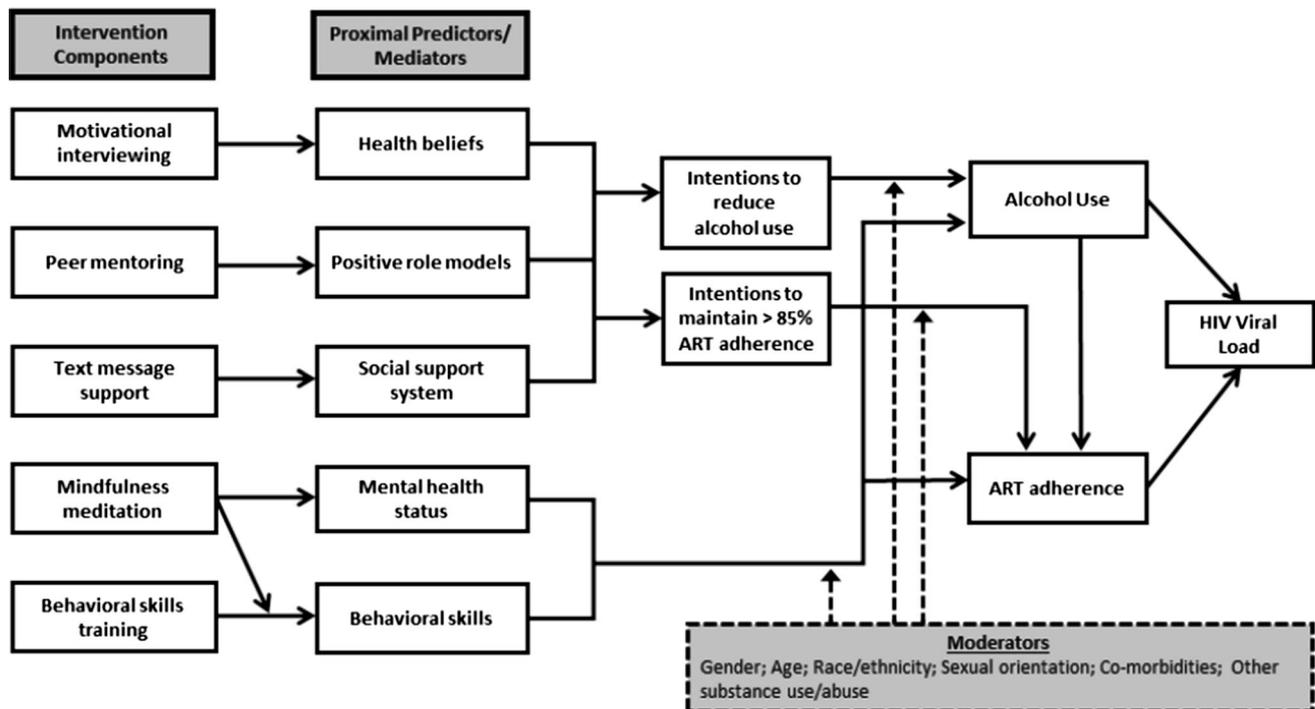


Fig. 1 Conceptual model of alcohol use and ART adherence among persons living with HIV/AIDS

change and is closely linked to normative health beliefs and intentions [37]. Consistent with this integrated conceptual model, the counseling approach used in the intervention components integrates cognitive-behavioral techniques to influence attitudes, foster interactions with role models, and build behavioral skills to influence intentions and behavior, along with motivational interviewing to trigger or enhance durable, high quality motivation for behavior change [35, 40].

In this model, both individual-level and social-level factors influence alcohol use patterns, ART adherence rates, and ultimately HIV viral load levels. First, *health beliefs* [38] are hypothesized to affect intentions to reduce alcohol use and intentions to improve ART adherence [36]. For example, health beliefs about possible toxic interactions between alcohol and ART have been found to influence intentions to take ART while using alcohol and to be related to missed ART doses during drinking episodes [41]. Next, intentions to modify both problem alcohol use and inadequate ART adherence are improved by access to both *positive role models* [38, 42, 43], including PLWHA who successfully manage both their alcohol use and adherence to ART, and a strong *social support system* [44, 45]. Intentions to reduce alcohol use and to improve ART adherence are hypothesized to directly affect alcohol use and ART adherence behaviors, respectively.

Mental health status, that is, perceived stress, anxiety, and depressive symptoms, is a potentially modifiable

individual-level characteristic that directly influences problem alcohol use and adherence to ART. Research has shown, for example, that individuals with untreated depression are more likely to use alcohol [46] and less likely to adhere to ART than those without depression [47–49]. Further, *behavioral skills* [38] to refuse/reduce alcohol and manage ART adherence are also hypothesized to directly influence use of alcohol and ART adherence. There is evidence that individuals who have developed these behavioral skills are less likely to use alcohol and more likely to be adherent than those who have not [50].

In addition, factors such as gender, race/ethnicity, sexual orientation, co-morbidities, and substance use [51, 52] are expected to moderate the effect of intentions, mental health status, and behavioral skills on behavioral outcomes. We wish to note that this is a hypothetical model developed for illustrative purposes, and is not intended to include all possible constructs or inter-relationships among domains. Nonetheless, to be plausible and well-grounded in the empirical literature, we focus on the primary factors that drive the inter-related problems of hazardous drinking and poor ART adherence.

Hypothetical Intervention Components

The leftmost column of Fig. 1 shows five intervention components. Each is designed to affect one of the individual and social proximal factors (described above); in

turn, these proximal factors will affect intentions to reduce alcohol use and/or increase adherence to ART and, ultimately, reduce HIV viral load. In this example, all intervention components will be delivered by trained interventionists primarily over the telephone or through a program on smartphones.

Motivational interviewing sessions component will affect health beliefs by engaging the participant in the process of examining alcohol use patterns in the context of HIV infection and ART adherence; normative beliefs about alcohol toxicities and ART; perspectives on alcohol use and ART adherence [33]; and intentions to use alcohol and adhere to the ART regimen. These sessions will seek to increase positive outcome expectancies for ART and help the individual plan strategies to reduce alcohol use and improve ART adherence, if appropriate [35]. This intervention component will comprise three 1-h sessions guided by an interventionist and conducted over the phone.

The *peer mentoring* component is intended to provide participants with a peer who has been successfully managing HIV and can serve as a positive role model. This component will include weekly contact over 12 weeks with a trained and supervised peer who is living with HIV, has experienced alcohol problems in the past, is presently managing alcohol at a non-problem level, and has been taking ART with high adherence for at least 12 months. The peer mentor will provide his/her own “story” regarding alcohol and ART; provide practical tips for managing alcohol use and ART adherence, based on his/her personal experience; elicit the participant’s experiences and concerns; offer informal counseling; and provide encouragement to link to other needed services, including substance use and mental health treatment [53].

The *text message support* component uses short message service (SMS) technology to increase the participant’s experience of social support for gains made in improving alcohol reduction and/or ART adherence intentions, reducing alcohol use behavior, improving ART adherence behavior, or other relevant attitudinal or behavioral changes. Text messaging has the potential to accomplish this in an efficient manner as compared to one-on-one structured sessions. The text message component will be personalized (not automated), and a staff member will provide two types of text messages: query messages (a prompt to begin a brief text discussion with participants) and support messages (responding to participants’ messages with reflective listening and information support, tangible assistance, esteem support, network support, and emotional support, as appropriate). Query and support messages will be drawn from a “bank” of messages grounded in motivational interviewing and developed with a community advisory board. Messages will prompt the participant to reflect on

personal alcohol and ART goals, recent drinking, and ART adherence in the context of HIV infection, and will provide support and encouragement for gains made. Staff will also engage in non-scripted interaction with participants. Text message sessions will last 5–10 min. Staff will query participants at varying intervals over a 4-month period, approximately every other day, and participants will be encouraged to text their staff member at any time. The availability of frequent messaging with staff, as well as content of the messages, which will be focused on the various types of support found in past research to be salient for PLWHA, will help foster the experience of connection to a social support system [54–56].

It is expected that the vast majority of the sample will exhibit mental health distress in the form of perceived stress, anxiety, or depression at clinically significant levels as a result of the demands of coping with HIV diagnosis and ART, stigma (including sexual minority status), and, typically, low socio-economic status [48, 57–59]. The *mindfulness meditation training* component is intended to improve poor mental health status and reduce perceived stress. Over the course of three 1-h sessions, participants will be trained in meditation techniques, including meditation practice together with the intervention facilitator. They will also be provided with a workbook to guide home meditation activities [60].

The *behavioral skills training* component will be created using best practices in training to improve adherence skills. Items in this component will include identification of personal barriers to and facilitators of adherence, development of individualized adherence reminders, mapping the daily schedule plans to integrate ART into existing medication regimens, and receipt of a pill box or other adherence aid. This component will include an initial 1-h session conducted in person, at which time participants will be provided with a manual, followed by 6 weekly sessions conducted over the phone [61].

Figure 1 shows an arrow originating at mindfulness meditation and ending at the arrow joining behavioral skills training and the proximal predictor/mediator behavioral skills. This indicates that an interaction is hypothesized. We hypothesize that the presence of the mindfulness meditation component will increase the effect of the behavioral skills training component on behavioral skills, by increasing the individual’s receptivity to the training.

In addition to the five components listed above, the intervention also includes a single introductory *informational session*, lasting approximately 1 h. The purpose of the informational session is to begin engagement into the intervention and increase knowledge about the related problems of hazardous drinking and poor HIV outcomes. In this session participants will receive educational materials and referrals for alcohol problems, ART adherence,

and mental health treatment, as well as practical advice on how to maintain adherence and reduce drinking.

We acknowledge that the components described above, when combined into a single intervention, would result in a complex intervention of substantial duration. Yet an intervention of this nature might be necessary, given the multi-faceted and multi-level nature of the causes of hazardous alcohol use and poor ART adherence, and the inter-relationships between alcohol use and ART adherence. Further, many behavioral interventions are similarly comprised of numerous components [62]. This is particularly relevant in light of the growing awareness of the need for multi-level interventions for these very complex behaviors [63] and the poor or short-lasting effects of many, if not most, adherence interventions [64–66].

Comparing and Contrasting the Standard Approach and MOST

The Treatment Package Approach

Figure 2 compares the classical treatment package approach and MOST. Suppose our hypothetical investigator uses the treatment package approach, as depicted in the

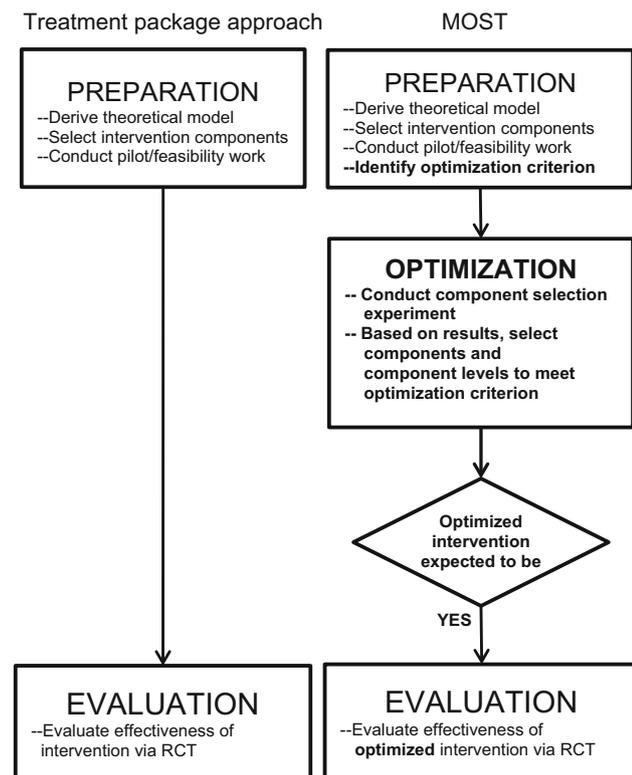


Fig. 2 Comparison of treatment package approach and multiphase optimization strategy (MOST). Differences are in *bold*

left column of Fig. 2. The investigator would prepare by developing a conceptual model such as the one shown in Fig. 1 and identifying the components to be included in the intervention. Then, the investigator would generally pilot test² the components, either individually or, more commonly, as a package. The purpose of the pilot test would be to verify that the intervention is acceptable to participants, feasible, and implementable, although in some cases the pilot test would examine the acceptability and feasibility of each component. Based on the pilot test, any necessary adjustments to the intervention package or components would be made. The investigator would then finalize the treatment package and proceed directly to evaluation of the treatment package in an RCT. This RCT would address the critical question of whether the intervention performs significantly better than a control or comparison condition.

MOST

The right column of Fig. 2 provides a brief summary of how the investigator would proceed through the preparation, optimization, and evaluation phases of the MOST framework. The initial steps of preparation are similar to the treatment package approach: a conceptual model is developed, and discrete intervention components are selected. However, in comparison to the classical treatment package approach MOST requires an increased emphasis on distinct components with respect to theoretical target, as is reflected in Fig. 1. The components are then pilot tested for acceptability, feasibility, evidence of effectiveness, and implementability and refined as needed.

An additional step included in the preparation phase of MOST is identification of an optimization criterion. The optimization criterion is an operational definition of the “best” intervention, subject to specific resource limitations or other constraints. Let us say that in our hypothetical example, the investigator has determined that to be scalable the ART adherence intervention must be implementable at a cost of no more than \$400 per person. Thus our optimization criterion will be “lowest average HIV viral load that can be obtained for less than \$400 per person.” Alternatively, if time is limited in busy HIV clinic settings, an appropriate optimization criterion might be “lowest HIV viral load that can be obtained using no more than 5 h of clinic time.” The optimization criterion also can be expressed in terms of cost-effectiveness, where cost is expressed in terms of money, time, or any resource or set of

² We subscribe to the definition of the role of pilot testing in intervention science that has been outlined by Leon et al. [68]: “The fundamental purpose of conducting a pilot study is to examine the feasibility of an approach that is intended to ultimately be used in a larger scale study... A pilot study... is not used for hypothesis testing” (p. 626).

resources. These should be considered examples only; there are many possibilities for optimization criteria. The optimization criterion is selected based on the objectives of a particular intervention and a realistic assessment of the constraints within which it will be operating.

The investigator working within the MOST framework would next proceed to the optimization phase. In this phase, which is typically not included in the treatment package approach, a *component selection experiment* (sometimes called a component screening experiment) is conducted. In this experiment, data are collected and analyzed with the specific objective of providing information on the individual effects of *each component* and on interactions between components. In other words, this experiment assesses the effectiveness of each individual component and whether its presence or absence, or its presence at a particular setting or dosage, has an impact on the effectiveness of other components. Additional relevant data, such as the cost of or time required for each component, may be collected in this phase. The composition of the optimized intervention package is then selected based on the chosen optimization criterion and the empirical results of the component selection experiment. Weak or non-performing components are eliminated. In our hypothetical example, the investigator would use the results to identify the optimized intervention by selecting the set of active components and component levels that produce the lowest HIV viral load without exceeding an implementation cost of \$400 per person. The optimization phase is discussed in more detail below.

After the investigator has identified the optimized intervention, the next step is evaluation of this optimized intervention in a standard RCT. Unlike investigators using the treatment package approach, at this point the investigator using MOST has a good sense of the likely performance of the intervention to be evaluated. Suppose given the set of components under consideration, the best intervention that can be delivered for \$400 per person or less is made up of so few components, or such weak components, that an evaluation via an RCT would likely produce a non-significant effect. In this case it would not make sense to devote resources to an RCT; instead, those resources could be devoted to going “back to the drawing board” by beginning a new study, reconsidering the conceptual model, and devising new components. Note that because the component selection experiment revealed which components were performing well, the investigator can build on this knowledge in the new study by focusing on rethinking, revising or replacing the components that performed poorly, thereby producing a materially improved intervention.

Now suppose that instead of the outcome just described, the results from the optimization phase indicate that a set of

components with substantial individual effects has been selected for inclusion in the intervention. Then it would make sense to proceed to the evaluation phase and subject the optimized intervention to an RCT. (We return to these ideas below.)

The Optimization Phase

Because the optimization phase may be unfamiliar to intervention scientists, we will discuss it in more detail here, using the hypothetical behavioral intervention discussed above and presented in Fig. 1 as an example.

Selection of an Experimental Design for the Optimization Phase of MOST

The optimization phase of MOST involves conducting a component selection experiment. The results of this experiment will form the basis for selection of the components to include in the optimized intervention. Choosing the most appropriate design for this component selection experiment is a critical part of the optimization phase of MOST. Any reasonable experimental design is a candidate for the component selection experiment. The only requirement is that the choice of design be based on the resource management principle [67]. This principle states that the appropriate experimental design is one that addresses the critical research questions while making the best use of the resources available for the experiment. Different experimental designs are suited to different research questions and make different resource demands [67]. In our view, component selection experiments are not pilot studies [68] because they are carefully controlled, fully powered experiments to be used for hypothesis testing.

Suppose our hypothetical investigator is considering various experimental designs for the component selection experiment. The investigator has decided that because knowledge about the effects of alcohol on HIV-related health and ART adherence is a necessary foundation for any intervention in this area, the informational session (described above) will be delivered to all participants and thus will not be examined via experimental manipulation (although this session could be informally evaluated by giving participants a brief knowledge assessment test). Thus there remain five components to be examined experimentally. These are listed in Table 1. For purposes of experimentation each component can take on two levels: on (included in the intervention package) and off (not included). The investigator has decided that to be eligible for inclusion in the intervention, a component must demonstrate a standardized effect size d of at least .25.

Thus it is necessary to power the experiment to detect $d \geq .25$.

One option is to conduct a separate experiment to examine each component, for a total of five individual experiments. Each experiment would essentially be an RCT, made up of an experimental condition in which the informational session is provided, one component is set to on, and the remaining components are set to off; and a control condition in which only the informational session is provided. For example, the experiment to examine the effect of motivational interviewing would have an experimental condition in which subjects received only the informational session and the motivational interviewing component. The control condition would be identical to the experimental condition, except that motivational interviewing would not be included. A total of ten experimental conditions would be implemented in the five experiments. A standard power analysis performed using the SAS macro FactorialPowerPlan ([69]; <http://methodology.psu.edu/downloads>) indicates that $N = 505$ is sufficient to achieve a power of .80 to detect $d \geq .25$ in a single experiment. Thus to conduct all five experiments and maintain this level of statistical power would require a total sample size of $5 \times 505 = 2525$.

The remaining experimental design options to be considered here are variations on the factorial design. When the objective is to examine the effects of individual components, a factorial experimental design can be an efficient and economical choice. When another important objective is determining whether the presence of one component enhances or reduces the effect of another, a factorial experiment is the only choice, because other experimental designs do not permit estimation of interaction effects. In our hypothetical example, a complete factorial experiment would require 32 experimental conditions, considerably more than the ten required by the individual experiments approach. However, the factorial experiment would require considerably fewer subjects than the individual experiments approach. A standard power analysis conducted in the same manner as above shows that statistical power of approximately .80 will be maintained in the factorial

experiment with an overall $N = 505$, which is only one-fifth of the subjects required by the individual experiments approach.³

Let us take a closer look at the factorial experiment to see why in this situation it requires so many fewer subjects than the individual experiments approach. Table 2 lists the experimental conditions in the factorial design. There are 32 experimental conditions depicted here, so with an overall N of 505 there will be 15 or 16 subjects per condition. If this design were viewed as a 32-arm RCT, it would be woefully underpowered. However, viewed as a standard factorial experiment, this experiment achieves expected power of .80. This is because the logical underpinnings, and therefore the approach to determining statistical power, of the factorial experiment are quite different from those of the RCT and related approaches, such as the individual experiments approach described above.

In an RCT, the objective is direct comparison of two or more experimental conditions. By contrast, the objective of a factorial experiment is estimation of the main effect of each independent variable, or factor, and interaction effects involving two or more factors. This is accomplished not by making direct comparisons between experimental conditions, but by comparing means based on aggregate combinations of experimental conditions. For example, consider the main effect of motivational interviewing sessions. This would be estimated by comparing the average of all the conditions that include motivational interviewing, namely conditions 17–32, to the conditions that do not include motivational interviewing, conditions 1–16. The main effect of peer mentoring would be obtained by comparing the mean of conditions 9–16 and 26–32 to the mean of conditions 1–8 and 17–25. In this manner all subjects are involved in every effect estimate. More about this can be found in Collins et al. [70], Collins et al. [67], and experimental design textbooks such as Kuehl [71].

Fractional Factorial Designs

Suppose in our hypothetical example the investigator wishes to take advantage of the efficient use of experimental subjects offered by the factorial design, but has determined that no more than 16 experimental conditions are feasible. This is a good motivation to consider a fractional factorial design for the component selection

Table 1 Hypothetical intervention components to be examined in the optimization phase

Component	Low level	High level
Motivational interviewing	Off ^a	On
Peer mentoring	Off	On
Text message support	Off	On
Mindfulness meditation training	Off	On
Behavioral adherence skill training	Off	On

^a Off = not included in intervention; On = included in intervention

³ Everything said about factorial experiments in this article assumes that the data from the experiment are to be analyzed in a standard factorial analysis of variance (ANOVA) using effect coding rather than dummy coding. We strongly recommend the use of effect coding, which, unlike dummy coding, produces estimates of main effects and interactions that are always consistent with the classic definitions found in most statistics textbooks; see Collins et al. [3].

Table 2 Conditions in complete factorial experiment

Condition number	Informational session	Motivational interviewing	Peer mentoring	Text message support	Mindfulness meditation training	Behavioral adherence skill training
1	<i>On</i> ^a	Off	Off	Off	Off	Off
2	<i>On</i>	Ofa	Off	Off	Off	<i>On</i>
3	<i>On</i>	Off	Off	Off	<i>On</i>	Off
4	<i>On</i>	Off	Off	Off	<i>On</i>	<i>On</i>
5	<i>On</i>	Off	Off	<i>On</i>	Off	Off
6	<i>On</i>	Off	Off	<i>On</i>	Off	<i>On</i>
7	<i>On</i>	Off	Off	<i>On</i>	<i>On</i>	Off
8	<i>On</i>	Off	Off	<i>On</i>	<i>On</i>	<i>On</i>
9	<i>On</i>	Off	<i>On</i>	Off	Off	Off
10	<i>On</i>	Off	<i>On</i>	Off	Off	<i>On</i>
11	<i>On</i>	Off	<i>On</i>	Off	<i>On</i>	Off
12	<i>On</i>	Off	<i>On</i>	Off	<i>On</i>	<i>On</i>
13	<i>On</i>	Off	<i>On</i>	<i>On</i>	Off	Off
14	<i>On</i>	Off	<i>On</i>	<i>On</i>	Off	<i>On</i>
15	<i>On</i>	Off	<i>On</i>	<i>On</i>	<i>On</i>	Off
16	<i>On</i>	Off	<i>On</i>	<i>On</i>	<i>On</i>	<i>On</i>
17	<i>On</i>	<i>On</i>	Off	Off	Off	Off
18	<i>On</i>	<i>On</i>	Off	Off	Off	<i>On</i>
19	<i>On</i>	<i>On</i>	Off	Off	<i>On</i>	Off
20	<i>On</i>	<i>On</i>	Off	Off	<i>On</i>	<i>On</i>
21	<i>On</i>	<i>On</i>	Off	<i>On</i>	Off	Off
22	<i>On</i>	<i>On</i>	Off	<i>On</i>	Off	<i>On</i>
23	<i>On</i>	<i>On</i>	Off	<i>On</i>	<i>On</i>	Off
24	<i>On</i>	<i>On</i>	Off	<i>On</i>	<i>On</i>	<i>On</i>
25	<i>On</i>	<i>On</i>	<i>On</i>	Off	Off	Off
26	<i>On</i>	<i>On</i>	<i>On</i>	Off	Off	<i>On</i>
27	<i>On</i>	<i>On</i>	<i>On</i>	Off	<i>On</i>	Off
28	<i>On</i>	<i>On</i>	<i>On</i>	Off	<i>On</i>	<i>On</i>
29	<i>On</i>	<i>On</i>	<i>On</i>	<i>On</i>	Off	Off
30	<i>On</i>	<i>On</i>	<i>On</i>	<i>On</i>	Off	<i>On</i>
31	<i>On</i>	<i>On</i>	<i>On</i>	<i>On</i>	<i>On</i>	Off
32	<i>On</i>	<i>On</i>	<i>On</i>	<i>On</i>	<i>On</i>	<i>On</i>

^a Off = not included in intervention; On = included in intervention. Italic data indicates higher level of factor

experiment. In a fractional factorial design, a subset, or fraction, of the experimental conditions that make up the complete factorial are run. A fractional factorial design is powered exactly the same as its complete factorial counterpart, so the overall sample size and therefore any subject costs will be exactly the same as those of the complete factorial counterpart. The appeal of fractional factorial designs is that they cut the number of experimental conditions that must be implemented, usually by half or more depending on the design selected, thereby reducing overall costs associated with implementing experimental conditions. Fractional factorial designs are commonly used in engineering, and examples of their use in intervention

science are increasing (e.g., Collins et al. [3]; Pellegrini et al. [72]; Strecher et al. [73]).

The increased economy of fractional factorial designs comes at a cost. Whenever experimental conditions are removed from a factorial experiment, some effects are combined, or, in the term used in the statistical literature, *aliased*. There are a wide variety of fractional factorial designs, and statisticians have determined which effects are aliased in each of them. Thus the investigator can select a fractional factorial design strategically so as to control which effects are aliased with which, thereby ensuring that the effects of primary scientific interest are aliased only with effects that can be assumed to be negligible in size

and are of much lesser scientific interest. The effects that are of primary scientific interest are usually main effects and lower-order interactions. To make effective use of fractional factorial designs, it is necessary to assume that the higher-order interactions, which in this example are the three-way, four-way, and five-way interactions, are negligible in size. Then a design can be selected in which the effects of primary scientific interest, namely the main effects and two-way interactions, are aliased with higher-order interactions. The logic here is that if an effect estimate is a combination of, say, a main effect and a four-way interaction, and the four-way interaction is assumed to be negligible, then the effect can be attributed primarily to the main effect.

The conceptual model is essential in determining whether or not it is reasonable to consider a fractional factorial design for the component selection experiment. If the conceptual model suggests that the effects of scientific interest are main effects and two-way interactions, and no higher-order interactions are specified, then a fractional factorial design may be worth considering. Figure 1 shows that the conceptual model guiding the hypothetical intervention does not include any higher-order interactions, so a fractional factorial experiment may be worth considering if it greatly increases economy and feasibility. In contrast, if the conceptual model specifies that there are a priori reasons to expect large higher-order interactions, then a fractional factorial experiment would probably not be a good idea.

The experimental conditions to be included in a fractional factorial design are not selected on conceptual grounds; rather, they are selected on purely statistical grounds. Selection of experimental conditions can be accomplished in a straightforward manner using software routines available in SAS[®], Minitab, or R. To use any of these routines, the investigator specifies such aspects of the experiment as the number of factors, which effects are expected to be negligible in size and which are expected to be sizeable and of scientific interest, and the maximum number of experimental conditions desired. The software then returns a suggested experimental design. In our hypothetical example the user would specify that there are five factors in the experiment; main effects and two-way interactions are expected to be sizeable and of scientific interest; and an experimental design that requires a maximum of 16 conditions, that is, a half fraction, is desired. An example of the kind of design that would be suggested by such software appears in Table 3. The experimental conditions in a fractional factorial design are always a subset of the experimental conditions in the complete factorial; it is evident that the conditions listed in Table 3 are a subset of those listed in Table 2. In the design in Table 3, each main effect is aliased with one four-way interaction, and

each two-way interaction is aliased with one three-way interaction. It is difficult to determine exactly which effects are aliased with which simply by perusing the list of experimental conditions in Table 3; fortunately, this information can be provided by the software used to select the design.

Power Analysis for Factorial Experiments with Several Factors

A power analysis for a factorial experiment can be accomplished readily using standard software for power analysis, such as PROC POWER or the FactorialPowerPlan macro (which is based on PROC POWER) in SAS[®] [74]. When conducting a power analysis for a factorial experiment it is necessary to be clear about whether effect coding, dummy coding, or some other kind of coding is to be used in the analysis. A main effect or interaction as modeled in effect coding is different from its counterpart as modeled in dummy coding, and therefore the hypothesis tests of these individual effects may be associated with different levels of power (for a more detailed explanation of the differences between effect-coded and dummy-coded effects, see Kugler et al. [75]). It may be helpful to note two characteristics of effect-coded estimates. First, effect coding produces effect estimates that are uncorrelated if there are equal numbers of subjects in each experimental condition, and very nearly uncorrelated as long as the sample sizes are approximately equal [67, 70]. Second, when effect coding is used, expected power for a regression coefficient of a particular size is the same regardless of whether the coefficient represents a main effect or interaction [67].

Comparing the Three Approaches

A comparison of the scientific yield and resource requirements of each design alternative is shown in Table 4. Each of the three designs provides an estimate of the individual effect of every component (main effects are aliased with other effects in not only the fractional factorial design but also the individual experiments approach; see Collins et al. [67] for an explanation). Only the factorial designs enable estimation of interaction effects. The complete factorial enables estimation of all of the interactions, whereas the fractional factorial design enables estimation of only two-way interactions. In our hypothetical example, the selected design enables estimation of two-way interactions; some other fractional factorial designs, typically those involving larger numbers of factors, enable estimation of selected interactions involving more than two factors.

To compare the resource requirements of the three designs it is necessary to identify approximate per-subject

Table 3 Conditions in fractional factorial experiment

Condition number from Table 2	Informational session	Motivational interviewing	Peer mentoring	Text message support	Mindfulness meditation training	Behavioral adherence skill training
2	<i>On</i> ^a	Off	Off	Off	Off	<i>On</i>
3	<i>On</i>	Off	Off	Off	<i>On</i>	Off
5	<i>On</i>	Off	Off	<i>On</i>	Off	Off
8	<i>On</i>	Off	Off	<i>On</i>	<i>On</i>	<i>On</i>
9	<i>On</i>	Off	<i>On</i>	Off	Off	Off
12	<i>On</i>	Off	<i>On</i>	Off	<i>On</i>	<i>On</i>
14	<i>On</i>	Off	<i>On</i>	<i>On</i>	Off	<i>On</i>
15	<i>On</i>	Off	<i>On</i>	<i>On</i>	<i>On</i>	Off
17	<i>On</i>	<i>On</i>	Off	Off	Off	Off
20	<i>On</i>	<i>On</i>	Off	Off	<i>On</i>	<i>On</i>
22	<i>On</i>	<i>On</i>	Off	<i>On</i>	Off	<i>On</i>
23	<i>On</i>	<i>On</i>	Off	<i>On</i>	<i>On</i>	Off
26	<i>On</i>	<i>On</i>	<i>On</i>	Off	Off	<i>On</i>
27	<i>On</i>	<i>On</i>	<i>On</i>	Off	<i>On</i>	Off
29	<i>On</i>	<i>On</i>	<i>On</i>	<i>On</i>	Off	Off
32	<i>On</i>	<i>On</i>	<i>On</i>	<i>On</i>	<i>On</i>	<i>On</i>

^a Off = not included in intervention; On = included in intervention. Italic data indicates higher level of factor

Table 4 Comparison of scientific yield and costs of three experimental design alternatives

	Individual experiments	Complete factorial design	Fractional factorial design (half fraction)
Interaction effects	Cannot be estimated	All can be estimated	Two-way interactions can be estimated
Number of experimental conditions	10	32	16
Experimental condition costs @ \$2,000/condition	\$20,000	\$64,000	\$32,000
Number of subjects	2,525	505	505
Subject costs @ \$100/subject	\$252,500	\$50,500	\$50,500
Total costs	\$272,500	\$114,500	\$82,500

and per-condition costs. Suppose the per-subject costs are expected to be approximately \$100, and the overhead expenses associated with each experimental condition are expected to be approximately \$2,000. Under these cost assumptions the individual experiments approach is the most costly by a wide margin, and the fractional factorial experiment is the least costly. As compared to the individual experiments approach, the use of a complete factorial design would save \$173,000, and the use of a fractional factorial design would save \$205,600. The relative differences in cost among these experimental designs is driven by the ratio of per-condition costs to per-subject costs [67]. In this hypothetical example the per-condition costs are 20 times the per-subject costs. For the fractional factorial experiment to approach the cost of the individual experiments approach, the per-condition costs would have

to be over 150 times the per-subject costs. (Readers are invited to recalculate this comparison using figures that are more realistic for their research. More detail can be found in Collins et al. [67], and the Relative Costs SAS macro with documentation to help with the comparison can be found at <http://methodology.psu.edu/downloads>.)

Suppose our hypothetical investigator is uncomfortable with using a fractional factorial design and prefers a complete factorial for the component selection experiment. Given that 16 experimental conditions is the maximum that resources will allow, if a complete factorial is to be conducted the investigator will have to reduce the scope of the component selection experiment to include only four factors. Thus the choice is between a fractional factorial experiment that enables examination of five components and will produce aliased effect estimates, and a complete

factorial experiment that will produce effect estimates with no aliasing but enables examination of only four components. Only the investigator can determine which is more appropriate in a given endeavor. In this article we do not advocate one strategy over another, only that investigators make an informed decision based on the resource management principle, taking into account the objectives of a particular study, the conceptual model, the experimental design alternatives, and the available resources.

Data Analysis and Decision Making

After the data from the component selection experiment have been collected, they can be analyzed in whatever manner is most appropriate. If a factorial experiment has been conducted, the data can be analyzed using a classic factorial ANOVA. As mentioned above, we recommend the use of effect coding for this analysis.

Once the main effects of the components and interactions between components have been estimated in the ANOVA, this information, along with other information, such as data on cost, can be used to decide which set of components best meets the optimization criterion. Decision making based on the results of the component selection experiment is an open research area. One approach, outlined in Collins et al. [76], involves beginning by making a preliminary selection of components that have achieved main effects exceeding a predetermined criterion for statistical significance. This preliminary selection is then systematically re-evaluated in light of any substantial interaction effects that have been detected to gain an understanding of how the components work in combination. Depending on the optimization criterion identified, this would then be combined with other information to make a final selection of components. Recall that the optimization criterion identified for the hypothetical example is “lowest average viral load that can be obtained for less than \$400 per person.” The investigators could use the approach outlined in Collins et al. to select viable components for inclusion in the intervention based on the ANOVA results. They could then compute the expected implementation costs of and expected outcome for various treatment packages made up of these components, identify the combinations that are expected to cost less than \$400, and determine which of these is expected to produce the lowest viral load.

In our experience it is not uncommon for both scientific theory and common sense to suggest that certain basic informational or educational material is an essential foundation for all intervention components under consideration. There also may be one or more components that have already earned a place in the intervention by having previously been demonstrated effective; ethics may dictate that these should not be withheld from participants. Such

material or components can be provided to all participants, and treated as a constant in the experiment. Later, when decisions are being made about what to include in the optimized intervention, it should be borne in mind that all of the observed effects are predicated on provision of any constant materials or components. For example, in the designs considered for the hypothetical experiment the informational session was treated as a constant (see Tables 2, 3). Suppose a factorial experiment is conducted and a main effect of adherence skill training is found. The experiment has examined adherence skill training only in the presence of the informational session. The design of this experiment does not permit examination of whether the effect of adherence skill training is about the same whether the informational session is provided or not, so it must be assumed that the informational session is necessary for adherence skill training to be effective. Thus it would be inappropriate to omit the informational session from the final version of the intervention. For this reason, it is best to select any constants that are to be included in a component selection experiment very carefully.

The conceptual model includes a number of hypothesized nonexperimental, that is, observed, moderators (see box on lower right of Fig. 2). Hypotheses about moderation by observed variables can be tested empirically by coding interactions between any of the moderator variables and the factors in the experiment and adding these to the set of predictor variables in the analysis. In general, statistical power for tests of hypotheses involving observed moderators is likely to be lower than power for tests of hypotheses about main effects of experimentally manipulated factors or interactions between them.

Why is the Evaluation Phase Necessary?

Above we mentioned that at the end of the optimization phase, the investigator may or may not decide to proceed to the evaluation phase, depending upon whether the results of the component selection experiment indicate that the optimized intervention is likely to be effective. This may prompt the question, why is the evaluation phase needed at all if it is to be undertaken only when the investigator is already confident that the intervention is effective?

To address this question, it is necessary to compare the kinds of experimental designs that are most useful in the optimization and evaluation phases. Consider our hypothetical example, in which an efficient factorial experiment was selected for the optimization phase. Suppose based on this information, an optimized intervention is constructed consisting of the informational session plus peer mentoring, mindfulness meditation training, and behavioral adherence skill training. Now the investigators would like to assess the performance of this optimized intervention.

It is tempting to consider conducting a sort of mini-RCT based on the factorial experiment, by comparing the experimental condition corresponding to the optimized intervention (in this example, experimental condition 12 in Table 2) directly to the experimental condition that corresponds to all of the components set to off (condition 1 in Table 2). There are two reasons why this may be impractical. First, as was discussed above, factorial experiments are powered for estimates of main effects and interactions, not for direct comparison of individual experimental conditions. Recall that the factorial experiment is fully powered with $N = 505$, or 15 or 16 per condition. This means that the comparison of conditions 1 and 12 would be based on a total N of about 31. The estimate of the effect size would have a large standard error and there would be insufficient statistical power for the hypothesis test. In general, the only way to enable sufficient precision and power for this kind of comparison within a factorial experiment is to power the experiment as if it were a 32-arm RCT, which would be impractical in most cases. Second, if a fractional factorial design was used for the component selection experiment, it may not include experimental conditions corresponding to either the combination of components that make up the optimized intervention or a condition in which all the components are set to the lowest level. For example, the design shown in Table 3 does not include a condition in which all of the components are set to off. Such a condition is not needed in this balanced fractional factorial experiment, but it would be needed to make the kind of comparison that is made in an RCT.

There may be times when the results of the optimization phase are so compelling that evaluation of the intervention via an RCT is deemed unnecessary. But when evaluation of the performance of the optimized intervention package is necessary, and in particular when an estimate of the intervention's effect size is desired, evaluation via an RCT must be undertaken.

Discussion

Optimization of an Existing Intervention

In this article we have discussed how the MOST framework can be used to develop, optimize, and evaluate behavioral and biobehavioral interventions for prevention and treatment of HIV/AIDS, using as an illustration a hypothetical example involving building a brand-new intervention. MOST can also be used to optimize an existing intervention to make it more effective, economical, efficient, or scalable. The starting point for this would be much the same as that for development of a new optimized intervention, namely the conceptual model and selection of intervention components to examine. Given

that the components will have been implemented previously as part of the existing treatment package, it may not be necessary to conduct any pilot testing before the component selection experiment.

In some cases the treatment package may be considered satisfactory, and the target for optimization is instead the fidelity of delivery of the intervention. Here the objective is to reduce the decrement in intervention performance between efficacy and effectiveness. An example of this can be found in Caldwell et al. [77], who were interested in optimizing the fidelity of delivery of HealthWise [78], a school-based intervention for prevention of drug abuse and HIV developed for South African youth. They investigated three components hypothesized to promote fidelity: enhanced teacher training, enhanced support to teachers who delivered the HealthWise program, and measures to enhance the school climate so that it would be more supportive of HealthWise.

Levels of Components to be Examined Experimentally

In the hypothetical example all of the component levels examined in the experiment were either on or off. In many cases it may be desirable to compare a low and a high level of a component. For example, Collins et al. [3] described the use of MOST to develop an optimized smoking cessation intervention. In their component selection experiment, several factors had levels representing low and high rather than off and on. For example, one factor was duration of nicotine replacement therapy, with levels shorter (8 weeks) and longer (16 weeks). When the levels of a component that are included in the component selection experiment are low and high rather than off and on, the results of the experiment cannot be used to decide whether or not the component should be included in the intervention. Instead, an assumption is made that the component is to be included at either the low level or the high level; the purpose of the component selection experiment is to help determine which level will be selected.

It is possible to include factors with more than two levels in a factorial experiment. However, including even one factor with more than two levels greatly increases resource requirements, both in terms of number of subjects and number of experimental conditions. We recommend using two levels per factor in component selection experiments wherever possible.

Different Approaches to Experimentation in the Optimization Phase

As mentioned above, any of a variety of experimental design approaches may be selected for the optimization

phase. The only requirement made by MOST is that the design be selected based on the resource management principle. Often this will lead an investigator to a standard factorial or fractional factorial experiment, but there are instances in which careful consideration of objectives and economy will suggest that a different experimental design is called for.

In time-varying adaptive interventions [79], the participant is assessed periodically, and the amount of intervention or even the intervention strategy may be varied depending on the individual's progress. For example, McKay [80] evaluated the effectiveness of an adaptive intervention for alcohol dependence, consisting of components such as telephone calls, counseling, and periodic assessment of relapse risks. When relapse risk exceeded predetermined levels, the intensity of the intervention was increased. To optimize a time-varying adaptive intervention, the most appropriate design for the component selection experiment is frequently a sequential, multiple assignment, randomized trial (SMART). The SMART is an innovative variation on the factorial experiment. More about how the SMART can be used in the optimization phase of MOST can be found in Collins, Nahum-Shani, and Almirall [4].

In some circumstances a time-varying adaptive intervention can be viewed as a dynamical system, opening up the possibility of approaching optimization from the perspective of control engineering [81, 82]. In other words, a controller could be derived for a time-varying adaptive intervention, in much the same way as controllers are built for automobile cruise-control systems, household heating and ventilation systems, and the like. The initial applications in this exciting new area are occurring in the obesity field [83, 84], but there is much potential for the HIV/AIDS field.

Potential Benefits of Using MOST in HIV/AIDS Intervention Research

We see several potential long-term benefits of using MOST in HIV/AIDS intervention research. First, when MOST is used, key constraints that operate in real-world settings can be incorporated from the beginning. Recall that in our hypothetical example, the investigator set about to develop the most effective intervention that could be obtained with the intervention components under consideration, mindful that the intervention is most likely to be implemented successfully in community settings if it costs no more than \$400 per person to deliver. This was translated into the optimization criterion, which means that the resulting intervention is guaranteed to be scalable, assuming that the \$400 per person limit guarantees scalability.

Second, MOST enables intervention scientists to engineer an intervention to attain a desired level of cost-effectiveness, by making cost-effectiveness a part of the optimization criterion. By contrast, using the treatment package approach, cost-effectiveness must be assessed after the intervention has been developed and evaluated, at which point it is a *fait accompli*.

Third, optimization could be done in a transparent manner by maintaining a web site upon which investigators can post the results of the component selection experiment, including data on resource requirements such as money and time that would be accessible by other intervention scientists. (We refer here to aggregate results, such as the results of a factorial ANOVA, that would not threaten confidentiality.) This would enable anyone to use the results to develop a different intervention based on a different optimization criterion more relevant to a particular situation. For example, suppose a community wishes to implement our hypothetical intervention, but they know they can spend no more than \$300 per person. The results from our component selection experiment can be used to determine which components to retain to arrive at the most effective intervention that can be delivered without exceeding this limit. Although it is natural to think in terms of removing one or more of the components that make up the intervention that was optimized using the criterion of no more than \$400 per person, in fact an intervention optimized using the \$300 per person criterion may or may not be made up of a subset of those components.

Fourth, using MOST will enable intervention science to look inside the “black box” of behavioral interventions to develop a coherent knowledge base about which components are effective and which are not, and which components enhance or reduce the effect of which other components. In particular, at this writing little is known about interactions between intervention components, because so few factorial experiments have been conducted. Thus MOST is responsive to Johnson et al.'s [85] call for approaches that “help refocus the field of HIV prevention on improved research strategies to further improve future interventions by discerning the content design factors related to success for particular populations, rather than merely to assess whether the interventions have been successful” (p. S259).

Fifth, once an intervention has been optimized to an explicitly operationalized criterion, future research can focus on measurably improving the intervention. The objective could be to develop an intervention that is just as effective as its predecessor, but shorter or less expensive; or that costs no more, but is more effective; or is better in some other clearly defined way. The coherent knowledge base mentioned in the preceding paragraph will help to speed up this effort. If a particular component's

effectiveness has been established in prior research, it may be possible to include it in subsequent interventions without further experimentation. Alternatively, it may be desirable to include it in the component selection experiment to determine whether it interacts with other components that are under consideration. In this manner, by working systematically, it may be possible to make steady, incremental progress over time in improving behavioral interventions for prevention and treatment of HIV/AIDS.

MOST and the Five-Year Grant Cycle

Funding from the US National Institutes of Health (NIH), as well as many funding agencies inside and outside the US, usually is provided for a maximum of 5 years. Depending on the endeavor it may be difficult to complete all three phases of MOST in a single 5-year study, even if much of the preparation phase is completed at the time of writing a grant proposal. Whether a 5-year plan of work can reasonably include the evaluation phase in addition to the optimization phase is largely dependent on two considerations. One is how rapidly research subjects can be recruited. In some domains, such as school-based research, large numbers of subjects can be recruited rapidly all at once; in other domains, such as many studies in medical settings, recruitment is on a rolling basis, meaning that it is dependent on the rate at which patients walk into a particular clinic or are admitted to a hospital.

The other consideration is the event horizon for the main outcome. In our hypothetical example, the outcome of viral load is something that can reflect component effects within a few weeks. By contrast, if the main outcome is sexual risk behavior, it might take a year or more to observe enough behavior to draw conclusions about component effectiveness. When the effects of the intervention components on the primary outcome are expected to occur in the distant future, it may be practical to use measures of the mediators as short-term outcomes for the component selection experiment, and optimize the intervention based on these measures rather than the primary outcome. This is another way in which the conceptual model plays a critical role in MOST; such a strategy will be successful only if the conceptual model is an accurate depiction of the process that is being intervened upon. The primary outcome variable would be used in the RCT conducted in the evaluation phase.

Thus there are two general strategies: (a) propose to complete all three phases of MOST, or (b) propose to complete the preparation and optimization phases, and indicate that future funding will be sought to complete the evaluation phase. If (a) is chosen, the investigator may be criticized for proposing an overly ambitious timeline. Another potential criticism is that it is unclear what course of action should be

taken if so few of the components show effects that an RCT is not warranted. (Our view is that sufficient flexibility could be allowed for the investigator to consult with the relevant NIH program staff and, with their consent, reconsider the conceptual model and components and conduct another component selection experiment. However, such a flexible strategy, although consistent with how most scientists have been taught to conduct research, is not typically encouraged in today's program announcements and can make reviewers uncomfortable). If (b) is chosen, reviewers who are accustomed to expecting delivery of a fully evaluated treatment package at the end of a 5-year funding period may feel shortchanged. We have received both types of critiques, but also have successfully proposed (a) and (b) in different projects as we deemed appropriate. Our suggestion is to lay out both options clearly and explain the rationale for the choice that was made.

Limitations and Future Directions

As mentioned above, one open area of research is approaches to making decisions about which components and component levels should be included to make up the optimized intervention. More work is needed on how to incorporate data on cost, time, and other resource demands. In addition, guidelines are needed in some cases on how to collect this kind of data. For example, it is not always clear how to collect data on resource limitations that are imposed when an intervention goes to scale.

More work is needed on experimental design. For example, many intervention scientists wish to examine some kind of group therapy as one of the components being considered, or to compare group therapy to individual therapy. When an intervention component is delivered in a group setting, dependence between observations is introduced. This dependence is not there at pretest if the subjects have been randomly assigned, but it can grow as a result of group therapy [86]. Thus an intra-class correlation often must be modeled in one level of the factor and not in the other level of the factor in which group therapy is not provided. Methods of powering such studies and guidelines for analyzing the resulting data are needed.

Conclusions

Now may be the time to take a longer view on the science of behavioral and biobehavioral interventions for HIV/AIDS. By implementing new methodological approaches for intervention development and evaluation such as MOST, intervention science can arrive at approaches for prevention and treatment of HIV/AIDS that are more effective, economical, efficient, and scalable. Moreover, as

time goes on, each intervention will be better than its predecessors along clearly articulated dimensions. In this manner, the public health impact of behavioral interventions for prevention and treatment of HIV/AIDS will increase, systematically and incrementally making progress in moving society toward an end to the pandemic.

Acknowledgments The work described herein was supported by Grant R03 HD079711 from the Eunice Kennedy Shriver Institute for Child Health and Human Development; grants P50 DA010075, R01 DA032083, and P30 DA011041 from the National Institute on Drug Abuse; grant R01 DK097364 from the National Institute for Diabetes and Digestive and Kidney Diseases; and grant P01 CA180945 from the National Cancer Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the institutes mentioned above.

References

- 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 e-health interventions. *Am J Prev Med.* 2007;32:S112–8.
- Collins LM, Baker TB, Mermelstein RJ, Piper ME, Jorenby DE, Smith SS, et al. The multiphase optimization strategy for engineering effective tobacco use interventions. *Ann Behav Med.* 2011;41(2):208–26.
- Collins LM, Nahum-Shani I, Almirall D. Optimization of behavioral dynamic treatment regimens based on the sequential, multiple assignment, randomized trial (SMART). *Clin Trials.* 2014;11:426–34.
- Seeley J, Watts CH, Kippax S, Russell S, Heise L, Whiteside A. Addressing the structural drivers of HIV: a luxury or necessity for programmes? *J Int AIDS Soc.* 2012;15(Suppl 1):17397.
- Latkin C, Kuramoto S, Davey-Rothwell M, Tobin K. Social norms, social networks, and HIV risk behavior among injection drug users. *AIDS Behav.* 2010;14(5):1159–68.
- Stangl AL, Lloyd JK, Brady LM, Holland CE, Baral S. A systematic review of interventions to reduce HIV-related stigma and discrimination from 2002 to 2013: how far have we come? *J Int AIDS Soc.* 2013;16(3Suppl 2):18734.
- Beyrer C, Baral SD, van Griensven F, Goodreau SM, Chariyalertsak S, Wirtz AL, et al. Global epidemiology of HIV infection in men who have sex with men. *Lancet.* 2012;380(9839):367–77.
- Hagan H, Jenness S, Wendel T, Murrill C, Neaigus A, Gelpi-Acosta C. Herpes simplex virus type 2 associated with HIV infection among New York heterosexuals living in high-risk areas. *Int J STD AIDS.* 2010;21(8):580–3.
- Oster AM, Sternberg M, Nebenzahl S, Broz D, Xu F, Hariri S, et al. Prevalence of HIV, sexually transmitted infections, and viral hepatitis by urbanicity, among men who have sex with men, injection drug users, and heterosexuals in the United States. *Sex Transm Dis.* 2014;41(4):272–9.
- Adimora AA, Auerbach JD. Structural interventions for HIV prevention in the United States. *J Acquir Immune Defic Syndr.* 2010;55(02):S132.
- Degenhardt L, Mathers B, Vickerman P, Rhodes T, Latkin C, Hickman M. Prevention of HIV infection for people who inject drugs: why individual, structural, and combination approaches are needed. *Lancet.* 2010;376(9737):285–301.
- Kurth AE, Celum C, Baeten JM, Vermund SH, Wasserheit JN. Combination HIV prevention: significance, challenges, and opportunities. *Curr HIV/AIDS Rep.* 2011;8(1):62–72.
- Centers for Disease Control and Prevention. HIV Transmission 2014. <http://www.cdc.gov/hiv/basics/transmission.html>. Accessed 8 Dec 2014.
- Dosekun O, Fox J. An overview of the relative risks of different sexual behaviours on HIV transmission. *Curr Opin HIV AIDS.* 2010;5(4):291–7.
- Grinsztejn B, Hosseinipour MC, Ribaud HJ, Swindells S, Eron J, Chen YQ, et al. Effects of early versus delayed initiation of antiretroviral treatment on clinical outcomes of HIV-1 infection: results from the phase 3 HPTN 052 randomised controlled trial. *Lancet Infect Dis.* 2014;14(4):281–90.
- Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, Kumarasamy N, et al. Prevention of HIV-1 infection with early antiretroviral therapy. *N Engl J Med.* 2011;365(6):493–505.
- Gwads MV, BCAP Collaborative Research Team. STTR with heterosexuals at high risk for HIV: Preliminary findings. In: Annual Meetings of the Society for Prevention Research; San Francisco, CA2013.
- Baliunas D, Rehm J, Irving H, Shuper P. Alcohol consumption and risk of incident human immunodeficiency virus infection: a meta-analysis. *Int J Public Health.* 2010;55(3):159–66.
- Rehm J, Shield KD, Joharchi N, Shuper PA. Alcohol consumption and the intention to engage in unprotected sex: systematic review and meta-analysis of experimental studies. *Addiction.* 2012;107(1):51–9.
- Kahler CW, Wray TB, Pantalone DW, Kruis RD, Mastroleone NR, Monti PM, et al. Daily associations between alcohol use and unprotected anal sex among heavy drinking HIV-positive men who have sex with men. *AIDS Behav.* 2014;19:422–30.
- Kelly JA, DiFranceisco WJ, Lawrence JSS, Amirkhanian YA, Anderson-Lamb M. Situational, partner, and contextual factors associated with level of risk at most recent intercourse among Black men who have sex with men. *AIDS Behav.* 2014;18(1):26–35.
- Wolf SE, Maisto SA. Alcohol use and risk of HIV infection among men who have sex with men. *AIDS Behav.* 2009;13(4):757–82.
- Azar MM, Springer SA, Meyer JP, Altice FL. A systematic review of the impact of alcohol use disorders on HIV treatment outcomes, adherence to antiretroviral therapy and health care utilization. *Drug Alcohol Depend.* 2010;112(3):178–93.
- Braithwaite RS, Bryant KJ. Influence of alcohol consumption on adherence to and toxicity of antiretroviral therapy and survival. *Alcohol Res Health.* 2010;33(3):280.
- Hendershot CS, Stoner SA, Pantalone DW, Simoni JM. Alcohol use and antiretroviral adherence: review and meta-analysis. *J Acquir Immune Defic Syndr.* 2009;52(2):180.
- Glass TR, Bategay M, Cavassini M, De Geest S, Furrer H, Vernazza PL, et al. Longitudinal analysis of patterns and predictors of changes in self-reported adherence to antiretroviral therapy: Swiss HIV Cohort Study. *J Acquir Immune Defic Syndr.* 2010;54(2):197–203.
- Grant RM, Lama JR, Anderson PL, McMahan V, Liu AY, Vargas L, et al. Preexposure chemoprophylaxis for HIV prevention in men who have sex with men. *N Engl J Med.* 2010;363(27):2587–99.
- Attia S, Egger M, Müller M, Zwahlen M, Low N. Sexual transmission of HIV according to viral load and antiretroviral therapy: systematic review and meta-analysis. *Aids.* 2009;23(11):1397–404.

30. Shuper PA, Neuman M, Kanteres F, Baliunas D, Joharchi N, Rehm J. Causal considerations on alcohol and HIV/AIDS—a systematic review. *Alcohol Alcohol*. 2010;45(2):159.
31. Wu ES, Metzger DS, Lynch KG, Douglas SD. Association between alcohol use and HIV viral load. *J Acquir Immune Defic Syndr*. 2011;56(5):e129.
32. Lewis MA, Patrick ME, Litt DM, Atkins DC, Kim T, Blayney JA, et al. Randomized controlled trial of a web-delivered personalized normative feedback intervention to reduce alcohol-related risky sexual behavior among college students. *J Consult Clin Psychol*. 2014;82(3):429.
33. Parsons JT, Golub SA, Rosof E, Holder C. Motivational interviewing and cognitive-behavioral intervention to improve HIV medication adherence among hazardous drinkers: a randomized controlled trial. *J Acquir Immune Defic Syndr*. 2007;46(4):443.
34. Office of National AIDS Policy. National HIV/AIDS strategy: Update of the 2014 federal actions to achieve national goals and improve outcomes along the HIV care continuum. Washington, DC: 2014.
35. Miller WR, Rollnick S. Motivational interviewing: helping people change. New York: Guilford Press; 2012.
36. Ajzen I. From intentions to actions: a theory of planned behavior. New York: Springer; 1985.
37. Vansteenkiste M, Sheldon KM. There's nothing more practical than a good theory: integrating motivational interviewing and self-determination theory. *Br J Clin Psychol*. 2006;45(1):63–82.
38. Bandura A. Social learning theory. Englewood Cliffs: Prentice Hall; 1977.
39. Bandura A. Social foundations of thought and action. Englewood Cliffs: Prentice Hall; 1986.
40. Ewart CK. Social action theory for a public health psychology. *Am Psychol*. 1991;46(9):931.
41. Kalichman SC, Grebler T, Amaral CM, McNerey M, White D, Kalichman MO, et al. Intentional non-adherence to medications among HIV positive alcohol drinkers: prospective study of interactive toxicity beliefs. *J Gen Intern Med*. 2013;28(3):399–405.
42. Nyamathi A, Flaskerud JH, Leake B, Dixon EL, Lu A. Evaluating the impact of peer, nurse case-managed, and standard HIV risk-reduction programs on psychosocial and health-promoting behavioral outcomes among homeless women. *Res Nurs Health*. 2001;24(5):410–22.
43. Purcell DW, Latka MH, Metsch LR, Latkin CA, Gómez CA, Mizuno Y, et al. Results from a randomized controlled trial of a peer-mentoring intervention to reduce HIV transmission and increase access to care and adherence to HIV medications among HIV-seropositive injection drug users. *J Acquir Immune Defic Syndr*. 2007;46:S35–47.
44. Brown JL, Littlewood RA, Venable PA. Social-cognitive correlates of antiretroviral therapy adherence among HIV-infected individuals receiving infectious disease care in a medium-sized northeastern US city. *AIDS Care*. 2013;25(9):1149–58.
45. Langebeek N, Gisolf EH, Reiss P, Vervoort SC, Thóra B, Richter C, et al. Predictors and correlates of adherence to combination antiretroviral therapy (cART) for chronic HIV infection: a meta-analysis. *BMC Med*. 2014;12(1):142.
46. Sullivan LE, Goulet JL, Justice AC, Fiellin DA. Alcohol consumption and depressive symptoms over time: a longitudinal study of patients with and without HIV infection. *Drug Alcohol Depend*. 2011;117(2):158–63.
47. Cruess DG, Kalichman SC, Amaral C, Swetzes C, Cherry C, Kalichman MO. Benefits of adherence to psychotropic medications on depressive symptoms and antiretroviral medication adherence among men and women living with HIV/AIDS. *Ann Behav Med*. 2012;43(2):189–97.
48. Gonzalez JS, Batchelder AW, Psaros C, Safren SA. Depression and HIV/AIDS treatment nonadherence: a review and meta-analysis. *J Acquir Immune Defic Syndr*. 2011;58(2):181–7.
49. Springer SA, Dushaj A, Azar MM. The impact of DSM-IV mental disorders on adherence to combination antiretroviral therapy among adult persons living with HIV/AIDS: a systematic review. *AIDS Behav*. 2012;16(8):2119–43.
50. Norton WE, Amico KR, Fisher WA, Shuper PA, Ferrer RA, Cornman DH, et al. Information–motivation–behavioral skills barriers associated with intentional versus unintentional ARV non-adherence behavior among HIV + patients in clinical care. *AIDS care*. 2010;22(8):979–87.
51. Lazo M, Gange SJ, Wilson TE, Anastos K, Ostrow DG, Witt MD, et al. Patterns and predictors of changes in adherence to highly active antiretroviral therapy: longitudinal study of men and women. *Clin Infect Dis*. 2007;45(10):1377–85.
52. Oh DL, Sarafian F, Silvestre A, Brown T, Jacobson L, Badri S, et al. Evaluation of adherence and factors affecting adherence to combination antiretroviral therapy among White, Hispanic, and Black men in the MACS Cohort. *J Acquir Immune Defic Syndr*. 2009;52(2):290.
53. Harris GE, Larsen D. HIV peer counseling and the development of hope: perspectives from peer counselors and peer counseling recipients. *AIDS Patient Care STDs*. 2007;21(11):843–60.
54. Guillory J, Chang P, Henderson CR Jr, Shengelia R, Lama S, Warmington M, et al. Piloting a text message-based social support intervention for patients with chronic pain: establishing feasibility and preliminary efficacy. *Clin J Pain*. 2015;31(6):548–56.
55. Mo PK, Coulson NS. Exploring the communication of social support within virtual communities: a content analysis of messages posted to an online HIV/AIDS support group. *Cyberpsychol Behav*. 2008;11(3):371–4.
56. Uhrig JD, Lewis MA, Bann CM, Harris JL, Furberg RD, Coomes CM, et al. Addressing HIV knowledge, risk reduction, social support, and patient involvement using SMS: results of a proof-of-concept study. *J Health Commun*. 2012;17(sup1):128–45.
57. Safren SA, Reisner SL, Herrick A, Mimiaga MJ, Stall R. Mental health and HIV risk in men who have sex with men. *J Acquir Immune Defic Syndr*. 2010;55(Suppl 2):S74.
58. McIntosh RC, Rosselli M. Stress and coping in women living with HIV: a meta-analytic review. *AIDS Behav*. 2012;16(8):2144–59.
59. Duncan LG, Moskowitz JT, Neilands TB, Dilworth SE, Hecht FM, Johnson MO. Mindfulness-based stress reduction for HIV treatment side effects: a randomized, wait-list controlled trial. *J Pain Symptom Manage*. 2012;43(2):161–71.
60. Riley KE, Kalichman S. Mindfulness-based stress reduction for people living with HIV/AIDS: preliminary review of intervention trial methodologies and findings. *Health Psychol Rev*. 2014:1–20.
61. AidsInfo. Guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents: National Institutes of Health. <http://aidsinfo.nih.gov/guidelines>. Accessed 20 Dec 2014.
62. Kalichman SC, Cherry C, Kalichman MO, Amaral CM, White D, Pope H, et al. Integrated behavioral intervention to improve HIV/AIDS treatment adherence and reduce HIV transmission. *Am J Public Health*. 2011;101(3):531–8.
63. Kaufman MR, Cornish F, Zimmerman RS, Johnson BT. Health behavior change models for HIV prevention and AIDS care: practical recommendations for a multi-level approach. *J Acquir Immune Defic Syndr*. 2014;66:S250–8.
64. Binford MC, Kahana SY, Altice FL. A systematic review of antiretroviral adherence interventions for HIV-infected people who use drugs. *Curr HIV/AIDS Rep*. 2012;9(4):287–312.
65. Charania MR, Marshall KJ, Lyles CM, Crepaz N, Kay LS, Koenig LJ, et al. Identification of evidence-based interventions for promoting HIV medication adherence: findings from a systematic review of US-based studies, 1996–2011. *AIDS Behav*. 2014;18(4):646–60.

66. Simoni JM, Nelson KM, Franks JC, Yard SS, Lehavot K. Are peer interventions for HIV efficacious? A systematic review. *AIDS Behav.* 2011;15(8):1589–95.
67. 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.
68. Leon AC, Davis LL, Kraemer HC. The role and interpretation of pilot studies in clinical research. *J Psychiatr Res.* 2011;45(5):626–9.
69. Dziak JJ, Collins LM, Wagner AT. *FactorialPowerPlan SAS macro suite users' guide.* 1st ed. University Park: The Methodology Center, Penn State; 2013.
70. Collins LM, Dziak JJ, Kugler KC, Trail JB. Factorial Experiments: efficient Tools for Evaluation of Intervention Components. *Am J Prev Med.* 2014;47(4):498–504.
71. Kuehl RO, Kuehl R. *Design of experiments: statistical principles of research design and analysis.* Pacific Grove, CA: Duxbury Press; 2000.
72. Pellegrini CA, Hoffman SA, Collins LM, Spring B. Optimization of remotely delivered intensive lifestyle treatment for obesity using the multiphase optimization strategy: Opt-IN study protocol. *Contemp Clin Trials.* 2014;38:251–9.
73. Strecher VJ, McClure JB, Alexander GL, Chakraborty B, Nair VN, Konkel JM, et al. Web-based smoking-cessation programs: results of a randomized trial. *Am J Prev Med.* 2008;34(5):373–81.
74. *Sas S. Guide SUs.* Cary: SAS Institute Inc; 1990. p. 355.
75. Kugler K, Trail J, Dziak J, Collins L. Effect coding versus dummy coding in analysis of data from factorial experiments. Technical Report, 2012.
76. Collins LM, Trail JB, Kugler KC, Baker TB, Piper ME, Mermelstein RJ. Evaluating individual intervention components: making decisions based on the results of a factorial screening experiment. *Transl Behav Med.* 2014;4:238–51.
77. Caldwell LL, Smith EA, Collins LM, Graham JW, Lai M, Wegner L, et al. Translational research in South Africa: evaluating implementation quality using a factorial design. *Child & youth care forum* (2012). Springer, New York
78. Caldwell LL, Patrick ME, Smith EA, Palen L, Wegner L. Influencing adolescent leisure motivation: intervention effects of HealthWise South Africa. *J Leis Res.* 2010;42(2):203–20.
79. Collins LM, Murphy SA, Bierman KL. A conceptual framework for adaptive preventive interventions. *Prev Sci.* 2004;5(3):185–96.
80. McKay JR. Is there a case for extended interventions for alcohol and drug use disorders? *Addiction.* 2005;100(11):1594–610.
81. 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:S31–40.
82. Lagoa CM, Bekiroglu K, Lanza ST, Murphy SA. Designing adaptive intensive interventions using methods from engineering. *J Consult Clin Psychol.* 2014;82(5):868.
83. Navarro-Barrientos J-E, Rivera DE, Collins LM. A dynamical model for describing behavioural interventions for weight loss and body composition change. *Math Comput Model Dyn Syst.* 2011;17(2):183–203.
84. Savage JS, Downs DS, Dong Y, Rivera DE. Control systems engineering for optimizing a prenatal weight gain intervention to regulate infant birth weight. *Am J Public Health.* 2014;104:e1–8.
85. Johnson BT, Michie S, Snyder LB. Effects of behavioral intervention content on HIV prevention outcomes: a meta-review of meta-analyses. *J Acquir Immune Defic Syndr.* 2014;66:S259–70.
86. Pals SL, Murray DM, Alfano CM, Shadish WR, Hannan PJ, Baker WL. Individually randomized group treatment trials: a critical appraisal of frequently used design and analytic approaches. *Am J Public Health.* 2008;98(8):1418.