Statistics for Social and Behavioral Sciences

Linda M. Collins

Optimization of Behavioral, Biobehavioral, and Biomedical Interventions

The Multiphase Optimization Strategy (MOST)

Springer

Chapter 1
Conceptual Introduction to the Multiphase Optimization Strategy (MOST)

Abstract The multiphase optimization strategy (MOST) is an engineering-inspired framework for development, optimization, and evaluation of behavioral, biobehavioral, and biomedical interventions. This chapter provides a conceptual overview of MOST and discusses how it is different from the classical approach. The focus of the classical approach is on developing an intervention a priori and then evaluating it in a randomized controlled trial (RCT). By contrast, the focus of MOST is on a phased approach: first developing and optimizing an intervention and then evaluating the optimized intervention in an RCT. The optimization is based on carefully conducted and fully powered optimization trials. The objective of MOST is to arrive at an intervention that not only demonstrates effectiveness in an RCT but also is efficient, economical, and scalable. Readers are encouraged to read the preface before this chapter, because it provides a rationale for and orientation to the book.

Contents

L. M. Collins, *Optimization of Behavioral, Biobehavioral, and Biomedical Interventions*, Statistics for Social and Behavioral Sciences,
https://doi.org/10.1007/978-3-319-72206-1_1

1.1 Introduction

This book and its companion volume (Collins & Kugler, 2018) are about MOST, a framework for development, optimization, and evaluation of behavioral, biobehavioral, and biomedical interventions. Here the term *intervention* refers to a program with the objective of improving and/or maintaining human health and well-being, broadly defined. Behavioral interventions use a strategy based on modification of affective, cognitive, or behavioral factors. Biomedical interventions use a strategy based on pharmaceuticals, surgery, and the like. Biobehavioral interventions use a strategy based on both behavioral and biomedical approaches. Interventions may have any of a variety of purposes. Examples include preventing and treating physical and mental health disorders, promoting physical and mental health, improving family functioning, preventing violence, improving learning, and promoting academic achievement. An intervention may be aimed at the individual, family, school, organizational, or community level or at a combination of levels.

The classical approach to developing and evaluating interventions has been about the same for many years. The intervention scientist identifies a set of components that potentially can be included in the intervention; perhaps conducts a pilot study[1] on these components to determine whether they are acceptable to participants,[2] safe, and reasonably practical to implement; and then assembles the components into an intervention package and evaluates the package by means of a randomized controlled trial (RCT). In a typical RCT, subjects are randomly assigned to one of two or more experimental conditions or "arms." In a two-arm RCT, subjects in the

[1]This book will use the term pilot study to refer to a study that is aimed at examining feasibility in preparation for a more formal study and not intended for hypothesis testing (Leon, Davis, & Kraemer, 2011).

[2]This book will use the term participant to refer to an individual who is taking part in an implementation of an intervention for clinical, as opposed to research, purposes and the term subject to refer to an individual who is taking part in, and providing data for, a research study.

treatment arm are provided with an intervention package to be evaluated, and subjects in the control arm are provided with a suitable comparison intervention, such as the current standard of care. If the difference between the treatment arm and the control arm is found to be statistically significant, then the intervention package is considered efficacious (if the RCT is conducted under controlled experimental circumstances) or effective (if the RCT is conducted under real-world circumstances). Often the investigator conducts mediation analyses of the data collected during the RCT, with the objective of revealing which variables mediate the effect of the treatment package on the outcome, thereby gaining an understanding of how the intervention operates.

1.1.1 An Approach Inspired by Ideas From Engineering

MOST offers a perspective on intervention development and evaluation that is different from the classical approach, because it has been inspired by ideas from engineering. Let us begin by considering how an industrial engineer might go about developing an improved process for manufacturing vehicle leaf springs, which are a part of the suspension system of most cars and trucks. Manufacturing may appear on the surface to have little relevance to intervention science, but closer examination shows that there are important conceptual similarities between what an industrial engineer sets out to accomplish when developing and evaluating a manufacturing process and what an intervention scientist sets out to accomplish when developing and evaluating an intervention. For now, there is no need to be concerned about details or definitions of terms; that will come later in this chapter and in the remainder of the book. The purpose of this introduction is to impart a conceptual feel for the difference between the way an engineer's training and an intervention scientist's training lead them to approach a problem and to illustrate how and why an engineering-inspired approach may have some value in intervention science.

Suppose the owners of a plant that manufactures leaf springs wish to improve the manufacturing process. There is always some variation in the length of the leaf springs when they come off the assembly line. The owners would like to improve the manufacturing process so that the leaf springs are closer to a particular desired ideal length. An industrial engineer, Dr. E, is brought in to study the problem.

Based on prior research and experience, Dr. E hypothesizes that the following components are critical in the leaf spring manufacturing process: furnace temperature, heating time, time on the conveyor belt, time in the high-pressure press, and range of quench oil temperatures. (This scenario and list of components are very loosely based on Pignatiello and Ramberg, 1985, as described in Wu and Hamada, 2011). In the current process, these components are set to the following levels: low furnace temperature, short heating time, short time on the conveyor belt, short time in the high-pressure press, and lower quench oil temperature range. Dr. E hypothesizes that the manufacturing process would be improved—that is, the leaf springs coming off the assembly line would be closer to the desired length—if the settings were changed so that the furnace temperature was higher; the heating time, time on

the conveyor belt, and time in the high-pressure press were longer; and the quench oil temperature range was higher.

If Dr. E had been trained as an intervention scientist, that training might suggest approaching this problem by relying primarily on the RCT. Dr. E would create a new manufacturing process involving higher furnace temperature; longer heating time, time on the conveyor belt, and time in the high-pressure press; and higher quench oil temperature range. Then the new process as a package would be compared to the standard of care, that is, the old process.

However, this is not how Dr. E would proceed. Let us consider why.

1.1.2 The Objective: Optimized Rather Than Best

Imagine Dr. E were to come up with the very best manufacturing process currently possible, one that produces leaf springs so uniform that differences from the target length are virtually undetectable, and proudly presents this to the plant owners. The plant owners ask how much the new manufacturing process costs, and Dr. E replies that the process costs $1 million per spring. From one perspective, this is undeniably an improved manufacturing process. However, from the plant owners' perspective, it is anything but an improvement. They inform Dr. E that to stay competitive in the industry, the company must contain the selling price of the leaf springs, and therefore the manufacturing cost cannot exceed $25 per spring. Thus although the new manufacturing process produces leaf springs that are uniformly close to the desired length, it is useless to the company because of its prohibitive cost. Dr. E would probably be sent packing!

In reality this misunderstanding would be unlikely to happen because, as an industrial engineer, Dr. E knows that the plant owners are primarily interested in a manufacturing process they can actually use. In other words, the objective is to arrive at an *optimized* manufacturing process rather than the best manufacturing process. An optimized manufacturing process represents the best expected manufacturing process that can be obtained *subject to constraints*, for example, constraints on the per spring cost. Soon after being hired, and before starting any research, Dr. E would discuss the key constraints with the plant owners. Constraints may be expressed in terms of money, time, personnel, equipment, or any other quantity. At this time the plant owner would tell Dr. E about the upper bound of $25 per leaf spring in manufacturing costs. Dr. E and the plant owners acknowledge that there may be manufacturing processes that are better in the sense that they produce more uniform leaf springs, but if they cost more than $25 per leaf spring, they are of no more than academic interest in this situation. An optimized manufacturing process will not produce leaf springs that are closest to the ideal that can be obtained in any absolute sense; rather, it will produce leaf springs that are the closest to the ideal that the plant can afford and thus will be genuinely practical.

1.1.3 The Kind of Information Needed for Optimization

Dr. E will conduct experimentation to obtain the information necessary to identify the optimized manufacturing process. In general, only experimental designs that enable Dr. E to address the following questions will be considered:

What are the size and direction of each component's effect?
Which components demonstrate a positive effect? A null effect? A negative effect?
Is the performance of one component affected by the presence or level of one or more other components?

There are several reasons why the answers to these questions are important for development of the optimized manufacturing process.

First, to ensure the efficiency of the new manufacturing process, Dr. E wishes to select only components and component levels that make a positive contribution toward obtaining leaf springs that are as close to the target length as possible. No resources in the new process are to be wasted on components that have very small or null effects, and it is particularly important not to include components and component levels that are counterproductive.

Second, in selecting components and component levels, Dr. E knows it is important to account for how one component may affect the performance of another. The effect of one component may be enhanced or undermined by the presence or level of one or more other components. For example, maybe when the manufacturing process allows more time on the conveyor belt, more time in the high-pressure press has a much larger effect on uniformity in the length of the leaf springs. A component may be effective only when certain other components are included or may be effective *unless* combined with certain other components. For example, maybe if the furnace temperature is set to low, the longer heating time results in leaf spring lengths that are closer to the ideal, but when the furnace temperature is turned to high, the longer heating time does not make any difference.

Third, Dr. E is mindful that different components may be associated with different costs. For example, suppose the plant owners have told Dr. E that the fuel needed to increase the furnace heat is expensive; by contrast, it does not cost much to increase time spent on the conveyor belt. To arrive at the best manufacturing process that can be obtained for a cost of $25 per leaf spring or less, it will be important to take cost differences in components into account. A very expensive component may have to demonstrate a correspondingly large effect on manufacturing precision to be included. This is another reason it will be important to be able to estimate both individual component effects and component interactions. It will also be important to have information on cost.

Fourth, in engineering there is an expectation that improvement is an ongoing process. This means that once the new manufacturing procedure is identified, almost immediately Dr. E will be expected to start working on producing the next one. In this way the manufacturing process will become incrementally better and better over time. To know where to focus efforts to create an even better manufacturing process

the next time, one must build on a previously obtained understanding of which components work well, which are weak and could be improved, which are not working at all or are counterproductive and should be abandoned, and which enhance or undermine the effectiveness of others.

1.1.4 Using Research Resources Strategically to Obtain the Needed Information

For all of the reasons listed above, Dr. E needs to be able to estimate the individual effects of components and how the components may affect each other (and, if not already known, the costs associated with each component). Dr. E must take stock of the amount and type of resources the plant owner is willing to spend on research to obtain this information. Resources might be money, time in the plant, technician time, space, materials, etc. Because different experimental designs can make very different resource demands, Dr. E will compare the resource requirements of several different designs, carefully selecting a highly efficient experimental alternative that makes use of whatever resources are available to obtain as much reliable information as possible. The idea is to work strategically to gather the most, and the highest quality, information that can be obtained with the available resources.

In approaching the task of developing the new manufacturing process, Dr. E would be unlikely to select a two-arm RCT or similar experimental design, because it would not yield the needed data. In a two-arm RCT, the components would be manipulated as a set. In the treatment group, all five components would be set to the higher/more intense level, and in the control group, all five would be set to the level specified in the current manufacturing process. In other words, in a two-arm RCT, the components all would be confounded with each other. To obtain the information needed, it is necessary to manipulate the components individually. Rather than using the RCT at this point, Dr. E will look for a highly efficient experimental design that can deliver a lot of scientific information about individual component effects in relation to the resources it requires.

Once Dr. E has selected an experimental design, conducted the experiment, and analyzed the data, the results will provide estimates of the effect of each of the five components and how each component's effects may be impacted by the presence or absence of other components. This information will form the basis for identifying which set of components and component levels produces the best outcome (i.e., leaf springs closest in length to the target) while staying within the limit of $25 per leaf spring in manufacturing costs.

After an optimized manufacturing process has been developed, Dr. E may directly compare its performance to that of the current process, if this is deemed necessary. In this case the final decision about whether or not the new process will replace the current one would be made on the basis of the outcome of this experiment.

1.2 Optimizing an Intervention: A Brief Hypothetical Example

Suppose an intervention scientist, Dr. B, wishes to develop and evaluate an approach for reducing viral load in HIV-positive individuals who drink heavily. Based on prior scientific literature, clinical experience, and a well-specified conceptual model (example presented in Collins, Kugler, & Gwadz, 2016, and reviewed in detail in Chap. 2), Dr. B has identified five intervention components that are hypothesized to be critical in helping HIV-positive individuals to reduce their drinking and improve their adherence to antiretroviral therapy (ART). They are (a) motivational interviewing to engage the participant in the process of examining his or her alcohol use and ART adherence, (b) peer mentoring to provide contact with a positive role model, (c) text messaging to provide access to a support system, (d) mindfulness meditation to improve mental health, and (e) behavioral skills training to improve behavioral skills for managing alcohol use and ART. Suppose there is currently no standard of care for the first four components, but physicians and HIV clinics routinely give HIV-positive individuals who drink heavily and have poor ART adherence a workbook to complete to help improve their behavioral skills.

The training of most of today's intervention scientists would suggest taking the classical approach. The five components would first be pilot tested and then assembled into an intervention package that would be evaluated in a two-arm RCT. Experimental subjects would be randomly assigned to either the treatment arm or the control arm. The treatment arm would receive motivational interviewing, peer mentoring, text messaging, mindfulness meditation, and behavioral skills training. The control arm would receive only the workbook. For the purposes of this example, the primary outcome variable is number of days of adherence to ART in the 30-day period following the conclusion of the intervention. This will be abbreviated *adhere*. After the RCT has been completed and the primary test of intervention efficacy or effectiveness has been performed, mediation analyses would probably be undertaken to provide a better sense of how the intervention worked.

The premise of this book and the companion volume is that the pace of progress toward better interventions can be accelerated by using MOST to develop, optimize, and evaluate interventions. MOST is similar to the approach taken by Dr. E to arrive at an optimized leaf spring manufacturing process but adapted for use in intervention science. Instead of using the classical approach, Dr. B can use MOST to develop, optimize, and evaluate the ART adherence intervention. The next section will illustrate how Dr. B's thinking parallels that of Dr. E. The discussion provides merely a thumbnail sketch; it is superficial and intended only to provide the reader with a conceptual sense of how MOST works. Moreover, it describes only *one way* that an investigator might apply MOST. It is important to realize that because MOST is a framework, not a specific procedure, different applications of MOST can proceed very differently. In particular, they can involve different approaches to experimentation, depending on what is most appropriate and efficient for that application. Much more detail on various approaches that can be used within the

MOST framework is provided in the subsequent chapters in this book and in the companion volume.

1.2.1 The MOST Perspective

Recall that Dr. B has identified five intervention components. Rather than immediately including all of these components in the intervention, Dr. B views the five components as a set of *candidates* to be selected for inclusion in the intervention. Like Dr. E, Dr. B will conduct experimentation to obtain the information necessary to identify the optimized manufacturing process.

Dr. B starts by taking stock of what resources will be available to implement the optimized intervention. Suppose Dr. B contacts several government-funded HIV clinics and engages in extensive discussions about what kind of intervention would be immediately scalable. It is determined that an intervention aimed at HIV-positive individuals who drink heavily that effectively improves adherence to ART and, thereby, reduces viral load would be immediately scalable if it cost no more than $500 per individual to implement. Thus, the clear goal is to develop the most effective intervention that can be implemented for $500 per individual or less. This will serve as Dr. B's definition of the optimized intervention, in other words, the *optimization criterion*.

Next, Dr. B turns to planning the experimentation that will be done to obtain the information needed for optimization. Dr. B is interested primarily in experimental designs that enable estimation of the size and direction of individual component effects and also assessment of whether the effect of one component is affected by the presence or level of any of the other components, for the following reasons.

First, to increase the efficiency of the new intervention, Dr. B wishes to select only components and component levels that are making a positive contribution toward increasing *adhere* (West, Aiken, & Todd, 1993). No resources in the new intervention are to be wasted on components that have very small or null effects, and it is particularly important not to waste resources by including components that appear to decrease *adhere*. The term resources here is broadly defined to include money, staff time or expertise, participant time or burden, space, equipment, or any other quantifiable resource.

Second, in selecting components and component levels, Dr. B knows that it is important to account for how one component may affect the performance of another. The effect of one component may be enhanced or undermined by the presence of one or more other components. For example, maybe use of mindfulness meditation enhances the effect of behavioral skills training because the meditation practice helps the individual be more open to the training.

Third, Dr. B knows that different components may be associated with different resource demands. Therefore, to increase the economy and scalability of the new intervention, it may be important to take these resource demands into account in relation to the contribution a component makes. For example, text messaging may be

relatively inexpensive, whereas motivational interviewing by a trained interviewer may be very expensive.

Fourth, like Dr. E, Dr. B knows that improvement is an ongoing process. Once the new ART adherence intervention has been identified, almost immediately Dr. B will start working on producing the next, even better one. To know where to focus efforts to create an even better intervention, Dr. B needs to build on a previously obtained understanding of which components work well, which are weak and could be improved, which do not work at all or are counterproductive and should be abandoned, and which enhance or undermine the effectiveness of others. In this way a coherent base of knowledge about what works and what does not work will be accumulated, and the ART adherence intervention will become incrementally better and better over time with repeated efforts to improve it.

Thus, to obtain the information needed in this phase of research, that is, the information needed to select what components/component levels will make up the optimized intervention, it is necessary to conduct experimentation that enables manipulation of the components individually. Dr. B takes stock of what resources are available to conduct this experimentation. Dr. B is in a somewhat different situation from Dr. E, in that it is necessary to write a grant proposal to obtain funding for the project. The proposal will be subject to whatever limitations the funding source may place on how much money can be requested. Rather than proposing use of the RCT design at this point, Dr. B will look for a highly efficient experimental design that can deliver a lot of scientific information for the resources it requires. Dr. B knows the grant proposal will be viewed as stronger by the review committee if it includes a clear rationale for why the experimental design selected is the one that will make the best use of the funds requested. Therefore, the proposal will contain a comparison of the resource requirements of a few experimental designs that could be used (comparison of resource requirements of different experimental designs is discussed in Chap. 6).

Once Dr. B has selected an experimental design, obtained the necessary funding, conducted the experiment, and analyzed the data, the results will provide estimates of the effect of each of the five components and how each component's effects may be impacted by the presence or absence of other components. This information will form the basis for identifying (a) which components should be eliminated from consideration and (b) out of those that remain, which set of components and component levels meets the optimization criterion, in other words, produces the best outcome on *adhere* without exceeding an implementation cost of $500.

Dr. B remains committed to evaluating the effectiveness of the optimized intervention as a package. Once the optimized intervention has been identified, Dr. B will directly compare its performance to that of a suitable control treatment in an RCT. The results will form the basis for the final decision about whether or not the new intervention is effective.

1.2.2 Gathering the Information Needed to Optimize the Intervention

To obtain the information required to decide what set of components and component levels should be selected to comprise the optimized intervention, Dr. B needs to gather several different kinds of information. Information is needed on the individual performance of each component. Information is also needed on how components perform together. This is represented statistically in interaction terms. Are there synergistic interactions between components—in other words, does one component enhance the effect of another? Are there antagonistic interactions—in other words, does one component undermine the effect of another?

An experiment designed to collect the information needed to optimize an intervention is called an *optimization trial*. In these two companion volumes, many pages are devoted to various approaches to experimentation that are appropriate for an optimization trial. In the example being discussed here, Dr. B needs to screen out components that do not show sufficient effectiveness and then, from the set of components remaining, select which components and component levels should be included in the optimized intervention. Dr. B needs to find an efficient experimental design that will provide the information needed to accomplish this.

For now, let us say that Dr. B decides to conduct a factorial experiment. Dr. B selects this experimental design because, as will be discussed at length in subsequent chapters in this volume, the factorial experiment is a highly efficient and economical way of assessing the performance of individual intervention components and determining how components may affect each other's performance. (The idea that factorial experiments can be highly efficient and economical may be counterintuitive to some intervention scientists. An introduction to factorial experiments, along with an explanation of their efficiency, is provided in Chap. 3). The experiment will include a factor corresponding to each component. Each factor will have two levels. The factors corresponding to the first four components can be set to no, that is, not included in the intervention, or yes, that is, included. The factor corresponding to the behavioral skills component can be at either a low level, consisting of the workbook only (i.e., standard of care), or a high level, consisting of the workbook plus training delivered by a behavioral skills counselor. Dr. B has selected *adhere* as the primary outcome variable.

In the course of conducting the experiment, Dr. B obtains information not only on the primary outcome and any secondary outcomes but also on implementation costs, because cost will be an important consideration when the final components and levels are selected. As explained above, Dr. B seeks an intervention that can be implemented for no more than $500 per person. Data will be collected on the cost of implementation of each component and whether there are economies of scale or additional expenses when certain components are combined.

1.2.3 From Experimental Results to Optimized Intervention

After conducting the factorial experiment, Dr. B analyzes the data using a standard factorial analysis of variance (ANOVA) to obtain estimates of the main effect of each component on *adhere* and of interactions between components. The main effects provide an assessment of the average performance of each component. The interactions provide an assessment of the extent to which components affect each other's performance. The ANOVA results form the basis for selecting components and component levels that will make up the optimized intervention. It is up to the investigator to use the information provided by the ANOVA to make decisions about which components and component levels will form the optimized intervention. This book follows an approach to decision-making based on the one outlined by Collins et al. (2014). The approach is described in Chap. 7.

In the approach to decision-making used in this book, Dr. B would first divide the components into a *screened-in set* and a *screened-out set*. Components are selected for the screened-in set because the experimental evidence suggests they have demonstrated an effect on *adhere* that is large enough to be considered important and is in the desired direction. A component may be selected for inclusion in the screened-in set because in the experiment it demonstrates an effect as an individual component or operates synergistically with one or more other components to enhance their effects. A component is selected for the screened-out set because the results of the experiment suggest it has only a weak effect on *adhere*, has a strong and positive main effect but performs poorly when combined with other effective components, or has an iatrogenic effect (i.e., the lower level performs better than the higher level).

Once the components have been sorted into the screened-in and screened-out sets, the investigator can decide which component levels will make up the optimized intervention. Why is this phrased in terms of selecting component levels rather than components? It may seem more natural to think in terms of selecting components for inclusion in the intervention, but in a sense one is always ultimately selecting component levels. To see why, compare the mindfulness meditation and behavioral skills training components. For the mindfulness meditation component, the two levels of the corresponding experimental factor are no and yes. Choosing the higher level, yes, means this component will be included in the intervention, and choosing the lower level, no, means it will not be included. For the behavioral skills training component, the two levels are workbook only and behavioral skills training plus the workbook. Here choosing the higher level means the intervention will provide both in-person behavioral skills training and a workbook, but choosing the lower level does not mean that behavioral skills training will be omitted from the intervention. Instead, choosing the lower level means that behavioral skills training will consist of the workbook only. This is why this book refers to selection of components and component levels for an intervention, rather than just selection of components. Later in this chapter and in Chap. 2 the concept of a component is discussed in more detail.

Now let us review how the optimized intervention is built by selecting component levels from the screened-in and screened-out sets. The components in the screened-out

set will all be set to the lower level in the intervention. The lower level may represent the absence of the component, or it may represent the inclusion of the component at a low level, depending on what levels were examined in the experiment.

Depending on the optimization criterion and the observed results, the optimized intervention may or may not include the higher level of all of the components in the screened-in set. If not for the need to arrive at an intervention that costs no more than $500 to implement, it would be appropriate for Dr. B simply to construct an optimized intervention that includes the higher level of all of the components in the screened-in set. This straightforward approach to optimization is suitable for situations in which cost is not an explicit consideration, but there is still a desire to avoid devoting resources—including participant time and energy—to inactive or counterproductive components.

In Dr. B's case, the optimization criterion calls for identifying the combination of component levels from the components in the screened-in set that is expected to produce the best outcome on *adhere* that can be obtained without exceeding the $500 limit. Dr. B can identify this set by (a) using the results of the ANOVA to arrive at the predicted outcome for each combination of component levels and (b) using the data that were collected on cost to determine which of these combinations can be implemented for no more than $500 per person. (Exactly how to do this will be discussed in Chap. 7.) The combination of components and component levels that produces the best expected outcome on *adhere* without exceeding the $500 cost limit comprises the optimized intervention.

Once the optimized intervention has been identified, it may then be evaluated in an RCT.

1.3 Definition of Optimization of an Intervention

Up to this point, optimization of interventions has been discussed without much specificity, but now a definition of optimization is needed. In this book optimization will be defined as follows:

> *Optimization of an intervention is the process of identifying an intervention that provides the best expected outcome obtainable within key constraints imposed by the need for efficiency, economy, and/or scalability.*

Let us examine this definition closely.

1.3.1 Optimization Is a Process: The Continual Optimization Principle

The continual optimization principle is one of the fundamental principles of MOST. To understand this principle, consider the development of consumer products, such as automobiles, appliances, and the like. When development has reached a point at

which the product is ready to market, the engineer does not declare "mission accomplished!" and walk away. Instead, work soon starts on next year's model, which will be better in a measurable, incremental way. For example, it may improve gas mileage by 3 miles per gallon, or include a new safety feature, or be more ergonomic.

The MOST framework provides an opportunity for intervention scientists to engage in the same sort of ongoing improvement. According to the continual optimization principle, optimization is a process of moving toward an ever-better intervention. Although in discussing an intervention that has been developed using MOST it may be correct to say it has been optimized, this does not mean the intervention cannot be improved further; saying an intervention has been optimized merely means it has been through one or more rounds of optimization. In fact, in theory an intervention could be optimized any number of times, with each round of optimization making it better and better by improving its effectiveness, increasing its efficiency or economy, or adapting and updating it in response to different or changing circumstances.

Suppose Dr. B has completed one round of MOST, that is, has optimized the ART adherence intervention and evaluated it in an RCT, and now has an effective intervention that can be implemented for $500 or less. According to the continual optimization principle, while the current optimized intervention is being implemented, Dr. B can use the MOST framework to begin thinking about what subsequent improvements can be made. Dr. B may wish to improve the effectiveness of the previous version of the intervention while staying within the $500 limit, perhaps by testing some new components that have been suggested by recent scientific literature. Or, the goal may be an intervention that comes as close as possible to the previous version but can be implemented for only $400; an intervention that is as effective as the previous version, but in addition to costing no more than $500, takes less than 240 min to complete; or some other quantifiable improvement.

How is Dr. B to proceed to develop the next, incrementally better version of the ART adherence intervention? Because MOST involves investigating the performance of individual intervention components and how components may affect each other's performance, it is straightforward for Dr. B to build directly on previous work. This previous work revealed which components worked well and which did not. This helps to illuminate the way forward. For example, suppose the peer mentoring component had an unacceptably small effect on *adhere* and was not included in the screened-in set, yet for theoretical reasons, peer mentoring remains an important part of the intervention in the eyes of Dr. B. In this case one next step would be to develop a revised peer mentoring component and test it, perhaps along with several other candidate additions to the intervention, in an experiment that enables assessment of the individual effects of each component and interactions between them.

Imagine that in the course of continual optimization, Dr. B and other scientists conduct more and more optimization trials and share the information gained. A body of knowledge will gradually accumulate about what intervention strategies work,

under what circumstances, and for whom, in the HIV field and other fields. The knowledge gained by one scientist can be built on by others, so that every new intervention is a measureable improvement on its predecessor. Scientists working in one domain will see what strategies work in others, providing valuable starting points for intervention development in new areas.

This is not intended to imply that it is always necessary to build directly on previous work, although frequently that is arguably the most efficient way to proceed. There are times in engineering and science when a paradigm shift comes along and a wholly new approach is proposed that is qualitatively different from what has gone before. In this case, even though all of the previously tested components may be discarded, it still makes sense to establish that the new approach is a measureable improvement over the old one along one or more specific dimensions.

1.3.2 Four Desiderata: Effectiveness, Efficiency, Economy, and Scalability

In this the desired characteristics of an intervention are reviewed. In the classical approach, the emphasis is primarily on effectiveness. By contrast, in MOST four desiderata for interventions are emphasized: effectiveness, efficiency, economy, and scalability. How much emphasis is placed on each relative to the others depends on the situation.

Effectiveness. Effectiveness is a critically important concept that can be manifested in several different ways. In this book the effectiveness of an intervention or a component of an intervention is defined as the degree to which the intervention or component produces an outcome in the desired direction. It is frequently important to make a distinction between effectiveness and efficacy, as suggested by Flay (1986), who defined these terms as follows:

> Efficacy trials provide tests of whether a technology, treatment, procedure, or program does more good than harm when delivered under optimum conditions... Effectiveness trials provide tests of whether a technology, treatment, procedure, intervention, or program does more good than harm when delivered under real-world conditions. (p. 451)

This is a helpful distinction. This book takes the position that both efficacy and effectiveness are important, but establishing effectiveness is the ultimate goal. MOST can be used in both efficacy and effectiveness trials and can even be used to reduce the loss of intervention potency that is often observed between efficacy and effectiveness trials of the same intervention (e.g., Caldwell et al., 2012). To avoid the tedium of having to repeat "efficacy or effectiveness," this book will simply refer to effectiveness when what is said applies to both and highlight the distinction between efficacy and effectiveness when it is important to do so.

Efficiency, economy, and scalability. Efficiency is the degree to which the intervention produces a good outcome while avoiding wasting money, time, or any other valuable resource. Economy is the degree to which the intervention

produces a good outcome without exceeding budgetary constraints (where a budget may be placed not only on money per se but time or any other resource) on implementation and the degree to which it offers a high degree of effectiveness in exchange for the resources required to implement it. Efficiency and economy are closely related but distinct. It is possible for an intervention to be efficient, that is, to be made up primarily of strongly performing components, but not to be economical because it is too expensive to be widely implemented. An intervention may be economical in the sense that it is affordable, but not efficient in the sense that it includes inactive or poorly performing components.

Scalability is defined in this book as the degree to which the intervention can be implemented widely in real-world settings exactly in the form in which it was evaluated, without the need for ad hoc adjustments. In other words, if an intervention is scalable, those tasked with implementing the intervention after its evaluation will not need to take measures to reduce its length, complexity, staff or participant burden, etc. To be scalable an intervention often has to be efficient and economical; sometimes it may need to be simple and straightforward as well. One goal of MOST is to produce interventions that are immediately scalable, that is, that can be implemented as is as soon as their effectiveness has been scientifically demonstrated in an RCT.

As will be discussed in Chap. 7, once an optimization trial has been completed, the information gathered can be used to develop different interventions that are scalable under different circumstances. In the example, Dr. B has established that to be scalable, the intervention must cost no more than $500 per participant. Suppose a health maintenance organization (HMO) that serves a resource-poor community has determined that to be scalable for them, an intervention like the one Dr. B is developing can cost no more than $400 per participant. The results of Dr. B's optimization trial could be used to identify the intervention with the best expected outcome that can be obtained for implementation costs of no more than $400 per person, provided that the results of the optimization trial can reasonably be assumed to generalize to the HMO's patient population.

1.3.3 Trade-Offs Among the Desiderata

MOST demands clarity about the relative importance of the four desiderata because, as will be discussed in detail in Chaps. 2 and 7, this clarity is necessary in order to proceed with optimization. In a public health crisis, say one involving a highly contagious and lethal disease, effectiveness might dwarf all other considerations, at least in the short term until the situation begins to be brought under control. However, in most cases, some or all of the remaining desiderata will be important too.

All else being equal, if there were unlimited resources for implementation—an intervention could cost any amount of money, take any amount of staff and participant time, and consume any amount of other resources—in most cases the result

would be a more effective intervention. However, once considerations of efficiency, economy, and scalability are introduced, constraints are imposed that are likely to reduce effectiveness to some extent. Any constraints on the amount of money, time, or other resources available to implement an intervention may mean that an effective component must be eliminated to enable implementation within the allocated budget. This means there is a fundamental tension between effectiveness on the one hand and efficiency, economy, and scalability on the other. From one perspective, optimization is the process of establishing clear trade-offs and thereby resolving this tension.

The definition of optimization refers to the best expected outcome *obtainable within key constraints.* Because of the necessity for trade-offs, this is not the same as the best outcome possible in an absolute sense. Let us return for a moment to the hypothetical example. To arrive at the optimized ART adherence intervention, Dr. B starts with the screened-in set and then looks for the subset of components and component levels that is expected to produce the best outcome on *adhere* that can be obtained for an implementation cost of $500 or less. Depending on the cost associated with each component, to arrive at an intervention that will cost $500 or less, Dr. B may have to reject components that have demonstrated an effect on *adhere*. For example, suppose based on the results of the optimization trial, motivational interviewing is selected for the screened-in set. Further suppose this component is expensive to implement, because of the costs associated with training the interviewers and monitoring their performance and paying them to conduct a lengthy session with each participant. It may be that because of the high cost of motivational interviewing, the combination of components and component levels producing the best expected outcome for no more than $500 will omit this component. Including motivational interviewing would produce a more effective intervention but one that would exceed the implementation budget. As this example illustrates, an intervention that meets necessary standards of, in this instance, economy may not produce the best outcome in an absolute sense.

The quest for the ideal of absolute effectiveness without regard for practical considerations may be quixotic if the resulting intervention is so inefficient, expensive, or complex that it is never implemented widely in its intended form—or is never implemented widely at all. The reality is that nearly every intervention must operate within constraints. Optimization explicitly recognizes both the ideal and this reality and works within the reality to come as close as possible to the ideal.

1.4 The Resource Management Principle

Two fundamental principles underlie MOST. One is the continual optimization principle, discussed above. The other is the resource management principle.

The resource management principle concerns how the investigator uses whatever resources are available for conducting experimentation during the optimization phase. According to the resource management principle, an investigator using

MOST must strive to make the best and most efficient use of available resources when obtaining scientific information. Money is the primary resource because it typically can be exchanged for other resources, but time, personnel, equipment, space, experimental subjects, or anything else needed to obtain the information necessary for intervention optimization can also be considered resources.

Note that the resource management principle specifies *available* resources, meaning resources that the investigator has or can reasonably expect to obtain by, for example, writing a successful grant proposal. Of course, having more resources is always better in any scientific endeavor, because it means the investigator can obtain more and better information. But MOST can be applicable even where resources are limited. Considering the resource management principle can help the investigator take a realistic and strategic look at how best to use limited resources to move intervention science forward. This will be discussed at length in Chap. 6. As the reader of this book will see, although MOST does not necessarily require an increase in resources, in many cases it requires a realignment of resources; in other words, the investigator working within the MOST framework uses resources somewhat differently than an investigator using the classical approach.

1.5 Some Differences Between the Classical and MOST Perspectives

By now it may be evident to the reader that there are some fundamental differences in perspective and priorities between the classical approach and MOST. These are summarized in a conceptual way in Table 1.1. Later in this chapter, there will be a more specific discussion of the three phases of MOST.

The first difference between the approaches is their objectives. The objective of the classical approach is to develop an intervention that demonstrates a statistically and clinically significant effect in an RCT. The objective of MOST is to build an intervention that meets specific predetermined standards and demonstrates a statistically and clinically significant effect in an RCT. As will be discussed further, the predetermined standards must be clearly operationalized. These standards may pertain to any or all of the desiderata: effectiveness, efficiency, economy, and scalability.

An investigator using the classical approach proceeds differently from an investigator who uses MOST. As discussed above, the investigator who uses the classical approach identifies a set of intervention components and after pilot testing immediately assembles the components into an intervention and conducts an RCT to evaluate the intervention as a package. The investigator using MOST does not go directly to an RCT after pilot testing of the intervention components and instead conducts an optimization trial aimed at gathering the information needed to optimize the intervention. Only after the intervention has been optimized does the investigator consider evaluating it in an RCT. How to conduct the research necessary for optimization is a major focus of this book and the companion volume.

Table 1.1 Some differences in perspective between the classical approach and MOST

	Classical approach	MOST
Objective	To develop an intervention that demonstrates a statistically and clinically significant effect in an RCT	To build an intervention that meets specific predetermined standards of effectiveness, efficiency, cost-effectiveness, and/or scalability and demonstrates a statistically and clinically significant effect in an RCT
Next steps after identification and pilot testing of components	Intervention is assembled and then evaluated as a package in an RCT	Optimization trial is conducted; an optimized intervention is built
Experimental designs used	Primarily the RCT	For the optimization trial, experimental designs selected based on resource management principle; for evaluation of intervention as a package, primarily the RCT
Examination of effectiveness of individual components	Conducted primarily via post hoc analyses on data from RCT	Conducted primarily via experimental manipulation of components
Examination of interactions between intervention components	RCT does not permit this	Experimental designs for optimization trial selected to enable this wherever possible
Inclusion of inert or counterproductive components or unnecessarily high component levels	Generally tolerated as long as overall effectiveness of intervention can be demonstrated	Generally not tolerated because this reduces the efficiency of the intervention
Scalability of intervention	Usually dealt with after evaluation of intervention, sometimes via ad hoc modifications	Intervention built for immediate scalability

The two approaches differ in the kinds of approaches to experimentation that are used. The classical approach relies primarily on the RCT and its variants, generally to the exclusion of other kinds of experimental designs. By contrast, MOST explicitly recognizes that it is unrealistic to expect that the RCT, as valuable as it is for evaluation, is always the experimental design of choice for every research question. Thus for use in the optimization trial, which can vary considerably across applications in different settings and content areas, MOST calls for selecting from among a broad array of approaches to identify the one that is most appropriate and efficient for the particular research questions at hand, in other words, to select a design based on the resource management principle. Which is the most appropriate and efficient experimental design will depend on the exact goals of the optimization, the type of intervention that is to be optimized (this is discussed briefly below and in more detail in Chap. 8), the research questions at hand, and the resources available to conduct the research. After the intervention has been optimized, MOST relies on the RCT to address the more circumscribed question of whether the optimized intervention demonstrates a statistically and clinically significant effect.

Perhaps because of its heavy reliance on the RCT, the classical approach generally has focused on establishing whether the intervention as a package has a significant effect, rather than looking inside the intervention to understand whether and how individual components operate. When the effectiveness of individual components has been investigated, the emphasis has been on secondary analyses of the data from RCTs, such as "dose-response" analyses and mediation analyses. Post hoc "dose-response" analyses rely on naturally occurring variation in subject adherence to manipulate the dose. However, this naturally occurring variation is entirely due to self-selection; subjects decide whether to comply with any particular demand of an intervention. Thus, there are many inferential problems with such analyses. Mediation analyses on data from an RCT can reveal which mediators were affected by the treatment package and which, in turn, affected the outcome. However, they cannot reveal which components affected which mediators, so their use in determining which components had an effect is limited. In MOST, by contrast, investigation of the effectiveness of individual components is not coupled with the evaluation of the intervention as a package, but instead is accomplished in an optimization trial involving direct experimental manipulation of components.

Again perhaps because of the heavy reliance on the RCT, interactions between intervention components are seldom examined in the classical approach. This means it is unknown whether a particular component boosts or undermines the effect of another component or whether a set of two or more components should always be included together in an intervention or never included together. In MOST examination of interactions between components under consideration for inclusion in an intervention is a high priority, so experimental designs that enable this are employed wherever possible.

The two approaches also differ in their perspectives on considerations other than the effectiveness per se of the intervention. In the classical approach, inclusion of inert or counterproductive components, or inclusion of unnecessarily high levels of components, is certainly not desired. However, it is tolerated as long as the intervention as a package demonstrates an overall effect. Any adjustments to the intervention to remove inert or counterproductive components would be done after the RCT, probably informed by mediation analyses. By contrast, in MOST there is a low tolerance for inert or counterproductive components and unnecessarily high component levels. In fact, an important purpose of the optimization trial is often to identify such components and component levels so they can be eliminated from the intervention being developed.

As has been discussed, practical constraints frequently influence the success of translation of an intervention to its intended setting in homes, schools, communities, or health care. An intervention that is too expensive, lengthy, complex, or burdensome for participants or staff has little chance of being implemented widely or at least little chance of being implemented widely as it was designed and evaluated. In the classical approach, considerations related to scalability usually are taken most seriously after the intervention has been evaluated. At this point it is too late to do much about scalability. Any significant post-evaluation revision to the intervention to make it cheaper, shorter, simpler, or less burdensome will render it a different

intervention from the one that was evaluated. The revised intervention may or may not exert a treatment effect comparable to the original evaluated version. By contrast, in MOST, key factors expected to affect scalability can be built in from the outset, with the goal of immediate scalability of the optimized intervention. For example, suppose clinic staff say that they can devote at best only 30 min from their work day to implement a particular intervention. MOST suggests that under these circumstances, it makes sense to build the intervention to achieve the best expected outcome that can be obtained without demanding any more than 30 min of clinic staff time, so that as soon as it is evaluated, it will be immediately scalable.

1.6 Definitions of Some Important Terms

This book draws on and integrates ideas from intervention science, statistics, and engineering, as well as other fields. Across different fields the same term may be used to refer to different things, and different terms may be used to refer to the same thing. In an effort to maintain clarity, for the purposes of this book and the companion volume, important terms are defined explicitly. Below definitions of several terms are provided; other definitions are provided as they become relevant in later chapters. A glossary of terms appears at the end of this book.

1.6.1 Design

In this book the word "design" will be used in three different ways: experimental design, research design, and intervention design. To avoid confusion, this book and the companion volume will specify intervention design, experimental design, or research design, unless the immediate context makes it clear.

Experimental design refers to the design of an experiment, in this context, to gather information needed to develop, optimize, or evaluate an intervention. Here the word "experiment" refers to manipulation of one or more independent variables for the purpose of empirically observing the effect on an outcome variable. Research design refers more broadly to the specific details of the procedures to be used in a study, such as selection and timing of measures or inclusion criteria for subjects. If the study includes an experiment, then the experimental design is one aspect of the research design.

Intervention design refers to the specific details of the approach taken by an intervention, including the intervention components (e.g., motivational interviewing, peer mentoring), the settings of the components (e.g., two 1-hour sessions of motivational interviewing; weekly half-hour sessions with a peer mentor for 5 weeks), any eligibility requirements for participants (e.g., must be HIV-positive and drink the equivalent of at least 14 grams of pure alcohol at least five times per week), and so on.

Fixed vs. Adaptive Intervention Designs Intervention designs fall in two general categories. One is the fixed intervention. In a fixed intervention, all participants are offered the same set of intervention components in a uniform manner. This intervention design does not include any planned variability in the approach, content, or dosage of the intervention across participants. The intervention Dr. B is developing is fixed; all participants are to receive the same intervention components and levels.

Another category of intervention design is the adaptive intervention. The Almirall, Nahum-Shani, Wang, and Kasari (2018) chapter in the companion volume defines adaptive interventions as follows:

> An adaptive intervention is a sequence of pre-specified decision rules that can be used to guide whether, how, or when—and based on which measures—to alter an intervention or intervention component (e.g., treatment type, duration, frequency or amount) at critical decision points during the course of care.

In other words, in an adaptive intervention, the content, dose, or approach of certain aspects of the intervention are varied based on pre-specified decision rules. The decision rules determine how the intervention will be varied in response to measured tailoring variables. For example, suppose an adaptive version of Dr. B's intervention is developed, with the objective of achieving better HIV medication adherence by providing additional treatment to individuals who do not become adherent after an initial treatment. The tailoring variable is ART adherence assessed at 45 days, via a report provided by a medication event monitoring system (MEMS®) cap on the participants' pill bottles. Those who have been at least 80% adherent are considered to be adherent; all others are considered non-adherent. An adaptive intervention might involve the following set of decision rules:

> *All participants initially are provided with behavioral skills training, a weekly meeting with a peer mentor for four weeks, and text messaging for the entire duration of the intervention. At 45 days, those who are considered to be adherent step down to daily text messaging only. Those who are considered non-adherent are provided with four more weeks of peer mentoring and one session of motivational interviewing, in addition to the ongoing text messaging.*

The difference between fixed and adaptive intervention designs is important, because the experimental design chosen for the optimization trial may be different depending on whether the intervention to be optimized is fixed or adaptive. This will be discussed briefly below, in more detail in Chap. 8, and in still more detail in two chapters in the companion volume: Almirall, Nahum-Shani, Wang, and Kasari (2018) and Rivera, Hekler, Savage, and Downs (2018). The focus of the present volume is primarily fixed intervention designs. Everything that is covered in this book provides a necessary foundation for learning about optimization of adaptive interventions.

1.6.2 Component

Up to now the components of interventions have been discussed in conceptual terms. Let us now be a bit more specific about what is meant by an intervention component.

The definition of intervention component used in this book may seem a bit circular, but it is very practical: an intervention component is any part of an intervention that can be separated out for study. This means that if it can be separated out for experimental manipulation, and doing so will provide the answer to a research question and thereby potentially help improve the intervention, it is a component. Components are the building blocks of interventions.

A *candidate component* is a component from which a level (e.g., no or yes, off or on, low intensity or high intensity) is to be selected for inclusion in an intervention. This term may refer to any type of component.

Content components, the "active ingredients" (Michie et al., 2013, p. 82) of interventions, are perhaps the first type of component that will come to mind for most readers. As the name implies, these components make up the content of an intervention; that is, they are aimed at the behavioral or biological processes being intervened on. For example, in Dr. B's hypothetical ART adherence intervention described above, motivational interviewing, peer mentoring, text messaging, mindfulness meditation, and behavioral skills training are all content components. The components in the example are all behavioral, but content components can be pharmaceutical, medical, or surgical. For example, Piper et al. (2016) examined six components under consideration for inclusion in a smoking cessation intervention. Two of these were pharmaceutical content components: use of a nicotine patch and use of nicotine gum during the 3 weeks preceding the quit date. Investigators looking for a starting point to identify content components for behavioral and biobehavioral interventions are referred to the excellent behavior change taxonomy work of Michie and colleagues (e.g., Michie et al., 2013).

Engagement/adherence components are aimed at ensuring that participants remain engaged in the intervention for its entire duration, carefully follow the intervention's instructions, and otherwise comply with its requirements. For example, an electronically delivered intervention may incorporate a game of some kind, such as providing a brief joke every day the participant meets an intermediate goal. Involvement of community volunteers or peer leaders to provide encouragement and mentoring from someone who has "been there" may be an engagement/adherence component.

Schlam et al. (2016) examined five components under consideration for inclusion in a smoking cessation intervention, three of which concerned adherence to the recommended regimen of nicotine replacement therapy. The components were (a) counseling to promote adherence, (b) automated telephone calls to promote adherence, and (c) electronic medication monitoring with feedback and counseling. Because barriers to participation can interfere with engagement and adherence, some

intervention components may focus on removing these barriers. Provision of transportation and child care to enable participation in an intervention are examples of engagement/adherence components. In some cases, the primary purpose of engagement/adherence components may be keeping participants interested in the intervention so that they continue with it.

Fidelity components are aimed at maintaining a high level of fidelity of intervention delivery. Such components are usually aimed at those who deliver the intervention or the environment in which the intervention is to be delivered, rather than the individuals who are the target of the intervention. For example, Caldwell et al. (2012) described a study examining three components hypothesized to affect the fidelity of delivery of the HealthWise intervention, an intervention developed for South African schoolchildren to prevent drug abuse and risky sex. The content of the HealthWise intervention had been evaluated in a previous study. In the Caldwell et al. study, two of the components, (a) enhanced teacher training and (b) structure, support, and supervision, were aimed at the teachers who delivered the intervention. A third component was aimed at (c) enhancing the school environment to make it more welcoming to and supportive of HealthWise.

In an adaptive intervention, there may also be components corresponding to aspects of adaptive intervention design such as decision points, tailoring variables, and decision rules (Collins, Nahum-Shani, & Almirall, 2014). This is discussed further in Chap. 8. In general, it is not necessary to categorize intervention components, and some components can arguably be labeled more than one way. The purpose of this section has been to encourage broad thinking about the types of components that can play different roles in an intervention and to assert that any of them can potentially be examined for the purpose of optimization.

1.6.3 Multicomponent Interventions

The focus of this book is on multicomponent interventions, which involve a strategy made up of more than one, usually numerous, tactics aimed at achieving the intervention's end goal. Nearly every behavioral and biobehavioral intervention is multicomponent. Even biomedical interventions that at first appear to have only a single component, such as a surgical procedure or a pharmaceutical, can be considered multicomponent when behavioral and biobehavioral considerations that may affect the success of the intervention are taken into account. For example, ART can be considered a single-component intervention to reduce HIV viral load. However, to reduce viral load to undetectable levels, ART requires steady and careful adherence. This suggests that simply providing patients with ART may not be sufficient to reduce viral load; a multicomponent intervention like the one Dr. B is developing may be a more effective way to administer ART.

1.7 The MOST Framework: The Three Phases

This section and the next provide an overview of the three phases of the MOST framework, represented in the flow chart depicted in Fig. 1.1. The remaining chapters in this book and those in the companion volume go into much more detail about MOST.

In Fig. 1.1 each rectangle represents one of the phases of MOST, namely, preparation, optimization, and evaluation. These three phases are reviewed in this section. The figure also contains a diamond representing a decision point and two arrows representing possible ways of returning to the preparation phase. These are reviewed in the next section.

1.7.1 The Preparation Phase

As Fig. 1.1 shows, MOST begins with the preparation phase (discussed in more detail in Chap. 2). The purpose of this phase is to lay the groundwork for optimization of the intervention.

Development or revision of a detailed conceptual model occurs during the preparation phase. This model provides the foundation for critical decisions concerning intervention development and research design, including experimental design selection. The starting point for development or revision of the conceptual model is a review of all existing information. This information will come primarily from empirical scientific literature and theory but may also be drawn from other sources, such as secondary analyses of data from prior experiments or clinical experience. A good conceptual model serves two purposes. One purpose is to depict the causal process that produces the behavioral or biological process to be intervened on. The second purpose is to specify where and how the components of the

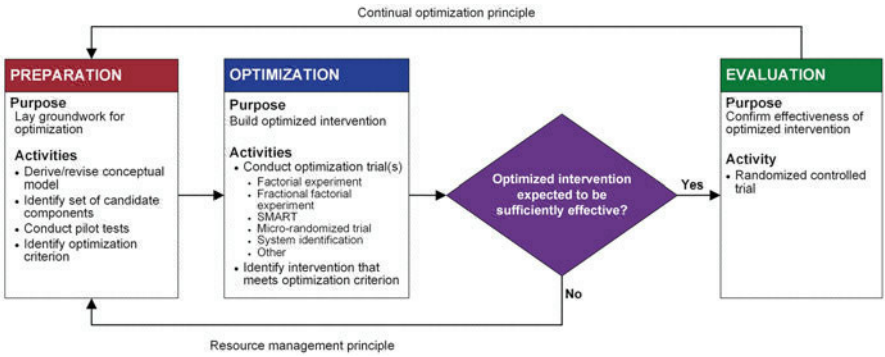

Fig. 1.1 Flow chart of the three phases of the multiphase optimization strategy (MOST)

intervention affect this causal process and how and why each component is expected to change the behavioral or biological process.

In some cases, all of the components featured in the conceptual model are selected for examination in the optimization phase, in other words, become part of the set of candidate components. In other cases, the effectiveness of some of the components may have been satisfactorily established in prior literature, and therefore it is a given that these components are to be included in the intervention. There may be no need to examine such components in the optimization phase. Or, the investigator may wish to determine whether and how the previously examined components interact with the candidate components, which would require examining them experimentally along with the others.

At this point it is necessary to determine how the components will be implemented and, usually, to pilot test the components and their implementation. After the pilot testing and any subsequent revision of components, there is one more necessary step in the preparation phase: identification of an optimization criterion. Recall the definition of optimization provided above:

> *Optimization of an intervention is the process of identifying an intervention that provides the best expected outcome obtainable within key constraints imposed by the need for efficiency, economy, and/or scalability.*

The above definition uses the terms "best expected outcome" and "constraints." The optimization criterion, which was discussed briefly earlier in this chapter, provides an operational definition of best expected outcome and also specifies the key constraints that are to be considered in optimization. In the leaf springs example, the optimization criterion is "leaf springs closest to the desired standard length that can be obtained without exceeding a manufacturing cost of $25 per leaf spring." The best outcome is "leaf springs closest to the desired standard length;" the specified constraint is on cost: the leaf springs must be produced "without exceeding a manufacturing cost of $25 per leaf spring." The ART adherence example used a conceptually similar optimization criterion, "largest expected value of *adhere* that can be obtained for an implementation cost of no more than $500 per person." There are many different types of optimization criteria; this is discussed further in Chaps. 2 and 7.

1.7.2 The Optimization Phase

The next phase of MOST is optimization. As the name implies, the purpose of this phase of MOST is to build an optimized intervention by selecting components and component levels from the set of candidates identified in the preparation phase. The decisions about selection of components and component levels are based on empirical data obtained by means of one or more carefully controlled and adequately powered optimization trials.

Choosing an Approach from the MOST Optimization Phase Toolbox There are many different approaches to experimentation that can be used in the optimization

phase. Factorial experiments, fractional factorial experiments, sequential, multiple assignment, randomized trials (SMARTs), micro-randomized trials, system identification, or any other suitable approach can be considered part of what Almirall, Nahum-Shani, Wang, and Kasari (2018) in the companion volume call the MOST optimization phase toolbox. MOST does not require any particular approach to experimentation in the optimization phase, only that the approach selected from the toolbox is the best one according to the resource management principle, that is, it is the most efficient way to obtain the information needed for optimization. The choice of approach depends on the type of intervention that is to be optimized, the exact empirical information that is needed, and the resources that are available to conduct the experimentation.

This volume emphasizes optimization of fixed interventions (the difference between fixed and adaptive interventions was discussed above). Typically, although not invariably, the most efficient way to obtain the data needed to optimize a fixed intervention is via a factorial or fractional factorial experiment. Chapters 3, 4, 5, and 6 cover these experimental designs, other designs that may be useful, and how to apply the resource management principle to determine which design is the most efficient for a given application. Optimization of adaptive interventions frequently requires an experimental approach other than the traditional factorial or fractional factorial design. This is where, depending on the type of adaptive intervention being developed, SMARTs, micro-randomized trials, and system identification may merit serious consideration. Optimization of different types of adaptive interventions is introduced in Chap. 8. Optimization of particular types of adaptive interventions is discussed in more depth in the Almirall, Nahum-Shani, Wang, and Kasari (2018) and Rivera, Hekler, Savage, and Downs (2018) chapters in the companion volume. Readers who are primarily interested in adaptive interventions are strongly encouraged to read Chaps. 3, 4, 5, and 6 in the current volume, because the information contained in these chapters provides a necessary foundation for understanding optimization of adaptive interventions. In particular, SMARTs and micro-randomized trials are closely related to the factorial experiment.

Decision-Making Once the experiment has been conducted, the data are analyzed in whatever manner is appropriate given the experimental design. Then decisions about the composition of the optimized intervention are made. These decisions are based on the experimental results, along with the optimization criterion that was selected in the preparation phase. As mentioned above, Chap. 7 discusses how to identify the components and component levels that will make up the optimized intervention.

Multiple Trials Within a Single Optimization Phase Under some circumstances, development of a highly effective, efficient, economical, and scalable intervention is best facilitated by conducting a series of trials within a single optimization phase. For example, sometimes it is possible to build a strong set of components by conducting a series of optimization trials in an iterative fashion. In this approach, the experiment is used in the usual manner to determine which components are performing

satisfactorily and which are performing unsatisfactorily. However, instead of building an optimized intervention after a single experiment and proceeding to the evaluation phase of MOST, an investigator using this iterative approach would revise the unsatisfactory components (or possibly replace them with new components) and then reevaluate them in a subsequent experiment. Once a set of satisfactory components has been arrived at, an optimized intervention is constructed in the usual manner, and the evaluation phase is begun. This approach has the potential to enable rapid progress in intervention development. An example of an iterative approach can be found in the Kugler, Wyrick, Tanner, Milroy, Chambers, Ma, and Collins (2018) chapter in the companion volume.

Depending on the situation, several optimization trials may be conducted. It may be expedient to use a measure of the mediator targeted by each component (see Chap. 2) as an outcome for that component, rather than the outcome of ultimate interest. Of course, MOST can help pinpoint which components are failing to have the intended effect on a mediator, but it is of limited use in determining exactly how to revise those components. Valuable leads for how to revise components may be gathered from collection of qualitative data, using sources such as focus groups and expert advisory panels.

A sequence of trials with different objectives may be conducted within a single optimization phase. Suppose an investigator who wishes to develop and optimize an adaptive intervention has the philosophy that, as a starting point, all of the intervention components to be included in the adaptive intervention should have a detectable overall effect when used in a fixed intervention. In this case the optimization phase could start with an experiment to identify which of a set of fixed components demonstrate a detectable overall effect. Then one or more subsequent experiments, perhaps SMARTs, could be conducted within the same optimization phase to build the adaptive intervention. The SMARTs could establish, for example, which of the components/component levels selected based on the first experiment represent the best initial treatment and which should be offered subsequently to those who respond to the initial treatment and to those who do not respond.

A single optimization phase of MOST can incorporate as many experiments as resources permit. To conduct several experiments within one optimization phase, it is necessary to have ready access to enough research subjects to provide sufficient power for the desired number of experiments. In addition, for each experiment the overall time frame must allow enough time for subject recruitment, conducting the experiment, data analysis, and planning of the next experiment based on the results. Investigators considering multiple experiments within a single optimization phase may be faced with the dilemma of which of two courses of action is a better use of resources within a given funding cycle: (a) completing the evaluation phase and conducting an RCT or (b) extending the optimization phase to include an additional experiment and applying for future funding to conduct the evaluation phase in a subsequent study. The former course of action has the advantage of providing a more definitive answer about the effectiveness of the optimized intervention, whereas the latter has the potential to produce an intervention that is more effective, efficient, economical, and/or scalable in the long run.

1.7.3 The Evaluation Phase

Once the optimized intervention has been identified, the investigator moves to the evaluation phase of MOST. The purpose of this phase is to confirm the effectiveness of the optimized intervention by means of an RCT. The objective of the RCT is to enable the investigator to decide whether the optimized intervention has a statistically and clinically significant treatment effect. Here effectiveness is expressed in terms of the size of the treatment/control difference. The treatment is the optimized intervention, and the control group could be current standard of care, a wait-list control, or any other suitable comparison group. If the results of the RCT indicate that the optimized intervention has a statistically and clinically significant effect, then the intervention may be released, in whatever manner is appropriate, for implementation in the intended setting.

1.8 Reasons for Returning to the Preparation Phase

1.8.1 Returning to the Preparation Phase Immediately After the Optimization Phase

The diamond in Fig. 1.1 represents a decision that must be made between the optimization and evaluation phases. The decision is whether or not the optimized intervention is expected to be sufficiently effective to justify continuing on to the evaluation phase.

It is possible for an intervention to be optimized and yet have an effect that is not expected to be large enough to be likely to achieve statistical significance in a reasonably sized RCT or to make much of a difference clinically. To understand how this can happen, let us return to the example. Suppose Dr. B identifies the set of intervention components that produces the best outcome that can be obtained without spending more than $500 per person, that is, optimizes the ART adherence intervention.

Now consider three different scenarios.

In the first scenario, the optimized intervention comprises a set of components that has many very potent members with strong individual and combined effects. In this happy scenario, it is likely that the optimized intervention will produce a statistically and clinically significant treatment effect when examined in an adequately powered RCT, and so the investigator would be justified in moving on to the evaluation phase of MOST.

In the second scenario, none of the components under consideration have very strong effects. It is possible to identify the best outcome that can be obtained without exceeding $500, but this best outcome is not likely to be much better than would be expected from a reasonable control or comparison treatment, and the cost is expected to be roughly the same. In this case, going to the trouble and expense of an RCT makes little sense. The resource management principle suggests that the resources

that would have been used in an RCT can more profitably be used to revise the conceptual model and develop and test a new set of components, with the objective of arriving at a new and more effective optimized intervention.

The third scenario is a bit more complicated. Here some of the components under consideration do have substantial effects, but the best combination of components that can be implemented within the $500 limit set by the insurer happens to be made up primarily of components with small effects. In other words, the optimized intervention, the one that produces the best outcome subject to the $500 constraint, is not very potent; there are combinations of components that do exert a potent effect, but they exceed the upper limit on cost. One possibility would be for Dr. B to ask the insurer whether the upper limit on cost can be raised, so that a more potent optimized intervention can be identified. A related possibility would be to suggest to the insurer that cost-effectiveness might be used in the optimization criterion rather than a fixed upper limit on cost. In other words, the optimized intervention would be the most cost-effective intervention, irrespective of absolute cost. These alternatives would be workable only if the insurers were willing to accept the idea of an intervention that will cost more than $500.

If the idea of a costlier intervention is acceptable, there is no need to redo the optimization trial. A new optimized intervention can be identified using the new criterion and the results of the experiment that was done to examine the existing components. Assuming the new optimized intervention looks more promising, it can be evaluated via an RCT. However, if the optimization criterion cannot be changed, it will be necessary to return to the preparation phase.

Whenever the results of the optimization phase suggest that the optimized intervention is likely to have an effect that will not achieve statistical or clinical significance, the resource management principle and common sense both suggest that it would be a poor use of resources to continue to the evaluation phase and subject this intervention to an RCT. Instead, it would be better to take the resources that would have been spent on an RCT and use them to return to the preparation phase to begin a new cycle of MOST focused on identifying some new, more effective components and, ultimately, arriving at a more effective intervention.

This option is represented in Fig. 1.1 by the arrow leading from the diamond immediately after the optimization phase back to the preparation phase. The investigator who follows this arrow is starting a new cycle of MOST but is not starting from square one. Whatever has been learned in the current cycle of MOST can be built upon and used to illuminate the way forward. Any components that performed well can be retained and do not necessarily have to be retested (although replicating their effects and seeing how they interact with any new components are probably good ideas); any components that performed poorly can be revised. Even if all of the components previously tested turned out to be poor performers, this at least has demonstrated that none of those components work.

Mediation analyses may be helpful in planning a strategy for building a more potent intervention in the next round of MOST (see the Smith, Coffman, & Zhu (2018) chapter in the companion volume). Depending on the experimental design used in the optimization phase, it may be possible to fit models that enable

examination of mediation of individual components. If a particular component fails to have an effect on a mediator but the mediator has an effect on the outcome, it may be helpful to try to revise the component or try a different strategy to affect the mediator. If the component affects the mediator but the mediator does not affect the outcome, then it may be necessary to rethink the conceptual model.

1.8.2 Returning to the Preparation Phase After the Evaluation Phase

In Fig. 1.1 there is an arrow leading from the evaluation phase back to the preparation phase, indicating that at the conclusion of the evaluation phase, there is always a return to the preparation phase. This return will occur after one of two possible outcomes: the optimized intervention either has or has not been determined to have a statistically and clinically significant effect.

If, based on the results of the optimization phase of MOST, a careful decision was made about whether or not to conduct the RCT, a null result is less likely than it would have been if the classical approach was taken. This is because in MOST, as discussed above, the decision about whether or not to go ahead with the RCT is informed by empirical evidence about the effectiveness of the components, which provides a sense of the expected potency of the intervention. This evidence is obtained not from pilot studies with low statistical power, but from one or more carefully controlled and adequately powered optimization trials. By contrast, in the classical approach, a set of components is typically assembled into an intervention a priori, informed by little, if any, empirical evidence about the likely effectiveness of the intervention except what has been obtained from pilot studies.

However, a null result on the RCT is always a possibility. This may occur because the statistical conclusion drawn from the RCT is the result of a Type II error; in other words, the intervention is effective in the population, but by chance a sample has been drawn that does not reflect this. Type II errors are always possible, particularly when the RCT is underpowered. Another possibility is that one or more of the conclusions upon which the optimization was based were the result of Type I errors; in other words, the investigator concluded that one or more components were effective when in reality they were ineffective. In this case the effect size of the optimized intervention in reality may have been smaller than would have been expected based on the results observed during the optimization phase. There are many other possibilities for why an intervention might fail to show a significant effect in an RCT. These include the usual issues that may affect RCTs, such as poor implementation of the intervention or compensatory behavior in the control subjects. A careful analysis of such issues should be completed as part of the preparation phase for the subsequent cycle of MOST.

If the results of the RCT indicate that the optimized intervention has a statistically and clinically significant effect, the return to the preparation phase is consistent with

the continual optimization principle, discussed above. This principle states that once an intervention has been optimized and evaluated—in other words, a cycle of MOST has been completed—subsequent work can begin to improve the intervention further. Interventions could even be assigned consecutive version numbers, and release notes could be provided, in much the way successive versions of software are released!

1.9 The Distinction Between Optimization and Evaluation

The optimization of an intervention and the evaluation of the resulting optimized intervention are related but distinct concepts. In this section some of the differences between optimization and evaluation are reviewed.

Optimization and evaluation have different objectives. The objective of optimization is to arrive at an intervention that meets the optimization criterion that was identified in the preparation phase. By contrast, the objective of evaluation is to determine whether an intervention has a statistically and clinically significant effect. An intervention may have demonstrated a statistically significant effect in an RCT and not have been optimized; in fact, at this writing this is true of most evidence-based interventions. Conversely, as discussed above, an intervention may have been optimized, and yet its anticipated effect may be small. A large sample size may be required to achieve adequate power for detection of such an effect in an RCT; moreover, a small effect may not be clinically meaningful.

Optimization and evaluation require different approaches to research. Optimization requires going "under the hood" to assess the performance of *individual components* of the intervention. This information enables the selection of the components and component levels that best meet the optimization criterion. In general, the standard two-arm RCT is not the most efficient experimental design for addressing the research questions that are posed in the optimization phase, although it is the most appropriate experimental design when it comes time to assess the performance of the intervention *as a package*. Much of the remainder of this book and the companion volume are devoted to research methods for the optimization of interventions.

1.10 A Different Way of Thinking About Intervention Research: MOST-Induced Dilemmas

In some ways, MOST calls for a different way of thinking about intervention research, on the part of investigators, funders, and the field in general. When an investigator is thinking about intervention research from a MOST perspective and the investigator's mentors, employers, and funders are not, this can lead to MOST-induced dilemmas.

For example, consider Dr. B, who is building the intervention to improve adherence to ART that has been discussed in this chapter. Suppose Dr. B has 5 years of funding, with the usual expectation of development of an intervention followed by its evaluation in an RCT. It has been made clear to Dr. B by several mentors that the greatest professional rewards come from an RCT that shows a statistically and clinically meaningful effect.

Recall that based on extended discussions with insurers, Dr. B has established that they are unwilling to pay more than $500 per person for implementation. Therefore, Dr. B has selected an optimization criterion that states the optimized intervention will be the one that produces the best expected outcome for $500 or less. Suppose an optimization trial reveals several effective components, but not many that are both effective and inexpensive. Thus because of the $500 limit, the optimized intervention is expected to have an effect that is so small it would be unlikely to be deemed statistically and clinically significant in an RCT. The results of the optimization trial suggest that if, instead of using the optimization criterion demanded by the insurers, cost was disregarded and all of the components were included at the higher level, the resulting intervention would be likely to demonstrate a statistically and clinically significant effect in an RCT. The insurers are not interested in this intervention; they are adamant that it is too expensive to be scalable, and the $500 limit is firm.

Thus, Dr. B is faced with a dilemma. The resource management principle of MOST, as depicted in Fig. 1.1, would suggest that available resources should not be spent on an RCT. Instead, they would be better spent on a return to the preparation phase and a search for additional candidate components that will contribute to intervention effectiveness without adding too much cost. But this means that in the current 5-year funding period at least, there will be no RCT. This is professionally disappointing to Dr. B and may be disappointing to the organization funding the research as well. The other course of action is to ignore issues of scalability and build and evaluate the more costly intervention. This is consistent with the classical approach and is a promising path to a successful RCT and all the professional rewards that go along with it. However, because Dr. B is familiar with MOST, there is the nagging concern that in the long run, it is best to avoid devoting resources to evaluation of an intervention unlikely ever to be widely implemented and therefore unlikely to make much of an impact on public health.

From one perspective, the root of this dilemma is a disconnect between advancing public health on the one hand and the reward structure for academic intervention scientists on the other. If publication of an RCT that reveals an intervention to have a statistically and clinically significant effect is a major career maker, to the exclusion of other kinds of publications, then it is not surprising Dr. B would seriously consider conducting an RCT on an intervention known not to be scalable. After all, at this time there is no expectation that a report of an RCT will make any argument that the intervention evaluated is efficient, economical, or scalable—only that it is effective.

By publishing the results of the optimization trial, Dr. B will increase the knowledge base about which strategies work and which do not. This is an important contribution and one that arguably should not be disappointing to funders or anyone else. Nevertheless, today such a publication may not be as highly valued, for example, by a promotion and tenure committee, as a report of an RCT.

Now suppose Dr. B wants to use the remaining funds in the grant to return to the preparation and optimization phases, in particular, to conduct another optimization trial. This leads to another dilemma: how to approach the funding agency to ask permission to replace the RCT with another optimization trial. Intervention science based on testing strategies to see whether they are effective, and then moving ahead if they are or going back to the drawing board if they are not, is intuitively appealing and consistent with how most of us have been taught to conduct scientific inquiry. Yet this approach requires a level of flexibility in resource management that at this writing is foreign to many funding agencies. Most grant proposals are expected to provide a plan for 5 years of research, with no provision for making a midcourse decision about how best to use resources based on results obtained.

It could be argued that in some areas, intervention science would be moved forward faster if investigators were encouraged to submit a research strategy to funders that contained a clearly specified if-then branching. One example of such branching might be

> *If at the conclusion of the optimization phase the empirical evidence suggests that the optimized intervention package is expected to have an effect size of at least $d = 0.3$, then its performance will be compared to a suitable control group in an RCT. Otherwise, we will return to the preparation phase, work on improving or replacing any poorly performing components, and build a new optimized intervention.*

If necessary, separate budgets could be provided for each branch. Such an approach to funding research could ultimately lead to better interventions.

1.11 What's Next

This chapter has provided a conceptual overview of MOST, without going into much detail. The remaining chapters in this book are intended to complete the picture and provide the reader with the necessary background to implement MOST in his or her own research. In the next chapter, Chap. 2, the first phase of MOST, preparation, is described. In the preparation phase, considerable attention is paid to the articulation of a conceptual model of the behavioral, biobehavioral, or biomedical process of interest and how the intervention under development is to intervene on this process. Chapter 2 also discusses the optimization criterion and its role in MOST.

References

Almirall, D., Nahum-Shani, I., Wang, L., & Kasari, C. (2018). Experimental designs for research on adaptive interventions: Singly and sequentially randomized trials. In L. M. Collins & K. C. Kugler (Eds.), *Optimization of behavioral, biobehavioral, and biomedical interventions: Advanced topics* (forthcoming). New York, NY: Springer.

Caldwell, L. L., Smith, E. A., Collins, L. M., Graham, J. W., Lai, M., Wegner, L., ... Jacobs, J. (2012). Translational research in South Africa: Evaluating implementation quality using a factorial design. *Child and Youth Care Forum*, *41*, 119–136.

Collins, L., Trail, J., Kugler, K., Baker, T., Piper, M., & Mermelstein, R. (2014). Evaluating individual intervention components: Making decisions based on the results of a factorial screening experiment. *Translational Behavioral Medicine*, *4*, 238–251.

Collins, L. M., & Kugler, K. C. (Eds.) (2018). *Optimization of multicomponent behavioral, biobehavioral, and biomedical interventions: Advanced topics*. New York, NY: Springer.

Collins, L. M., Kugler, K. C., & Gwadz, M. V. (2016). Optimization of multicomponent behavioral and biobehavioral interventions for the prevention and treatment of HIV/AIDS. *AIDS and Behavior*, *20*, 197–214.

Collins, L. M., Nahum-Shani, I., & Almirall, D. (2014). Optimization of behavioral dynamic treatment regimens based on the sequential, multiple assignment, randomized trial (SMART). *Clinical Trials*, *11*, 426–434.

Flay, B. R. (1986). Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. *Preventive Medicine*, *15*, 451–474.

Kugler, K. C., Wyrick, D. L., Tanner, A. E., Milroy, J. J, Chambers, B. D., Ma, A., ... Collins, L. M. (2018). Using the multiphase optimization strategy (MOST) to develop an optimized online STI preventive intervention aimed at college students: Description of conceptual model and protocol. In L. M. Collins & K. C. Kugler (Eds.), Optimization of multicomponent behavioral, biobehavioral, and biomedical interventions: Advanced topics (forthcoming). New York, NY: Springer.

Leon, A. C., Davis, L. L., & Kraemer, H. C. (2011). The role and interpretation of pilot studies in clinical research. *Journal of Psychiatric Research*, *45*, 626–629.

Michie, S., Richardson, M., Johnston, M., Abraham, C., Francis, J., Hardeman, W., ... Wood, C. E. (2013). The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: Building an international consensus for the reporting of behavior change interventions. *Annals of Behavioral Medicine*, *46*, 81–95.

Pignatiello, J. J., Jr., & Ramberg, J. S. (1985). Discussion of "off-line quality control, parameter design, and the Taguchi method" by Kackar, R.N. *Journal of Quality Technology*, *17*, 198–206.

Piper, M. E., Fiore, M. C., Smith, S. S., Fraser, D., Bolt, D. M., Collins, L. M., ... Baker, T. B. (2016). Identifying effective intervention components for smoking cessation: A factorial screening experiment. *Addiction*, *111*, 129–141.

Rivera, D. E., Hekler, E. B., Savage, J. S., & Downs, D. S. (2018). Intensively adaptive interventions using control systems engineering: Two illustrative examples. In L. M. Collins & K. C. Kugler (Eds.), *Optimization of multicomponent behavioral, biobehavioral, and biomedical interventions: Advanced topics* (forthcoming). New York, NY: Springer.

Schlam, T. R., Fiore, M. C., Smith, S. S., Fraser, S., Bolt, D. M., Collins, L. M., ... Baker, T. B. (2016). Comparative effectiveness of intervention components for producing long-term abstinence from smoking: A factorial screening experiment. *Addiction*, *111*, 142–155.

Smith, R. A., Coffman, D. L., & Zhu, X. (2018). Investigating an intervention's causal story: Mediation analysis using a factorial experiment and multiple mediators. In L. M. Collins & K. C. Kugler (Eds.), *Optimization of multicomponent behavioral, biobehavioral, and biomedical interventions: Advanced topics* (forthcoming). New York, NY: Springer.

West, S. G., Aiken, L. S., & Todd, M. (1993). Probing the effects of individual components in multiple component prevention programs. *American Journal of Community Psychology*, *21*, 571–605.

Wu, C. J., & Hamada, M. S. (2011). *Experiments: Planning, analysis, and optimization*. New York, NY: Wiley.