Chapter 14
Optimizing Family Intervention Programs: The Multiphase Optimization Strategy (MOST)

Linda M. Collins

Optimizing Family Intervention Programs: The Multiphase Optimization Strategy (MOST)

Behavioral interventions aimed at the family are an important vehicle for improving the well-being of the family as a unit and of its individual members. A few examples include the Family Bereavement Program (Hagan et al. 2012; Sandler et al. 2010), aimed at helping to improve adjustment in parentally bereaved families; the Family Check-Up (Shaw et al. 2006), an intervention for reducing child problem behavior by improving parenting; and Familias Unidas (Prado and Pantin 2011), aimed at preventing substance use and risky sexual behavior in Hispanic youth.

As different as these behavioral interventions are, they have some important features in common. First, each of them is multi-component. The Family Bereavement Program includes separate components targeted to children, adolescents, and caretakers. The Family Check-Up includes components focusing on parenting, family management, and contextual issues. Components of Familias Unidas include family visits, group sessions with parents, parent-teacher/counselor meetings, skill building activities, and supervised peer activities.

The second feature that these behavioral interventions have in common is that they are all evidence-based; that is, each has been empirically evaluated and demonstrated to have a statistically significant effect on key outcomes. The randomized clinical trial (RCT) is widely considered to be the gold standard for this evaluation. The purpose of an RCT is to compare experimentally the performance of a behavioral intervention to that of a suitable comparison intervention. The comparison intervention is frequently whatever constitutes the customary standard of care, or sometimes treatment is delayed in the control group and provided after the intervention is expected to have its effect in the treatment group.

L. M. Collins (✉)
The Methodology Center, Department of Human Development and Family Studies,
The Pennsylvania State University, University Park, PA, USA
e-mail: lmc8@psu.edu

Although evaluation of multicomponent behavioral interventions via the RCT is indispensable, much more can be done to develop highly effective, efficient, and cost-effective behavioral interventions (Collins et al. 2005, 2007, 2011). As an alternative to sole reliance on the RCT, Collins and colleagues suggest a different approach called the multiphase optimization strategy (MOST). MOST is a comprehensive framework that includes not only evaluation of behavioral interventions via the RCT, but also optimization of the intervention before evaluation. This is done by conducting a component screening experiment to gather empirical information on the performance of individual intervention components, and using this information to select the intervention components and component levels to form an intervention package that meets a prespecified optimization criterion. In other words, the objective is to build an optimized intervention, and then to bring this optimized intervention to an RCT. Examples of applications of the MOST framework include Caldwell et al. (2012), Collins et al. (2011), McClure et al. (2012), and Strecher et al. (2008).

The purpose of this chapter is to introduce MOST to the family research community. The chapter will review a hypothetical example of the use of MOST to develop a behavioral intervention, and then offer a discussion of considerations relevant to the application of MOST in family research.

A Hypothetical Example of MOST

Why Consider Using MOST?

Suppose a team of investigators is interested in developing a behavioral intervention for treatment of adolescent substance use. The investigators are considering four components for inclusion in the intervention. The first three components are aimed at the adolescent. They are (1) motivational enhancement and cognitive behavioral therapy (ME/CBT; Sampl and Kadden 2001); (2) academic skill building (ASB), in which adolescents are taught study and time management skills, and tutored in their weakest school subjects; and (3) contingency management (CM; Stanger et al. 2009), in which the adolescent undergoes a weekly urine toxicology screening and is given a monetary reward for each clean specimen. The final component is (4) multidimensional family therapy (MDFT; Liddle et al. 2008), in which a therapist works with the adolescent and the family in several different treatment domains, including parenting skills, family communication, and family competence.

The customary approach would be to form an intervention package by combining the four components, and then evaluate this intervention by means of a classic RCT. This approach would address an important question: Does the intervention, as a package, have a statistically significant effect on adolescent substance use compared to a suitable control or comparison group? However, the RCT would not directly address some other questions, which are also important. Suppose the RCT is conducted, and the results suggest that the intervention package has a statistically significant effect. The RCT would still not provide direct answers to the following questions:

Does each of the four components contribute to the overall effect, and therefore, is each essential to the intervention? It may be that one or more of the components is not effective, or does not provide added value over and above other components. In this case these components could be eliminated to save time and money, with little or no decrement in overall effectiveness.

Does each component produce an effect large enough to justify its cost? For example, CM can be very resource-intensive. In Stanger et al.'s (2009) CM procedure each adolescent could earn up to nearly $600. In addition, CM can be somewhat controversial politically, suggesting that there can be costs in terms of time and effort associated with convincing communities to support its use. If CM is more expensive than other components that are candidates for inclusion in the intervention, it is worthwhile to determine whether it also has a larger effect.

Now suppose the RCT suggests that the intervention package does not have a statistically significant effect. The RCT would not provide direct answers to the following question: *Are any of the individual components worth saving for inclusion in a future intervention?* If an RCT shows that a treatment package is ineffective, the investigator learns little about what went wrong. It would be helpful to know whether one or more of the components is worth saving. It is even possible that one or more components had an iatrogenic effect, offsetting the positive effects of other components and producing a net effect of zero.

It is not necessary to address any of these questions to obtain evidence about whether the treatment as a package is working as intended. However, it is necessary to address these questions to develop an efficient intervention, point the way toward the next steps that should be taken to improve a successful intervention, or amend an unsuccessful intervention and make it successful. These are reasons why our hypothetical investigators might decide to use the MOST framework to develop an adolescent substance use treatment intervention.

Optimization and the Optimization Criterion

The words "optimal" and "optimized" are sometimes used colloquially to mean "best." However, in MOST a more technical definition drawn from mathematics and engineering is used. According to the *Concise Oxford Dictionary of Mathematics*, the term "optimization" means "the process of finding the best possible solution . . . subject to given constraints" (Clapham and Nicholson 2009). Thus, according to this definition an optimized intervention is not the best possible intervention in some absolute sense. Rather, it is the best intervention that can be constructed subject to realistic constraints or limitations. In MOST, one limitation is always the finite set of intervention components under consideration. Other constraints may be the amount of money or the amount of time required to deliver the intervention.

MOST requires the investigator to choose an *optimization criterion*. The optimization criterion is an operational definition of an optimized intervention. In other words, it is the goal of MOST as applied to a particular intervention. In our example

we will use one simple optimization criterion: "an intervention that includes only active components operating in the desired direction." This criterion expresses the modest goal of arriving at an intervention that includes only components that have been demonstrated to contribute positively to any overall program effect, and is the optimization criterion that will be used in the hypothetical example described here. More sophisticated optimization criteria are possible. For example, suppose it is known that for pragmatic reasons the drug abuse treatment program must cost no more than $1,000 per family to implement. Then the desired optimization criterion might be "the most effective intervention (based on the four components at hand) that will not exceed $1,000 per family in implementation costs." It is also possible to express the optimization criterion in terms of an absolute effect size, for example, to come as close as possible to an overall effect size of $d = 0.5$.

The optimization criterion should articulate a goal for the intervention that is clear, is desirable, and includes realistic constraints. Suppose the investigators decided to select as an optimization criterion "the most effective intervention that will not exceed $5,000 per family in implementation costs." The resulting intervention would likely be more effective than an intervention developed using the constraint of $1,000 per family, but this means little if the $5,000 intervention is unlikely to be implemented broadly. It may be better to select an optimization criterion that uses a smaller upper limit on implementation costs and arrive at a somewhat less effective but readily implementable intervention. Of course, the resulting intervention must still demonstrate a statistically significant effect when it is evaluated via an RCT.

The Component Screening Experiment

The purpose of the component screening experiment is to gather information that will be used to determine which intervention components are eligible for inclusion in the intervention. (Please note that the term "screening" is used here the same way it is used in engineering: it refers to screening of components rather than screening of intervention participants based on characteristics such as psychiatric diagnosis.) Selection of an experimental design is critical. Taking a resource management perspective when selecting an experimental design is recommended (Collins et al. 2009). From this perspective, the best experimental design is the one that gathers the most, and most relevant, scientific information while making the most efficient use of available resources. The term "resources" is broadly defined here. Usually money and experimental subjects are the most important resources, but the list may include personnel, equipment, time, or any other resource that must be drawn upon in significant quantities to conduct the experiment.

In the example, there are four intervention components to be examined. One alternative would be to conduct four individual experiments, one for each of the intervention components. That is, one experiment would compare ME/CBT to a control, a second experiment would compare ASB to a control, and so on. In a sense,

Condition number	Intervention component			
	ME/CBT	ASB	CM	MDFT
1	On	On	On	On
2	On	On	On	Off
3	On	On	Off	On
4	On	On	Off	Off
5	On	Off	On	On
6	On	Off	On	Off
7	On	Off	Off	On
8	On	Off	Off	Off
9	Off	On	On	On
10	Off	On	On	Off
11	Off	On	Off	On
12	Off	On	Off	Off
13	Off	Off	On	On
14	Off	Off	On	Off
15	Off	Off	Off	On
16	Off	Off	Off	Off

Table 14.1 Experimental conditions in factorial component screening experiment

each of these experiments is a mini-RCT, in which a single treatment is compared to a control.

Another alternative is to conduct a factorial experiment (e.g., Fisher 1926; Kirk 2012). In a factorial experiment, multiple intervention components can be examined simultaneously, in a design that varies the levels of the factors in a systematic way. In the example, each of the intervention components can have two levels: on (included in the intervention) or off (not included in the intervention). Table 14.1 shows the experimental conditions that would be included in the factorial experiment. As Table 14.1 shows, in this factorial experiment every level of each intervention component is systematically combined with every level of every other intervention component. This results in a total of $2^4 = 16$ experimental conditions.

The logical underpinnings of a factorial experiment are different from those of an RCT and related experimental designs. In an RCT, individual experimental conditions are directly compared to assess the differences between them. By contrast, in a factorial experiment individual experimental conditions are usually not compared, and therefore statistical power is not based on comparisons of individual experimental conditions. Instead, combinations of experimental conditions are compared to estimate the main effect of each independent variable (e.g., intervention component), and the interactions between independent variables. For example, the main effect of ME/CBT would be obtained by comparing the mean of all of the experimental conditions in which ME/CBT is set to on (conditions 1–8) against the mean of all of the experimental conditions in which ME/CBT is set to off (conditions 9–16). Similarly, to obtain the main effect of ASB, the mean of the conditions in which ASB is set to on (conditions 1–4 and 9–12) would be compared to the mean of the conditions in which ASB is set to off (conditions 5–8 and 13–16). The main effects of the other two intervention components are obtained in the same manner. Note that

all of the subjects are included in all comparisons by "reshuffling" the subjects for each comparison. The same subject might be in the "off" group for ME/CBT, the "on" group for ASB, and so on.

The reader may wonder how it can be appropriate to conduct this comparison when subjects in, say, the ME/CBT "on" group are in different combinations of levels of the other intervention components. The answer is that complete factorial experiments (as opposed to fractional factorial experiments, which are described later in this chapter) are designed to eliminate confounding of effects. Examination of Table 14.1 shows that when ME/CBT is set to on, each of the other intervention components is on in half of the experimental conditions and off in half of the experimental conditions, and this pattern is repeated when ME/CBT is set to off. Thus, although the other intervention components may have an effect on the outcome, if they raise the mean when ME/CBT is set to on, they will raise the mean an equivalent amount when ME/CBT is set to off, and there will be no net effect on the estimate of the main effect. The same holds true for estimates of interactions between components.

Selecting an Experimental Design from a Resource Management Perspective

Now that the logic of the factorial experiment has been reviewed, we can compare the scientific yield and resource demands of four individual experiments vs. one factorial experiment. Both approaches yield estimates of the individual effect of each component, although the estimates are not equivalent (see Collins et al. 2009). Only the factorial experiment yields estimates of interactions between components.

To compare resource demands it is necessary to consider the sample size requirements to achieve the desired level of statistical power. The investigators who are to conduct the screening experiment must identify the minimum effect size that a component must demonstrate to render it eligible for inclusion in the intervention package. This effect size then serves as the starting point for a power analysis, because effects smaller than this predetermined minimum are not of interest for the purpose of intervention optimization. Suppose the investigators have decided that an intervention component must demonstrate a main effect size of at least $d = 0.3$ to be considered for inclusion in the intervention package. A power analysis indicates that achieving power of at least 0.80 requires a sample size of approximately 340 for each experiment, or a total N of 1,360 for the individual experiments approach. By contrast, the power analysis indicates that in the factorial experiment, a total N of approximately 340 is required to detect the same effect size at the same level of statistical power. In other words, the individual experiments approach requires four times the sample size of the factorial experiment. The overall N in the factorial experiment will be divided approximately equally among the 16 experimental conditions, for a per-condition sample size of about 21. Although the factorial experiment has relatively few subjects per condition, this does not matter because, as mentioned above, individual conditions are not compared. Thus, the power of a factorial experiment is driven by the overall N rather than the number of subjects in individual conditions.

Although a factorial experiment always requires many fewer subjects than would be required to conduct individual experiments on the same number of intervention components, the trade-off is that it will nearly always require more experimental conditions. As Table 14.1 shows, the factorial experiment will require 16 experimental conditions, as opposed to a total of eight in the individual experiments approach. The additional experimental conditions required by the factorial experiment may result in additional overhead costs, such as additional staff to deal with the more complicated logistics.

In choosing between these two alternative design approaches and others, the investigators must weigh the benefits against the costs. The factorial experiment yields both main effects and interactions, and in this case requires one quarter as many experimental subjects. However, the factorial experiment also requires implementation of twice as many experimental conditions. The decision hinges on what resources are available, and which approach makes the best use of these resources. A SAS macro that investigators can use to weigh the relative resource demands of different designs for component screening experiments may be found at http://methodology.psu.edu/downloads.

Decision Making Based on Component Screening

The data from the screening experiment (if the factorial design is chosen) or experiments (if the individual experiments approach is chosen) are analyzed to produce estimates of the individual component effects and, in the case of a factorial experiment, interactions between intervention components. Based on the results, the investigators can determine which of the components attain the prespecified threshold effect size of $d = 0.3$. If interaction effects are available, these can be used to determine whether the effect of a component is significantly enhanced or dampened by the presence of one or more other components. This is all the information that is needed to achieve the simple optimization criterion of "an intervention that includes only active components." Any components that achieve or exceed the threshold effect size will be included in the intervention package. If a more complex optimization criterion has been selected, it may be necessary to include additional data, such as implementation costs, in decision making.

If the intervention package is satisfactory, the next step is to conduct a standard RCT, with individual families randomly assigned either to receive the intervention package or to a control group, comparison group, or wait list. However, it is possible that the results of the screening experiment will indicate that it is not worthwhile to devote resources to an RCT. This conclusion might be drawn if too few or even no components reach the threshold effect size. In this case the resources that would have been devoted to an RCT could instead be devoted to another screening experiment to examine a new set of candidate intervention components. Any components that were demonstrated to be successful in the previous screening experiment could either be included in the new experiment to replicate their effects and determine whether they

interact with any of the new components, or set aside for inclusion in the treatment package when it is ultimately assembled.

Some Considerations Relevant to the Application of MOST in Family Research

What Constitutes an Intervention Component?

The concept of an intervention component is a broad one. In most cases, the investigator determines what constitutes a component, selecting aspects of the intervention that (1) can be separated out for examination and (2) are important enough to devote resources to examining. In the hypothetical example reviewed here, the intervention components were distinct strategies for adolescent drug abuse treatment. It would be possible to apply the MOST framework to examine a different set of intervention components. Four treatment domains within multidimensional family therapy (MDFT): adolescent, parent, interactional, and extrafamilial, are discussed by Liddle et al. (2008). If the treatments associated with each of these domains can be delivered independently, that is none requires the inclusion of any of the others, then MOST could be used to develop an optimized MDFT intervention. Sometimes there is a core component that provides a foundation upon which all of the others build, such as a component providing basic information about the effects of drugs in a drug abuse prevention or treatment program. An experimental design can be used in which this core component is provided to all subjects, as long as the remaining components are independent of each other. With this approach the effectiveness of the core component alone cannot be estimated.

The components examined in MOST do not necessarily have to be aspects of intervention content. A study in which the intervention itself, a school-based drug abuse and HIV prevention program focusing on positive use of leisure time, was not examined is described by Caldwell et al. (2012). Instead, the investigators examined three factors hypothesized to affect the fidelity with which teachers delivered the intervention: enhanced teacher training, teacher support, and enhanced school environment. It would also be possible to use MOST to examine components that promote adherence to the intervention, or that support sustainability of implementation.

Highly Efficient Fractional Factorial Experimental Designs

As was discussed in an earlier section, factorial experiments often require many fewer experimental subjects than alternative approaches. However, they usually require implementation of more experimental conditions. As the number of intervention components to be examined increases, the number of experimental conditions required rises rapidly (but not the number of subjects required; see Collins et al. 2009).

If another two-level intervention component were to be added to the hypothetical factorial experiment depicted in Table 14.1, the number of experimental conditions would be doubled, to $2^5 = 32$; examination of six components would require $2^6 = 64$ experimental conditions. Although the logistics of large factorial experiments in field settings should not be undertaken lightly, there are examples of such experiments being carried out successfully, even in challenging circumstances. One example provided by Caldwell et al. (2012) described an eight-condition factorial experiment in middle schools in a poor school district in South Africa. Another by Collins et al. (2011) described a 32-condition factorial experiment conducted in primary health care settings in the United States midwest. In some situations conducting a large factorial experiment may not be much more resource-intensive than conducting an RCT, for example, when the intervention is largely Internet-delivered.

When it is scientifically important to examine multiple intervention components, there is a design alternative that enables the investigator to take advantage of the factorial experiment's economical use of subjects while reducing the number of experimental conditions that must be implemented. Fractional factorial designs (Wu and Hamada 2009; Collins et al. 2009; Dziak et al. 2012) are a special type of factorial design with a long history in engineering and related fields, but to date they have rarely been used in the behavioral sciences. In fractional factorial designs, which are associated with exactly the same sample size requirements as complete factorial designs, a carefully selected fraction (hence the name) of experimental conditions is included. A fractional factorial design was used by Strecher et al. (2008) to examine the performance of five components of an online smoking cessation intervention. The design reduced the number of experimental conditions by half, from 32 to 16. A fractional factorial design was used by Collins et al. (2011) to examine six components of a clinic-based smoking cessation intervention. The design cut the number of experimental conditions by half, from 64 to 32. Depending on the number of factors in the experiment, there may be fractional factorial designs available that cut the number of experimental conditions by 75 % or more.

Fractional factorial designs are economical, but there is a tradeoff for this economy. Whenever experimental conditions are removed from a complete factorial design, the number of effects that can be estimated is reduced. As a result, effects that could be individually estimated in a complete factorial design become combined in a fractional factorial design. These effects can no longer be disentangled; they are estimated as a "bundle" (Collins et al. 2009; Chakraborty et al. 2009; Wu and Hamada 2009). In other words, they are confounded (or, to use the term that is more common in engineering, aliased). For every fractional factorial design it is known in advance which effects will be bundled together; in fact, the investigator can select a design strategically so as to bundle the effects of primary scientific interest with effects that are expected to be very small. Typically this means selecting a design in which main effects and two-way interactions are bundled with higher-order interactions. For example, in the fractional factorial design used by Collins et al. (2011), each main effect estimate is actually an estimate of that main effect plus one five-way or six-way interaction, and each estimate of a two-way interaction is actually an estimate of that interaction plus one four-way interaction. Because Collins and colleagues

had no reason to expect that the three-way and higher-order interactions would be large, the estimates of main effects and two-way interactions were expected to be approximately equivalent to those that would have been obtained in a much more expensive complete factorial experiment.

Fractional factorial designs with different numbers of factors, different numbers of conditions, and different patterns of bundling of effects have been tabulated and are available in books and on the Internet. One convenient method of selecting a fractional factorial design is to use a computer program such as PROC FACTEX in SAS. A brief tutorial on selecting a fractional factorial design, aimed at behavioral scientists, appears in Collins et al. (2009).

When Subjects are Sampled in Clusters

Sampling subjects in clusters is a reality of much behavioral science. Because the family is a cluster of people, cluster sampling is nearly always used to some extent in family research. In much family research, the clustering of individuals into families is itself of research interest. Thus, each family member as well as relationships within the family may play a unique role in the research, and the extent to which family members are alike or unalike is scientifically interesting.

However, some types of clustering are not of scientific interest and therefore are primarily a nuisance, as when families must be sampled from clusters such as schools, hospitals, health care clinics, and neighborhoods. In the hypothetical example reviewed in this chapter, families might have been sampled from several drug abuse treatment centers. Clustering is important because units within a cluster tend to be more alike than two randomly selected units. That is, families who live within the same community tend to be more alike than two randomly selected families will be. The degree of this similarity is expressed in the intra-class correlation.

When clusters are present, randomization to experimental conditions can take place in two ways (Dziak et al. 2012). In within-cluster randomization, the families within a cluster can be randomized to different experimental conditions. This approach is taken when there is little interaction between families within a cluster, and therefore there is little risk of contamination between experimental conditions. The experiment described in Collins et al. (2011) sampled subjects from health care clinics, and used within-cluster randomization because interaction among clinic patients was minimal. In between-cluster randomization, entire units are randomized to experimental conditions. This approach, taken when there is a risk of contamination, is used frequently in educational research and other research, conducted in educational settings, such as drug abuse prevention (Murray 1998; Raudenbush and Bryk 2002). In one study, schools were assigned to experimental conditions for two reasons (Caldwell et al. 2012). One reason was that two of the components were aimed at teachers, and it was believed that teachers within a school would be likely to share information. The other reason was that one of the components, school environment, was aimed at the school as a unit.

The implications of cluster structure and the choice of within-cluster or between-cluster randomization on factorial experiments are discussed by Dziak et al. (2012). One area in which cluster structure can have an effect is statistical power. When between-cluster randomization is used, the presence of a non-zero intra-class correlation results in what is known as the design effect (Murray 1998). Essentially, this means that as compared to a situation in which individual subjects are randomized, more subjects are required to maintain the same level of statistical power. All else being equal, a larger intra-class correlation and larger clusters (i.e., the same N divided into fewer clusters) produce a larger design effect. It has been shown that the design effect has the same effect on factorial experiments as it does in the standard RCT, and that standard approaches to estimating power are accurate when between-cluster randomization is to be used in a factorial experiment (Dziak et al. 2012). They also showed that the design effect is an issue primarily with between-cluster randomization. When within-cluster randomization is structured so that each unit within a cluster has the same experimental condition assignment probabilities, the design effect is typically negligible.

In addition to its effect on statistical power, cluster structure can impact choice of experimental design. Suppose a scientist who works in the area of school-based drug abuse prevention wishes to conduct a factorial experiment to examine the performance of five intervention components. This experiment will involve $2^5 = 32$ experimental conditions. The investigator has access to enough students to achieve an acceptable level of power, even after taking the design effect into account. However, the students are clustered into 30 schools, and it is determined that between-cluster randomization is necessary. Thus, the investigator has a sufficiently large sample of individual subjects, but does not have enough clusters to implement 32 experimental conditions.

The investigator must somehow reduce the number of experimental conditions. One approach would be to abandon the idea of conducting a factorial experiment and choose a different approach, such as individual experiments. This idea is unappealing for two reasons. First, it would rule out the possibility of examining interaction effects. Second, with a different approach the investigator would have to recruit many more subjects, and therefore many more schools. Another approach would be to reduce the number of intervention components being examined, which would reduce the scientific yield of the study and would probably require the same number of subjects as the larger experiment. Simply reducing the number of intervention components does not necessarily reduce the sample size requirements of a factorial experiment, and, conversely, increasing the number of components does not necessarily increase sample size requirements (See Collins et al. 2009). A third approach would be to retain the same number of intervention components for examination and select a fractional factorial design. This would reduce the number of experimental conditions by at least half and thereby make the experiment feasible with 30 schools. As mentioned above, bundling of effects must be considered carefully when evaluating a fractional factorial design. The use of fractional factorial designs with between-cluster randomization is discussed by Dziak et al. (2012).

Interactions

Two variables are said to interact when the effect of one is different depending on the level of the other. An investigator who raises the issue of interactions usually has one of two different kinds of interactions in mind. The first kind is an interaction between an intervention or intervention component and family characteristics, such as whether or not the family is headed by a single parent, below the poverty line, or of a particular ethnicity. These interactions are of interest because they suggest that an adaptive intervention (Collins et al. 2004; Lei et al. 2012) could be developed that would be tailored to improve performance in certain subgroups. For example, suppose there was evidence of an interaction between MDFT and family ethnicity, suggesting that this intervention component was successful with all families except Hispanics. The intervention package would have a larger overall effect if Hispanic families were provided with a component that better met their needs. The new component could be a specialized version of MDFT or an entirely different component.

The second kind of interaction involves two or more intervention components. In MOST, decisions about which components to retain for potential inclusion in the intervention and which to reject depend largely on main effects. Interactions are examined primarily to determine whether they indicate that these decisions should be reconsidered (Collins et al., in preparation). For example, if the main effect of Academic Skill Building (ASB) disappears when MDFT is included, this suggests that only one of the two components should be included in the intervention.

It may be helpful to clarify an issue about interpretation of main effects when interactions are present. The interpretation of any effects in ANOVA depends on what approach is used for coding the effects. In the behavioral sciences typically either dummy (0,1) or effect ($-1,1$) coding is used. Dummy coding produces effect estimates that technically are not main effects and interactions (at least according to the widely accepted textbook definitions, such as those found in Montgomery, 2009; for a detailed explanation, see Kugler et al. 2012). Moreover, these effects are often substantially correlated. Effect coding, which is recommended for component screening experiments (Collins et al. 2009), produces estimates of main effects and interactions that do correspond to the textbook definitions. Moreover, these estimates are orthogonal when the sample size is equal across experimental conditions (and nearly orthogonal otherwise in most cases), making it much more reasonable to consider each effect on its own merits.

Open Areas and Future Directions

One open area is how to make decisions about which intervention components to include in an intervention when some components may have different effects on different outcome variables. In family research different outcome variables may be relevant for different family members. In the drug abuse treatment example,

substance use might be the primary outcome for the adolescent, and parental effectiveness might be the primary outcome for the parent. In addition, there might be outcomes that pertain to dyadic relationships within the family, such as effectiveness of parental communication, or outcomes that pertain to the family as a unit, such as amount of family conflict. Decision making can be complicated under these conditions, particularly when experimental results are inconsistent across outcomes, and guidelines are needed. Another open area is how to use the MOST framework to develop the most cost-effective intervention. Additional research is also needed on experimental design, to increase the options available for intervention scientists.

It is interesting to consider what the future direction of family-based interventions might look like if the MOST framework were widely adopted. First, the pace with which the field accumulates a coherent base of scientific knowledge about which intervention approaches work, and which do not, might be accelerated, because every study would shed light on which components were effective. Second, scientists might begin engineering family interventions to meet specific and clearly stated optimization criteria, such as the most effective intervention that can be delivered under some particular dollar limit. Third, once this practice becomes common, the bar could be raised systematically and incrementally over time, so that one objective of new interventions would be to improve on the old by demonstrating that they were more effective, more efficient, or less costly. In this way, behavioral interventions could steadily make ever-increasing progress in improving the lives of families.

References

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.

Chakraborty, B., Collins, L. M., Strecher, V. J., & Murphy, S. A. (2009). Developing multicomponent interventions using fractional factorial designs. *Statistics in Medicine, 28,* 2687–2708.

Clapham, C., & Nicholson, J. (2009). Optimization. *The concise Oxford dictionary of mathematics.* www.oxfordreference.com/views/ENTRY.html?subview=Main&entry=t82.e2047. Accessed 3 April 2011.

Collins, L. M., Murphy, S. A., & Bierman, K. (2004). A conceptual framework for adaptive preventive interventions. *Prevention Science, 3,* 185–196.

Collins, L. M., Murphy, S. A., Nair, V., & Strecher, V. (2005). A strategy for optimizing and evaluating behavioral interventions. *Annals of Behavioral Medicine, 30,* 65–73.

Collins, L. M., Murphy, S. A., & Strecher, V. (2007). The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): New methods for more potent e-health interventions. *American Journal of Preventive Medicine, 32,* S112–S118.

Collins, L. M., Dziak, J. R., & Li, R. (2009). Design of experiments with multiple independent variables: A resource management perspective on complete and reduced factorial designs. *Psychological Methods, 14,* 202–224.

Collins, L. M., Baker, T. B., Mermelstein, R. J., Piper, M. E., Jorenby, D. E., Smith, S. S., & Fiore, M. C. (2011). The multiphase optimization strategy for engineering effective tobacco use interventions. *Annals of Behavioral Medicine, 41,* 208–226.

Dziak, J. D., Nahum-Shani, I., & Collins, L. M. (2012). Multilevel factorial experiments for developing behavioral interventions: Power, sample size, and resource considerations. *Psychological Methods, 17,* 153–175.

Fisher, R. A. (1926). The arrangement of field experiments. *Journal of the Ministry of Agriculture of Great Britain, 33,* 503–513.

Hagan, M. J., Tein, J. Y., Sandler, I. N., Wolchik, S. A., Ayers, T. S., & Luecken, L. J. (2012). Strengthening effective parenting practices over the long term: Effects of a preventive intervention for parentally bereaved families. *Journal of Clinical Child and Adolescent Psychology, 41,* 177–188.

Kirk, R. E. (2012). *Experimental design: Procedures for the behavioral sciences* (4th ed.). Los Angeles: Sage.

Kugler, K. C., Trail, J. B., Dziak, J. J., & Collins, L. M. (2012). *Effect coding versus dummy coding in analysis of data from factorial experiments (Technical report No. 12–120).* University Park: The Methodology Center, Penn State.

Lei, H., Nahum-Shani, I., Lynch, K., Oslin, D., & Murphy, S. A. (2012). A "SMART" design for building individualized treatment sequences. *Annual Review of Clinical Psychology, 8,* 21–48.

Liddle, H. A., Dakof, G. A., Turner, R. M., Henderson, C. E., & Greenbaum, P. E. (2008). Treating adolescent drug abuse: A randomized trial comparing multidimensional family therapy and cognitive behavior therapy. *Addiction, 103,* 1660–1670.

McClure, J. B., Derry, H., Riggs, K. R., Westbrook, E. W., St John, J., Shortreed, S. M., An, L. (2012). Questions about quitting (Q2): Design and methods of a multiphase optimization strategy (MOST) randomized screening experiment for an online, motivational smoking cessation intervention. *Contemporary Clinical Trials, 33,* 1094–1102.

Montgomery, D. C. (2009). *Design and analysis of experiments* (7th ed.). Hoboken: Wiley.

Murray, D. M. (1998). *Design and analysis of group-randomized trials.* New York: Oxford University Press.

Prado, G., & Pantin, H. (2011). Reducing substance use and HIV health disparities among Hispanic youth in the U.S.A.: The Familias Unidas program of research. *Psychosocial Interventions, 20,* 63–73.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks: Sage.

Sampl, S., & Kadden, R., (2001). *Motivational enhancement therapy and cognitive behavioral therapy for adolescent cannabis users: 5 sessions* (Vol. 1). Rockville: Center for Substance Abuse Treatment, Substance Abuse and Mental Health Services Administration.

Sandler, I. N., Ma, Y., Tein, J. Y., Ayers, T. S., Wolchik, S., Kennedy, C., & Millsap, R. (2010). Long-term effects of the family bereavement program on multiple indicators of grief in parentally bereaved children and adolescents. *Journal of Consulting and Clinical Psychology, 78,* 131–143.

Shaw, D. S., Dishion, T. J., Supplee, L., Gardner, F., & Arnds, K. (2006). Randomized trial of a family-centered approach to the prevention of early conduct problems: 2-year effects of the family check-up in early childhood. *Journal of Consulting and Clinical Psychology, 74,* 1–9.

Stanger, C., Budney, A. J., Kamon, J. L., & Thostensen, J. (2009). A randomized trial of contingency management for adolescent marijuana abuse and dependence. *Drug and Alcohol Dependence, 105,* 240–247.

Strecher, V. J., McClure, J. B., Alexander, G. W., Chakraborty, B., Nair, V. N., Konkel, J. M., Pomerleau, O. F. (2008). Web-based smoking cessation programs: Results of a randomized trial. *American Journal of Preventive Medicine, 34,* 373–381.

Wu, C. F. J., & Hamada, M. S. (2009). *Experiments: Planning, analysis, and optimization.* New York: Wiley.