

Intervention Optimization: A Paradigm Shift and Its Potential Implications for Clinical Psychology

Linda M. Collins,^{1,2} Inbal Nahum-Shani,³
Kate Guastaferro,¹ Jillian C. Strayhorn,¹
David J. Vanness,⁴ and Susan A. Murphy⁵

¹Department of Social and Behavioral Sciences, New York University, New York, NY, USA;
email: lmc9580@nyu.edu

²Department of Biostatistics, New York University, New York, NY, USA

³Institute for Social Research, University of Michigan, Ann Arbor, Michigan, USA

⁴Department of Health Policy and Administration, The Pennsylvania State University,
University Park, Pennsylvania, USA

⁵Departments of Statistics and Computer Science, Harvard University, Cambridge,
Massachusetts, USA

ANNUAL REVIEWS CONNECT

www.annualreviews.org

- Download figures
 - Navigate cited references
 - Keyword search
 - Explore related articles
 - Share via email or social media
-

Annu. Rev. Clin. Psychol. 2024. 20:21–47

First published as a Review in Advance on
February 5, 2024

The *Annual Review of Clinical Psychology* is online at
clinsy.annualreviews.org

<https://doi.org/10.1146/annurev-clinpsy-080822-051119>

Copyright © 2024 by the author(s). This work is
licensed under a Creative Commons Attribution 4.0
International License, which permits unrestricted
use, distribution, and reproduction in any medium,
provided the original author and source are credited.
See credit lines of images or other third-party
material in this article for license information.



Keywords

intervention optimization, multiphase optimization strategy, adaptive intervention, factorial experiment, sequential multiple assignment randomized trial, micro-randomized trial

Abstract

To build a coherent knowledge base about what psychological intervention strategies work, develop interventions that have positive societal impact, and maintain and increase this impact over time, it is necessary to replace the classical treatment package research paradigm. The multiphase optimization strategy (MOST) is an alternative paradigm that integrates ideas from behavioral science, engineering, implementation science, economics, and decision science. MOST enables optimization of interventions to strategically balance effectiveness, affordability, scalability, and efficiency. In this review we provide an overview of MOST, discuss several experimental designs that can be used in intervention optimization, consider how the investigator can use experimental results to select components for inclusion in the

optimized intervention, discuss the application of MOST in implementation science, and list future issues in this rapidly evolving field. We highlight the feasibility of adopting this new research paradigm as well as its potential to hasten the progress of psychological intervention science.

Contents

1. INTRODUCTION	23
2. OVERVIEW OF INTERVENTION OPTIMIZATION VIA THE MULTIPHASE OPTIMIZATION STRATEGY	24
2.1. Definition of Optimization	24
2.2. The Concept of Intervention Components	25
2.3. Brief Overview of the Multiphase Optimization Strategy	25
3. THE 2^k FACTORIAL OPTIMIZATION RANDOMIZED CONTROLLED TRIAL	27
3.1. Introduction to the Factorial Optimization Randomized Controlled Trial	27
3.2. A Few Words About Interactions.....	30
3.3. The Concept of Control in a Factorial Experiment.....	30
3.4. Conclusion-Priority Versus Decision-Priority Perspectives	30
4. DECISION MAKING AND INTEGRATION OF RESOURCE CONSIDERATIONS	31
5. OPTIMIZATION OF ADAPTIVE INTERVENTIONS	33
5.1. Fixed and Adaptive Intervention Strategies	33
5.2. Optimization of Adaptive Interventions with Relatively Slow Timescales of Adaptation: The Sequential Multiple Assignment Randomized Trial	34
5.3. Optimization of Adaptive Interventions with Relatively Fast Timescales of Adaptation: Just-in-Time Adaptive Interventions and Micro-Randomized Trials	36
6. OTHER EXPERIMENTAL DESIGNS USEFUL IN INTERVENTION OPTIMIZATION	37
6.1. Fractional Factorial Designs	37
6.2. Optimization Randomized Controlled Trials with Cluster Randomization ..	38
6.3. Evaluation-Optimization Trials	38
6.4. Multiple-Timescale Optimization Randomized Controlled Trial Designs for Optimization of Adaptive Interventions	39
6.5. System Identification	39
7. OPTIMIZATION AND IMPLEMENTATION SCIENCE: WRAPAROUND INTERVENTIONS	40
7.1. The Multiphase Optimization Strategy and Implementation Science	40
7.2. Implementation Strategies as Wraparound Multicomponent Interventions	40
7.3. Fixed and Adaptive Wraparound Interventions	40
8. FUTURE DIRECTIONS	41
9. CONCLUSIONS	42

1. INTRODUCTION

Many psychological interventions are made up of multiple components, each of which is expected to contribute to effectiveness. For example, Watkins et al. (2023) hypothesized that seven distinct components (e.g., functional analysis, concreteness training, self-compassion training) would contribute to the success of a cognitive behavioral therapy (CBT) intervention for treating depression. In this review we take a broad view of multicomponent psychological interventions, including interventions that aim to improve mental and/or physical health. From this perspective, psychological interventions are remarkably diverse. The components may all be behavioral or psychological, or the set of components may include one or more medical elements, such as an antidepressant. They may involve participants at any point along the life course, and they may be aimed at the individual or at an aggregate unit, such as a dyad (e.g., a patient and their care partner), family, neighborhood, medical practice, school, faith congregation, or other organization or entity.

This review is based on the premise that the goals of psychological intervention science are to (a) build a coherent knowledge base about what intervention strategies work, why, for whom, and under what circumstances, (b) leverage this knowledge to develop interventions that have positive societal impact, and (c) make ongoing improvements in intervention strategies to achieve ever-greater positive societal impact on mental and physical health. To date, these goals have been pursued primarily within the classical treatment package research paradigm, which emphasizes establishing whether an intervention as a package of components demonstrates a statistically and clinically significant effect compared to a suitable control. Within this paradigm, establishing that a treatment package has a statistically and clinically significant effect—that is, intervention evaluation—is often considered the logical first step in the process of arriving at an effective and implementable evidence-based intervention. This emphasis on interventions as a package has persisted despite calls for the clinical psychology field to study individual intervention components (Kazdin 2000).

We agree that intervention evaluation is essential. However, for several reasons, we question whether in the long run the classical treatment package paradigm is the best approach to advance the three goals of psychological intervention science. First, the expectation that the effect of a treatment package will be demonstrated before examining the performance of any of its components has the unintended consequence of incentivizing the development of complex, expensive, and inefficient treatment packages that may be too impractical to implement. Investigators have few reasons to limit the number of components, guard against the inclusion of weak, inactive, or even potentially counterproductive components, or consider how resource-intensive the resulting intervention will be, and they have every reason to include as many components as they can think of to guard against null results in an evaluation randomized controlled trial (ERCT). Second, if the intervention package produces a statistically and clinically significant effect but is not practical to implement, then under the classical treatment package paradigm it is unclear what to do to make it implementable. Making the intervention less resource-intensive by removing some components, or setting some to less resource-intensive levels, risks reducing or even eliminating the intervention's effect unless the decisions can be based on information about individual component performance—information that is generally not obtainable within this paradigm. Third, the classical treatment package paradigm provides few data that are useful in identifying the strong and weak parts of an intervention, and therefore there is little systematic basis for ongoing improvement.

In this article we review the multiphase optimization strategy (MOST; Collins 2018, Collins et al. 2021), a different way of thinking about the science of intervention development, evaluation, and implementation. MOST offers an alternative to the classical treatment package paradigm for

pursuing the goals of psychological intervention science. Integrating ideas from behavioral science, industrial and control engineering, implementation science, health economics, and decision science, MOST places intervention optimization as logically precedent to intervention evaluation. The purpose of optimization is to enable the investigator to identify the combination of components that best meets the objective at hand. If the objective is to identify the most effective intervention, optimization can be used to eliminate counterproductive components. However, the objective may not necessarily be to select the most effective intervention in an absolute sense; instead, it is often to identify the most effective intervention that is practical to implement. Implicit in intervention optimization is acknowledgment that placing limits on resource requirements to make an intervention implementable may take a toll on effectiveness to some degree. Strategic trade-offs are often necessary to achieve positive societal impact. To take an extreme, yet not uncommon, example, a highly effective intervention will have little societal impact if its high resource demands prevent its adoption.

Optimization of an intervention typically requires an optimization randomized controlled trial (ORCT). An ORCT uses any of a variety of experimental designs that are appropriate and efficient for this purpose. ORCTs are designed to systematically assess the performance of intervention components. The experimental results, along with cost and other information, are then used as a basis for identifying the set of components and component levels that represents the optimized intervention according to a clearly specified optimization objective. Only after optimization has been performed is the intervention evaluated as a package in an ERCT.

The information gained from ORCTs is useful in building a coherent knowledge base about what intervention strategies work and how those strategies can be improved. The investigator can then conduct subsequent ORCTs to examine new or revised components and reoptimize in an ongoing cycle. This continual optimization enables incorporation of new intervention and implementation strategies suggested by the research literature as well as adaptation to changing societal, public health, economic, and implementation circumstances. This approach has the potential to maintain and even incrementally increase an intervention's positive societal impact over time.

The purpose of this review is to introduce MOST as a paradigm shift in psychological intervention science, discuss its potential contributions to the field of clinical psychology, and provide enough basic information and suggestions for additional literature to enable readers to consider employing this paradigm in their own research. We begin with an overview of MOST (for a comprehensive introduction to MOST, see Collins 2018). We then go into more detail about the factorial ORCT and focus on the particularly efficient 2^k design. This is followed by a discussion of decision making based on the results of a factorial ORCT. Next, we consider optimization of adaptive interventions and two ORCT designs that are useful in this context: the sequential multiple assignment randomized trial (SMART) and the micro-randomized trial (MRT). We briefly review some other ORCT designs. We then take a closer look at the application of MOST in implementation science. We end by offering some ideas for future directions in the rapidly developing field of intervention optimization. Throughout, we highlight the feasibility of adopting this new research paradigm and emphasize its tremendous potential to hasten progress toward the three goals of psychological intervention science.

2. OVERVIEW OF INTERVENTION OPTIMIZATION VIA THE MULTIPHASE OPTIMIZATION STRATEGY

2.1. Definition of Optimization

Optimization is a broad concept that has different definitions in different fields and applications. We use a definition adapted from industrial and control engineering for application to intervention

science: Optimization is the process of achieving a strategic balance of intervention effectiveness,¹ affordability, scalability, and efficiency (Collins et al. 2021). MOST is a principled approach to intervention optimization—in other words, to development of interventions that achieve this strategic balance, which we call intervention *EASE*.

Here, affordability is defined as the “extent to which the intervention is deliverable within budget, and offers a good value” (Collins et al. 2021, p. 2000). Scalability is the “extent to which the intervention is implementable in the intended setting with no need for ad hoc modifications” (Collins et al. 2021, p. 2000). In practice, many factors (e.g., community support) determine whether an intervention ultimately goes to scale. Thus, intervention optimization does not guarantee that an intervention will be scaled up, but it can enable the investigator to balance competing considerations to identify an intervention that is practical to scale up. Efficiency is the extent to which the intervention achieves the desired outcome(s) with minimal resource demands relative to the alternatives. Efficient interventions do not demand participant effort that does nothing to further the intervention’s goals, and therefore they may be more ethical than less efficient alternatives. Efficient interventions may also be more effective if they prevent participant fatigue and are not overwhelming, thereby making it more likely that engagement will be maintained for the duration of the intervention. Affordability, scalability, and efficiency are related but distinct; for example, an intervention could be affordable if it costs less than the upper limit of an implementation budget, but inefficient if it contains inactive components.

2.2. The Concept of Intervention Components

Because intervention optimization using MOST is based on the idea of assessing the individual and combined contributions of intervention components, it is necessary for the investigator to identify intervention components. Intervention components must be conceptually and practically distinct, and although they may, and often will, interact, it must be possible to manipulate them independently in an ORCT. Components may be defined at any level of granularity, from relatively macro (e.g., text messaging as an overall strategy) through relatively micro (e.g., each individual text message). Nearly any aspect of an intervention can be considered a component as long as it is useful and feasible to separate it out for empirical study. For example, a component may correspond to an element of the treatment itself, to a feature included solely to promote adherence to or engagement in an intervention, or to a feature designed to affect the implementation of the intervention, such as specialized training offered to program delivery staff or psychotherapists.

Investigators may find frameworks such as the Behavior Change Wheel (Marques & Guastaferro 2022, Michie et al. 2014) and human-centered design (O’Hara et al. 2022) helpful in identifying distinct intervention components.

2.3. Brief Overview of the Multiphase Optimization Strategy

As **Figure 1** illustrates, MOST comprises three phases: preparation, optimization, and evaluation. In the preparation phase the investigator lays the groundwork for intervention optimization through a series of specific activities (Landoll et al. 2022). An essential part of this preparation is deriving or revising a detailed conceptual model. This is a model of the behavioral process to be intervened upon; it shows where and how each intervention component is hypothesized to have an effect on the process, and it specifies any important moderators hypothesized to alter component effects. It is possible that not all of the components in the conceptual model will be investigated

¹In this review we use the term effectiveness to include both efficacy and effectiveness because interventions can be optimized for either.

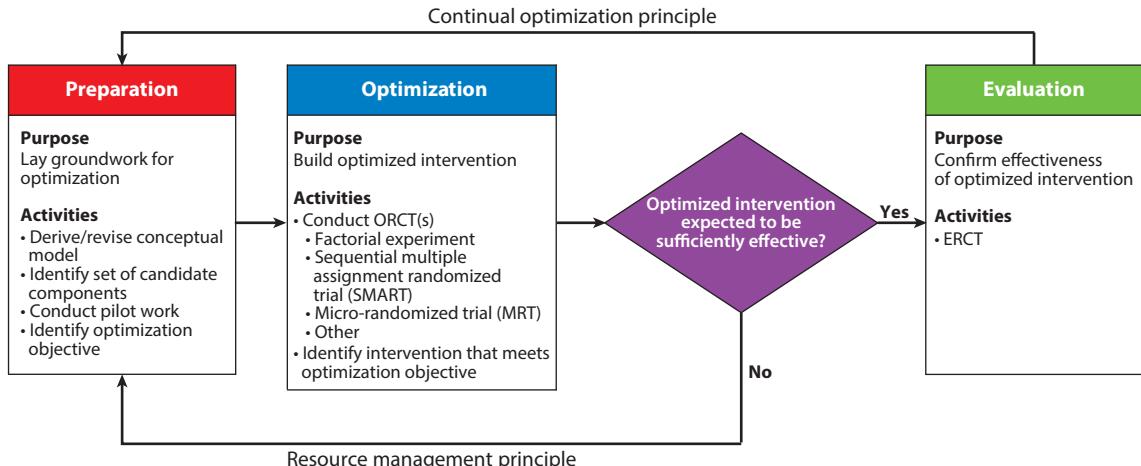


Figure 1

MOST consists of three phases: preparation, optimization, and evaluation. Abbreviations: ERCT, evaluation randomized controlled trial; MOST, multiphase optimization strategy; ORCT, optimization randomized controlled trial. Figure adapted from Collins (2018) with permission from the Springer Nature Customer Service Center.

in a given ORCT. Some components may already have been selected for inclusion in the intervention; for example, their effectiveness may have been established in a prior ORCT, or they may represent an established standard of care. The components that will be examined in the ORCT are called candidate components. The conceptual model is theory based wherever possible, and it may be informed by more than one theory. For detailed examples of conceptual models, readers are referred to Gwadz et al. (2017) and Kugler et al. (2018).

Suppose an investigator, who wishes to use MOST to optimize an intervention for depression based on CBT, plans to examine the following four components (adapted from Watkins et al. 2016): functional analysis (systematic reduction of undesired behaviors and increasing of desired behaviors), relaxation (e.g., progressive muscle relaxation), challenging negative thoughts (e.g., psychoeducation), and activity scheduling (e.g., reducing avoidance). Any pilot work in preparation for the ORCT—whether it is done to test the acceptability, feasibility, or practicality of these intervention components, to pilot the ORCT itself, or for some other reason—is conducted in the preparation phase of MOST.

In the preparation phase, the investigator also identifies the optimization objective (previously called the optimization criterion). This is the investigator's operational definition of intervention EASE. It contains a statement of the desired outcome(s) of the intervention and the salient constraints. For example, if it is known that no more than \$500 per participant will be available to implement the intervention, the optimization objective might be to arrive at the best expected outcome (expressed in terms of the outcome variables) that can be obtained within an upper limit of an implementation cost of \$500 per person. Or, it may be determined that staff can spare no more than 2 hours to implement the intervention. Then the optimization objective might be to arrive at the best expected outcome that can be obtained in 2 hours or less of implementation time. Our hypothetical intervention scientist wishes to identify a value-efficient intervention, which balances predicted effectiveness and implementation costs without placing an upper limit on costs. There are many different possible optimization objectives. In every case, the idea is to arrive at an intervention that represents a realistic and acceptable balance of effectiveness, affordability, scalability, and efficiency.

The ultimate purpose of the next phase of MOST, optimization, is to identify the optimized intervention. Optimization is always completed before the evaluation phase. It typically involves assessing the individual and combined performance of the candidate intervention components by means of an ORCT. In the hypothetical CBT example, the intervention scientist would conduct an ORCT on the four candidate intervention components. **Figure 1** lists three popular ORCT designs, all of which are efficient ways of examining individual components under consideration for inclusion in an intervention: the factorial experiment (Fisher 1946) and two other designs drawn from the family of factorial experiments, namely, the SMART (Kidwell & Almirall 2023, Murphy 2005) and the MRT (Liao et al. 2016, Qian et al. 2022). By contrast, for research in the evaluation phase, MOST uses an ERCT, which may be a standard two-arm clinical trial or use another design from this family. Both ORCTs and ERCTs meet the definition of a clinical trial maintained by the US National Institutes of Health (NIH 2017).

Once the ORCT has been conducted, our hypothetical investigator will use the results to decide which combination of the four components will make up the optimized intervention. As mentioned above, this investigator's optimization objective is to identify a value-efficient intervention—in other words, an intervention that offers good potential effectiveness in relation to anticipated implementation resource demands. This can be a complex decision-making process, particularly if there are interactions among the components or if there are multiple outcomes to be considered. Recent work in decision making for intervention optimization is reviewed in Section 4.

Two fundamental principles of MOST appear in **Figure 1**. One is the continual optimization principle, which states that the investigator has the opportunity to improve the intervention in an iterative fashion over time. This principle suggests that the hypothetical investigator will immediately return to the preparation phase after the completion of the evaluation phase, armed with information about the intervention's strong and weak points gained during the optimization phase. The other fundamental principle is the resource management principle, which suggests that the investigator must make efficient use of research resources. For example, immediately after the optimization phase the investigator should consider whether or not to proceed to the evaluation phase (this step is indicated by the purple diamond in **Figure 1**). If the ORCT data suggest that the optimized intervention is reasonably likely to demonstrate a statistically detectable and clinically significant effect, then the investigator is justified in devoting resources to an ERCT. On the other hand, if the optimized intervention does not appear potentially strong enough to warrant an ERCT, the resource management principle suggests returning to the preparation phase. However, this is markedly different from starting over after a nonsignificant effect in an ERCT. The data from the ORCT indicate which components performed satisfactorily and which did not, providing a firm basis for continued progress toward an intervention with high societal impact. The investigator can reconsider the conceptual model, identify the components that must be revised or replaced, and conduct another ORCT.

The resource management principle also suggests that the investigator must select an ORCT design carefully to make the best use of research resources and collect the highest-quality and most pertinent scientific data. We turn now to one particularly efficient experimental design, the 2^k factorial.

3. THE 2^k FACTORIAL OPTIMIZATION RANDOMIZED CONTROLLED TRIAL

3.1. Introduction to the Factorial Optimization Randomized Controlled Trial

In a factorial ORCT, each candidate component is represented by an independent variable called a factor. The factors are manipulated simultaneously in a coordinated fashion. In a 2^k factorial experiment, there are k factors, each of which has two and only two levels.

Table 1 Illustration of 2^4 factorial experiment

Condition number	Factor			
	<i>FA</i>	<i>RE</i>	<i>CN</i>	<i>AS</i>
1	No	Video	No	No
2	No	Video	No	Yes
3	No	Video	Yes	No
4	No	Video	Yes	Yes
5	No	Relaxation training	No	No
6	No	Relaxation training	No	Yes
7	No	Relaxation training	Yes	No
8	No	Relaxation training	Yes	Yes
9	Yes	Video	No	No
10	Yes	Video	No	Yes
11	Yes	Video	Yes	No
12	Yes	Video	Yes	Yes
13	Yes	Relaxation training	No	No
14	Yes	Relaxation training	No	Yes
15	Yes	Relaxation training	Yes	No
16	Yes	Relaxation training	Yes	Yes

Blue shading highlights the different levels of the factors. The main effect of each factor is the mean of the shaded experimental conditions minus the mean of the unshaded experimental conditions. For example, the main effect of *FA* is the mean of experimental conditions 9–16 minus the mean of experimental conditions 1–8. Abbreviations: *AS*, activity scheduling; *CN*, challenging negative thoughts; *FA*, functional analysis; *RE*, relaxation.

Recall that our hypothetical investigator wishes to examine the performance of four components that are candidates for inclusion in a CBT intervention to treat depression: functional analysis, relaxation, challenging negative thoughts, and activity scheduling. **Table 1** illustrates a factorial ORCT to investigate the performance of these four components. Each component is represented by a factor: *FA* (functional analysis), *RE* (relaxation), *CN* (challenging negative thoughts), and *AS* (activity scheduling). In this experiment, each factor has two levels (i.e., a $2 \times 2 \times 2 \times 2$, or 2^4 , factorial experiment), and there are 16 experimental conditions (for a helpful discussion of approaches to conducting random assignment, see Gallis et al. 2019). For factors *FA*, *CN*, and *AS*, the factor levels are yes (i.e., offered; indicated in the table by blue shading) and no (i.e., not offered). For *RE*, the levels are relaxation training (indicated in the table by blue shading) and video. Factorial experiments have a balanced structure, evident in **Table 1**, in which each level of a factor appears the same number of times at each level of every other factor. For example, consider the “no” level of *FA*. As the table shows, *RE* is set to video four times (conditions 1–4) and relaxation training four times (conditions 5–8). Now consider the “yes” level of *FA*. *RE* is again set to video four times (conditions 9–12) and relaxation training four times (conditions 13–16). A perfectly balanced factorial experiment has this characteristic structure and, in addition, has equal sample sizes across experimental conditions.

The logical underpinnings of any factorial experiment are different from those of an ERCT. When properly analyzed using factorial analysis of variance (ANOVA), a 2^k factorial experiment enables estimation of the main effect of each of the k factors ($k = 4$ in the example discussed here) and interactions involving two or more factors up to the k -way (in the example, four-way) interaction. When, as we recommend (Kugler et al. 2018), the analysis is based on effect coding [(-1, 1) in the case of two levels in a factor], two different types of quantities are estimated. The

main effect of a factor is an estimate of that factor's effect averaged across the remaining factors. The interaction between, say, factor *A* and factor *B* reflects the degree to which the effect of factor *A* varies depending on the level of factor *B*, averaged across the remaining factors. [An ANOVA based on dummy (0, 1) coding yields the same omnibus *F* but estimates of different quantities; see Kugler et al. 2018.]

Effect estimates based on an ERCT are obtained by direct comparison of the means of experimental conditions. By contrast, main and interaction effect estimates based on a factorial experiment are obtained by comparing the means of specific combinations of experimental conditions. This can be illustrated by returning to **Table 1**. The main effect of the factor *FA* is estimated by comparing the mean of the eight conditions in which *FA* is set to yes (indicated by blue shading in the table) to the mean of the eight conditions in which *FA* is set to no. Thus, the main effect of *FA* is defined as the effect of *FA* averaged across the remaining factors. In this 2^4 experiment, data from all participants contribute to this effect estimate. Similarly, the main effect of the factor *RE* is estimated by comparing the mean of the eight conditions in which *RE* is set to yes (indicated by blue shading in the table) to the mean of the eight conditions in which *RE* is set to no. Again, data from all participants contribute to this effect. The same logic holds for the estimate of the main effect of *CN* and the main effect of *AS*. An interaction involving, say, *FA* and *RE* is said to occur if the effect of *FA* differs across the levels of *RE*, averaged across all levels of the other factors in the experiment.

In a properly conducted factorial ANOVA, each participant contributes to the estimation of multiple effects; therefore, factorial experiments typically use many fewer participants than alternative approaches that have been used in clinical psychology, such as dismantling studies and other component designs. As Papa & Follette (2015, p. 1) explained, "Component designs include either adding or subtracting specific elements of a treatment protocol to assess the relative effects of the component in the context of the other elements in the design." Component designs involve direct comparison of individual experimental condition means and therefore have statistical properties similar to those of the ERCT. Consider a component design that comprises five experimental conditions: one control condition and four experimental conditions, each setting one of the components to yes and the others to no. Compared to a factorial experiment, the component experiment would require 2.5 times the number of participants to achieve the same level of statistical power. A separate experiment for each component would require four times as many participants. In both cases, the quantity being estimated confounds main effects and interactions (Collins et al. 2009). The SMART and the MRT, reviewed below, are in the general family of factorial experiments and are similarly analyzed by aggregating experimental conditions in a principled and efficient manner.

We wish to highlight two related characteristics of the factorial experiment, either of which may initially appear counterintuitive to those trained primarily in ERCTs. First, because effects are estimated based on combinations of experimental conditions, the power associated with a factorial experiment is only indirectly related to the per-condition sample size. Instead, power is directly related to the sample size per *level* of a factor. For this reason, factorial experiments can have very small per-condition sample sizes and nevertheless be well powered if the aggregate sample size per level is sufficiently large. Second, it is often possible to add one or more factors to a factorial experiment without needing to add more participants to maintain power if the additional effects to be estimated are no smaller than the ones upon which the original power analysis was based (assuming the results are to be interpreted from a decision-priority perspective; see Section 3.4). Although adding one or more factors increases the number of experimental conditions and therefore results in a smaller per-condition sample size (and potentially adds logistical complexity), if the sample size per level of a factor is identical, power will not be appreciably diminished.

The factors in a factorial experiment can, in theory, have any number of levels. However, as we have mentioned, the 2^k experiment is particularly efficient. Consider a factorial ANOVA conducted within the regression framework and using effect coding. In this case, to power a 2^k factorial experiment it is necessary only to identify the smallest regression coefficient of scientific interest (power for factorial experiments is usually based on main effects) and power the experiment to detect that one. Any regression coefficient equal or greater in magnitude will be detectable with at least the selected level of power (Cohen 1988). Compared to a 2^k design, a factorial design that includes more than two levels of even one factor can require a substantial sample size increase, often $\geq 50\%$.

3.2. A Few Words About Interactions

In our view, understanding interactions is essential for building a coherent intervention science knowledge base and for ongoing improvement of psychological interventions. Only the factorial family of experiments enables estimation of interactions among components. As mentioned above, in the particularly efficient balanced 2^k experiment, all else being equal, statistical power is identical for detection of any regression coefficient of a given size. This holds whether the regression coefficient corresponds to a main effect or an interaction of any order (again assuming that effect coding is used in the analysis; for more about statistical power for detection of interactions, see Collins 2018, chapter 4). For readers tempted to avoid complexity by choosing experimental designs that do not feature interactions, we wish to point out that if an interaction is present, it will manifest itself in data no matter what experimental design is used, but only factorial designs enable one to disentangle the interaction from other effects (Collins et al. 2009).

3.3. The Concept of Control in a Factorial Experiment

The concept of control is different in a factorial experiment compared to an ERCT. In an ERCT there is typically a single control arm, whereas in a factorial experiment there is no single experimental condition that serves as “the” control condition. Instead, one level of each factor is the designated control for that factor. In **Table 1**, the conditions in which the factor is set to the “no” level serve as the control for that factor. For the *FA* factor, the aggregate of conditions 1–8 serves as the control; for *RE*, the aggregate of conditions 1–4 and 9–12 serves as the control, and so on. The factors may have different types of controls; in this example, three of the factors use a “no” level as the control, whereas a video is used as the control for the *RE* factor.

3.4. Conclusion-Priority Versus Decision-Priority Perspectives

There are two related but distinct perspectives an investigator can take on the data from a 2^k experiment, or indeed from any ORCT (Collins 2018). From a conclusion-priority perspective, the purpose of an experiment is to draw scientific conclusions about a hypothesized effect, which can be put under the scrutiny of fellow scientists. This perspective requires, for example, adhering to well-established conventions about type I ($\alpha \leq 0.05$) and type II ($\beta \leq 0.20$) error rates. By contrast, from a decision-priority perspective, the purpose of an experiment is to use available resources efficiently to provide information that will enable the investigator to decide which components and component levels will constitute the optimized intervention, which may then be evaluated further. From this perspective, the investigator is less constrained by scientific conventions and instead may use the experimental data in any reasonable and transparent manner. The investigator may, for example, wish to raise the type I error rate above the conventional level to reduce the type II error rate, and accept an increased probability of mistakenly including an inert component to reduce the probability of overlooking a potentially useful one.

The same data may be approached from either perspective depending on the scientific objective. For example, we do not advocate using a Bonferroni or similar adjustment to maintain a specific type I error rate when using the results of an ORCT for decision-making purposes. When conducting analyses intended to test hypotheses about whether particular components have a statistically detectable effect, or when testing mediation (Piper et al. 2017, Strayhorn et al. 2022) or moderation (Piper et al. 2017) hypotheses with data from an ORCT, such an adjustment could be considered. A conclusion-priority perspective is typically appropriate when conducting an ERCT.

4. DECISION MAKING AND INTEGRATION OF RESOURCE CONSIDERATIONS

Suppose our hypothetical investigator has conducted the 2^4 factorial ORCT illustrated in **Table 1**. The 16 experimental conditions represent different combinations of components, each of which is an alternative version of the intervention, and the goal of the optimization process is to select the alternative that best meets the optimization objective. How can the investigator accomplish this?

Collins et al. (2014) recommended a component screening approach for identifying the optimized intervention based on the results of a factorial experiment. This approach is a systematic way of sorting through the many effect estimates provided by a factorial ANOVA to screen out underperforming components. However, a component screening approach has important limitations; for example, it cannot accommodate more than one outcome variable of interest, it can be laborious, and complex patterns of interactions can make decision making difficult. In this section we summarize some recent work (Strayhorn et al. 2023, 2024b) that proposes a new and demonstrably better approach for selecting the components that make up the optimized intervention. The new approach to decision making relies on relative comparisons among the posterior expected outcomes associated with each alternative intervention under consideration (Strayhorn et al. 2023).

When the optimization objective involves simply maximizing expected effectiveness on a single outcome variable of interest, choosing an optimized intervention can be relatively straightforward. Suppose the outcome variable in the hypothetical ORCT is a mental health inventory, where a higher score indicates better mental health. Strayhorn et al. (2023) recommended conducting a factorial ANOVA by fitting a regression equation using a Bayesian approach, thereby producing a posterior expected outcome—in other words, a \hat{Y} —for each experimental condition. In our example, the optimized intervention would be the combination of components and component levels associated with the best expected outcome—in this hypothetical case, the largest \hat{Y} on the mental health inventory. Strayhorn et al. (2023) demonstrated via statistical simulation that this vastly easier approach works better than the component screening procedure. If instead the optimization objective is to select, say, the combination of components that produces the best expected outcome subject to an upper bound on implementation cost, then the most favorable \hat{Y} corresponding to an alternative under that limit can be selected.

However, recall that our hypothetical investigator wishes to include cost in decision making. Suppose per-person implementation costs for each component level are as follows: functional analysis, \$400; relaxation, \$100 for the video and \$200 for relaxation training; challenging negative thoughts, \$300; and activity scheduling, \$500. The alternative intervention costs range from \$100 for the version where only the relaxation video is provided to \$1,400 for the version where all four components, including relaxation training, are provided. As mentioned above, this hypothetical investigator's optimization objective did not specify an upper limit on cost but instead stated the desire for a value-efficient intervention. In other words, this investigator's definition of

intervention *EASE* is a strategic balance of anticipated effectiveness (defined in terms of the mental health measure) against resource demands (defined in terms of per-person implementation cost).

This now becomes a multicriteria decision problem, in which the criteria are the outcome variable—the mental health inventory—and cost. In any multicriteria decision problem, the alternative interventions under consideration are likely to differ in their performance on the various criteria, often in ways that complicate the selection of an optimized intervention considerably. Because it may not be possible to simultaneously achieve the “best” on every criterion that matters to investigators, selecting one optimized intervention can require that investigators confront challenging trade-offs among criteria, such as the not-unusual trade-off between effectiveness and cost.

Decision analysis for intervention value efficiency (DAIVE; Strayhorn et al. 2024b) is an emerging decision-making framework that facilitates this sort of complex multicriteria decision making in the selection of an optimized intervention. To facilitate decision making, DAIVE (*a*) plots the alternative interventions according to their expected outcomes and their costs on an x - y plane and (*b*) highlights the “value-efficient” alternatives—those that cost less than alternatives while producing the same or better expected outcomes. **Figure 2** shows this plot for our hypothetical example (based on artificial data). Each plot point represents the expected outcome (normed to range between 0 and 1) on the x axis and anticipated per-person delivery cost on the y axis for 1 of the 16 combinations of components (i.e., alternative interventions) included in the 2^4 ORCT. The subset of value-efficient alternatives (highlighted and connected in the figure with a

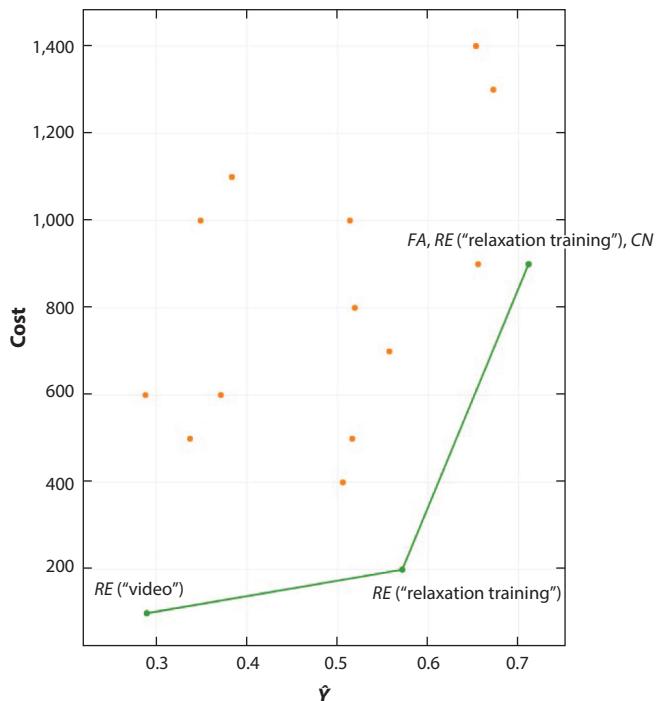


Figure 2

Graph showing expected outcome (x axis) and anticipated per-person delivery cost (y axis) of each of a set of alternative versions of an intervention to deliver cognitive behavioral therapy. Each point represents a combination of candidate components. The green line shows the value efficiency frontier. Figure based on artificial data. Abbreviations: CN, challenging negative thoughts; FA, functional analysis; RE, relaxation.

solid line) makes up the value efficiency frontier. The value efficiency frontier by definition begins with the alternative that costs the least and ends with the alternative that is associated with the best expected outcome. Each intervention on the value efficiency frontier could technically be chosen as the optimized intervention depending on the decision maker's willingness to commit additional resources to pay for improved effectiveness.

In **Figure 2** there are three alternative interventions on the value efficiency frontier: the minimal intervention, which consists only of the relaxation video; the intervention consisting of relaxation training only; and the intervention consisting of functional analysis, relaxation training, and challenging negative thoughts. This last intervention is associated with the best expected outcome, which makes it a great option for the optimized intervention. However, at \$900 it is also the most expensive alternative on the value efficiency frontier. The alternative that contains only relaxation training delivers an expected outcome that is relatively close to the best expected outcome—and at considerably lower cost (\$200). An investigator open to strategically trading some effectiveness for a lower delivery cost in this way might tentatively select the intervention made up of relaxation training only, rather than the one expected to be more effective but also costlier. Before a definitive selection is made, results would be further contextualized using examination of main and interaction effects and, in some cases, additional steps to determine whether continuing on to the evaluation phase of MOST and an ERCT would be justified.

The above example offers one simple illustration of DAIVE based on a single intervention outcome (the mental health measure) and one particular cost (monetary per-person cost of intervention delivery). One exciting advantage of DAIVE is the flexibility to accommodate optimization objectives in which effectiveness is defined in terms of multiple outcomes (Strayhorn et al. 2024b)—for instance, several health outcomes, or one health outcome plus one empirical implementation-related outcome variable, and so on. In these cases, the multiple outcomes are combined using a value function carefully selected to reflect the investigators' preferences, including which outcomes, if any, matter more than others, and under what circumstances. Selection of the optimized intervention now involves relative comparisons among the expected values associated with each alternative intervention based on the selected value function. Furthermore, because cost in DAIVE can be defined in terms of any limited resource that is relevant for the implementation of an intervention, efficiency can be considered in any number of ways. For example, DAIVE could be applied to strategically balance anticipated effectiveness on one or more outcomes with efficiency in the use of provider time.

5. OPTIMIZATION OF ADAPTIVE INTERVENTIONS

5.1. Fixed and Adaptive Intervention Strategies

It is helpful to distinguish between fixed and adaptive (Almirall & Chronis-Tuscano 2016, Collins et al. 2004) interventions because this can be a critical consideration in choice of ORCT design. Fixed interventions, like the hypothetical CBT example above, call for all participants to be offered the same treatment in a uniform manner. There may in practice be variability in treatment, such as when a participant does not engage in every aspect of treatment, but this unplanned variability is not a part of the treatment strategy. By contrast, an adaptive intervention calls for the treatment to be varied in a prespecified and principled fashion across individuals and within individuals over time. The decisions about what treatment to offer an individual are made according to a set of decision rules at key decision points on the basis of measured tailoring variables. Adaptive interventions vary in terms of the timescale of adaptation (Montgomery et al. 2011; Nahum-Shani et al. 2018b, 2022b). Interventions with slower timescales of adaptation have fewer decision points—that is, fewer opportunities to consider changing the treatment—spaced farther

apart in time, whereas interventions with faster timescales of adaptation have more decision points spaced closer together.

There are two key reasons to consider employing an adaptive rather than a fixed treatment strategy. One is to attempt to produce better outcomes—for example, by providing additional treatment for participants who do not obtain a satisfactory result with an initial treatment. A second reason is to attempt to conserve resources and increase efficiency by providing no more treatment than is needed; this has the potential to enhance intervention affordability, scalability, and efficiency and to produce better outcomes by minimizing participant burden and habituation.

As demonstrated below, in addition to all the types of components that are investigated in optimization of fixed interventions, in optimization of adaptive interventions the tailoring variable, the timing or number of decision points, or any aspect of the decision rule can be considered a component and investigated in an ORCT.

5.2. Optimization of Adaptive Interventions with Relatively Slow Timescales of Adaptation: The Sequential Multiple Assignment Randomized Trial

The SMART is an ORCT used for optimization of adaptive interventions with relatively slow timescales of adaptation. Consider a hypothetical adaptive intervention aimed at encouraging adherence to HIV treatment, consisting of taking a daily pill. Compliance is measured daily via an electronic pill bottle cap. The outcome, adherence, is operationalized as the number of daily pills taken during the final month of the intervention: days 90–120. All participants are offered encouraging text messaging, peer mentoring, and behavioral skills training extending over an initial 45-day period. At 45 days, adherence is assessed in terms of the number of daily pills taken during this time. Those who are designated adherent according to a prespecified cutoff score (discussed further below) are provided with encouraging text messaging going forward, and no other treatment. Those who are designated nonadherent continue to receive encouraging text messaging and peer mentoring and, in addition, are offered motivational interviewing. In this simple adaptive intervention, there is a single decision point, which occurs at 45 days. The measure of adherence serves as the tailoring variable. The decision rule is as follows: “If adherent, step down to text messaging only. If nonadherent, step up by adding motivational interviewing.”

Now suppose an investigator wishes to optimize an adaptive intervention similar to the one described above. The investigator has identified the following research questions:

1. Which operational definition of the tailoring variable results in more daily pills taken during days 90–120: adherence defined as taking the pill (a) ≥ 35 times or (b) ≥ 40 times during an initial 45-day period?
2. For those participants designated as adherent at day 45, which is a better second-line treatment: (a) step down to text messaging only or (b) continue with text messaging and peer mentoring?
3. For those participants designated as nonadherent at day 45, which is a better second-line treatment: (a) continue with text messaging and peer mentoring or (b) step up by adding motivational interviewing?

Thus, three candidate components are being investigated: the operational definition of the tailoring variable, adherence; the second-line treatment strategy for adherent participants; and the second-line treatment strategy for nonadherent participants.

These three research questions can be addressed in the SMART illustrated in **Figure 3**. In this experiment, initially all participants are offered text messaging, peer mentoring, and behavioral skills training; this is a given and is not examined in the experiment. In the first randomization on day 1, each participant is assigned to one of two operational definitions of adherence. The next

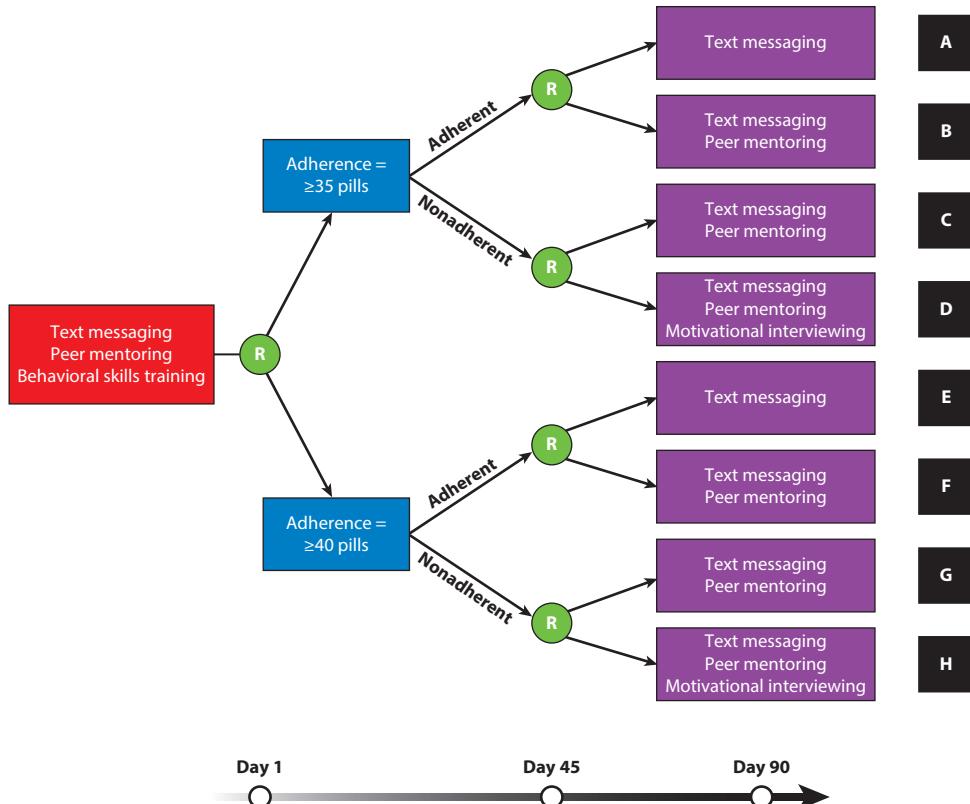


Figure 3

Illustration of a sequential multiple assignment randomized trial (SMART). Each participant is randomized twice. On day 1, each participant is provided with the same intervention and randomized to an operational definition of the tailoring variable, adherence. On day 45, each participant is randomized to second-line treatment, conditional on whether they are designated adherent or nonadherent based on their assigned definition of the tailoring variable. Figure adapted from Collins (2018) with permission from the Springer Nature Customer Service Center.

randomization, on day 45, is conditional on whether the participant is designated adherent or nonadherent based on their assigned operational definition. Those designated adherent are randomly assigned to one of two second-line treatments; similarly, those designated nonadherent are randomly assigned to one of two second-line treatments. This design includes eight experimental conditions (labeled A–H in **Figure 3**).

The SMART is a type of factorial design, and it uses the same logic of comparing means based on aggregates of experimental conditions. However, in a SMART some research questions pertain to all participants, whereas others pertain only to certain subgroups, reducing the sample size available for those questions. Question 1, the comparison of the two operational definitions of the tailoring variable, pertains to all participants and can be addressed by comparing the mean outcome across conditions A–D to the mean outcome across conditions E–H. Question 2 pertains only to participants designated as adherent. It can be addressed by comparing the mean outcome of conditions A and E to that of conditions B and F. Similarly, question 3, which pertains only to participants designated as nonadherent, can be addressed by comparing the mean outcome of conditions C and G to that of conditions D and H. Because some research questions

will be addressed based on a subset of participants, careful planning is required to ensure sufficient power in a SMART (Dziak et al. 2024, Kidwell et al. 2018, Yap et al. 2023).

5.3. Optimization of Adaptive Interventions with Relatively Fast Timescales of Adaptation: Just-in-Time Adaptive Interventions and Micro-Randomized Trials

Just-in-time adaptive interventions (JITAIs; Nahum-Shani et al. 2018b) have relatively fast timescales of adaptation—daily, hourly, even minute-to-minute. They use dynamically changing information about the individual’s internal state (e.g., craving, stress) and context (e.g., physical location), usually obtained from self-reported and/or sensed data from wearables and smart devices, to adapt intervention delivery to the individual’s state and context in daily life. JITAIs have particularly high potential to minimize burden and avoid habituation because technology enables them to deliver an intervention only when necessary (i.e., only when states of vulnerability and/or opportunity occur) and only when the individual is receptive and likely to engage with the intervention, where engagement is defined as the investment of energy in the intervention (Nahum-Shani et al. 2022c).

For illustrative purposes, consider a simple hypothetical JITAI designed to increase self-monitoring of dietary intake in overweight and obese adults. Obese/overweight adults who are motivated to lose weight are asked to specify three windows of time in which they typically eat their main meals (breakfast, lunch, and dinner). Then, participants receive a weight loss app for their mobile devices. The app enables each participant to record their dietary intake and to view relevant graphical feedback. At three decision points per day, at 1 hour after the person’s prespecified mealtime (for breakfast, lunch, and dinner), if dietary intake for the current meal was not recorded in the app, then a prompt is delivered immediately via the app (if the person is not driving), encouraging the individual to self-monitor their dietary intake. Otherwise, no prompt is delivered.

The MRT is an ORCT that is used in optimizing a JITAI. Like the SMART, the MRT involves sequential randomizations. However, compared to a SMART, in an MRT there are many more randomizations, and they occur much more rapidly: Each individual is randomized at each of many decision points over the course of the study, and because of the fast timescale of adaptation that characterizes the JITAI, the length of the time interval between randomizations is relatively short (e.g., hours).

Suppose investigators developing a weight loss intervention would like to answer the following question about a JITAI for promoting dietary intake self-monitoring: Does sending (versus not sending) a reminder 1 hour before a meal increase the likelihood of self-monitoring in the next 2 hours? To answer this question, the investigators design the following (hypothetical) MRT: At the beginning of the study, obese/overweight adults are provided with the weight loss app as described above and are asked to specify the windows of time in which they typically eat their main meals. During a 12-week period, the app passively collects information about each participant’s self-monitoring behaviors. Three times daily, 1 hour before their prespecified mealtime, each participant is randomly assigned (with 0.50 probability) to receive either no message or a text message (again, if they are not driving) encouraging them to self-monitor their dietary intake. The proximal outcome—participant self-monitoring of dietary intake during the following 2 hours (binary outcome)—is assessed following each randomization.

The hypothetical MRT discussed above includes a single experimental factor that can take either of two levels—prompt versus no prompt—on each of three decision points per day over the entire 12-week period. The question motivating this MRT can be addressed by pooling outcome data across all 252 decision points (3 decision points per day \times 7 days per week \times 12 weeks) and N individuals. Specifically, this will involve comparing the probability of self-monitoring in the next

2 hours following all the decision points in which a prompt was delivered (across all individuals) to the probability of self-monitoring in the next 2 hours following all the decision points in which a prompt was not delivered (across all individuals). Because MRTs enable investigators to answer scientific questions about intervention effects based on a combination of between-person contrasts and within-person contrasts in the proximal outcome, they typically require fewer experimental participants relative to comparable between-subject experiments (Qian et al. 2022).²

This hypothetical example includes a single factor, but MRTs can include multiple factors. Whereas in a standard factorial design like the 2^k discussed above, the main effect of a factor is computed by averaging over the levels of the other factors, in MRTs the main effect of a factor is computed by averaging over the levels of the other factors and all decision points (Qian et al. 2022). The meaning of main effects in an MRT is different from that in standard factorial designs. Because MRTs involve many decision points in which individuals are randomly assigned to intervention options, main effects are interpreted in reference to how intervention options were delivered prior to a given decision point (for more details, see Qian et al. 2022). In the current example, the schedule of intervention delivery before each decision point involves random assignment with 0.50 probability to either sending or not sending a reminder at all prior decision points. Hence, the main effect of sending (versus not sending) a reminder at a given decision point assumes that the reminder was sent about 50% of the times at previous decision points. This effect would not generalize well to settings in which a reminder was sent 10% or 90% of the times. Thus, the randomization probabilities in an MRT shape the meaning of the main effect; this is different from standard factorial designs where the randomization probabilities concern the design of the trial and are mainly guided by considerations related to power for detecting the effects of interest. In MRTs the randomization probabilities are not only part of the experimental design but also hold scientific meaning and practical implications for intervention design.

In an MRT, as in a SMART, random assignment may be restricted to specific conditions, conditional on a measured variable. Restrictions may be informed by prior empirical evidence or practical considerations suggesting that certain JITAI options should not be considered under specific conditions. For example, individuals may not be eligible to be randomized to be sent/not sent a reminder if they are driving a car or if they activate the “do not disturb” mode on their smartphone (Nahum-Shani et al. 2021).

In many ORCTs and ERCTs the primary outcome is distal—that is, intended to capture phenomena that unfold or change over a relatively long period of time. Thus, the primary outcome is typically measured at multiple later time points following randomization (e.g., weight measured at months 3, 4, 5, and 6). However, the primary outcome in an MRT is proximal—that is, intended to capture phenomena that unfold or change over a relatively short period of time. In the hypothetical example above, the intervention was aimed at distal weight loss, but the proximal outcome was self-monitoring. Thus, the primary outcome in an MRT is always measured soon after each randomization and is measured repeatedly because there are repeated randomizations.

6. OTHER EXPERIMENTAL DESIGNS USEFUL IN INTERVENTION OPTIMIZATION

6.1. Fractional Factorial Designs

Fractional factorial designs are used to address the same scientific questions as corresponding standard factorial designs but use only a reduced number, or fraction, of the experimental conditions

²Differences and similarities between MRTs and traditional within-subject designs are discussed by Walton and colleagues (2018).

while requiring the same sample size. Fractional factorial designs are worth considering when (a) reducing the logistical complexity of implementing a large number of experimental conditions would increase efficiency or (b) cluster randomization must be used (see Section 6.2), and the overall sample size provides adequate statistical power but the number of clusters is insufficient to populate the experimental conditions (Dziak et al. 2012). In selecting a fractional factorial design, the investigator explicitly trades precision of effect estimates for efficiency because fractional factorial designs result in some effects being combined with others (this phenomenon, called aliasing, occurs in designs from the ERCT family too; see Collins et al. 2009). Examples of fractional factorial ORCTs include those described by Collins et al. (2011), Gwadz et al. (2017), Watkins et al. (2016), and McClure et al. (2012). For an introduction to fractional factorial designs, readers may consult Collins (2018), and for a comprehensive overview, readers are referred to Wu & Hamada (2011). Before making a final selection of a fractional factorial design, it is a good idea to seek input from an expert on experimental design.

6.2. Optimization Randomized Controlled Trials with Cluster Randomization

Cluster randomization is frequently necessary in ORCTs in school-based research (Caldwell et al. 2012), implementation science (Kilbourne et al. 2014), and other areas because (a) frequent interaction between participants within a cluster would lead to contamination between experimental conditions and/or (b) outcomes at both the individual and cluster levels are of interest. Clusters may be naturally occurring (e.g., schools) or experiment-induced (e.g., therapy groups). Compared to randomization of individual participants, cluster randomization in an ORCT or ERCT usually requires more participants to achieve the same level of statistical power (Murray 1998). As mentioned in Section 6.1, because ORCTs tend to have many experimental conditions, there may be enough individual participants (e.g., students) to achieve a desired level of statistical power, but not enough units (e.g., schools) to populate the experimental conditions. In this case, a fractional factorial design may be a good choice. For more about cluster randomization in ORCTs, readers are referred to Dziak et al. (2012), Nahum-Shani & Dziak (2018), Almirall et al. (2018), and Nahum-Shani et al. (2018a).

6.3. Evaluation-Optimization Trials

The continual optimization principle of MOST suggests that investigators should be thinking about further improvements to their optimized interventions on a continual basis. Suppose an investigator's next step is an evaluation of an optimized intervention, and at the same time a set of components that are candidates to be added to the optimized intervention has been identified. A time-saving alternative to conducting a separate ERCT and ORCT is the evaluation-optimization trial (E-ORCT), which enables evaluation of the optimized intervention along with examination of the performance of a set of candidate components. For example, Tanner et al. (2021) had previously optimized itMatters, an intervention aimed at preventing excessive alcohol use and risky sex in college students (Wyrick et al. 2020). In the process of designing the ERCT, the investigators learned that college administrators who would ultimately adopt the intervention were keenly interested in adding a component on prevention of sexual victimization. As a response, Tanner et al. (2021) conducted a 2^2 experiment rather than a traditional two-arm ERCT. One factor, which examined the performance of the optimized intervention, had the levels "optimized intervention" and "control"; the other factor, which examined the performance of the sexual victimization component, had the levels "included" and "not included." It would have been arguably more efficient to include additional factors corresponding to additional candidate components. The E-ORCT is appropriate only when the candidate components are to be added to the optimized intervention, not when the candidate components are under consideration as substitutes.

In the E-ORCT, the ERCT is one of the factors in a factorial experiment—that is, it is crossed with all remaining factors. This is emphatically not the same as adding a separate control “arm” to an ORCT with the objective of conducting an ERCT by comparing its mean to that of whichever experimental condition ultimately corresponds to the optimized intervention. We do not in general advocate adding a separate control “arm” to any ORCT because it usually wastes resources; an efficient ORCT is unlikely to contain enough participants in any one experimental condition to provide adequate power for an ERCT.

6.4. Multiple-Timescale Optimization Randomized Controlled Trial Designs for Optimization of Adaptive Interventions

Adaptive interventions often involve multiple timescales. In some settings, decisions about certain components can be made only once (i.e., at a single decision point), whereas decisions about other components can be made several times during the intervention (i.e., at multiple decision points). The hybrid experimental design (HED) is a new experimental approach developed to help investigators answer scientific questions about the delivery and adaptation of multiple intervention components on multiple timescales—slow and fast (Nahum-Shani et al. 2022a). The HED involves sequential randomizations on multiple timescales corresponding to the proper timescale at which each component should be considered. The HED is highly flexible and can take various forms depending on the scientific questions motivating the study. For example, suppose that investigators would like to answer the following questions about the development of a 12-week weight loss intervention that integrates a mobile app and coaching:

- Question A: Does sending (versus not sending) a reminder before each main mealtime benefit self-monitoring in the next 2 hours?
- Question B: Does delivering (versus not delivering) a brief coaching session at weeks 0 (program entry), 4, and 8 amplify the proximal benefits of sending (versus not sending) a reminder before each main mealtime?

These questions and others can be answered via an HED in which participants are randomly assigned three times daily, before each main mealtime, to either receiving a text message that encourages them to self-monitor their dietary intake or receiving no message (with 0.50 probability). Additionally, the HED will randomly assign participants at weeks 0, 4, and 8 to either a brief coaching session or no coaching session (with 0.50 probability). Here, randomizations are employed at two timescales: every few hours and every 4 weeks. There are many other variations of the HED, which vary depending on the scientific questions motivating the study (Nahum-Shani et al. 2023).

6.5. System Identification

Compared to the other approaches discussed in this review, system identification represents a different perspective on intervention optimization, one that has emerged from control systems engineering (Guo et al. 2022, Rivera et al. 2018). From this perspective, human behavior can be viewed as a dynamical system, and adaptive interventions with fast timescales of adaptation can be viewed as engineering controllers built to influence the system to achieve a specified outcome. System identification, which typically requires collection of intensive longitudinal data by methods such as ecological momentary assessment (Shiffman et al. 2008) and frequently involves experimentation, forms the basis for modeling the dynamical system. Standard engineering approaches, such as model predictive control, can be used to develop a controller. Some applications of this approach include development of an adaptive intervention to moderate gestational weight gain (Downs et al. 2021) and identification of the appropriate intervention dose (Voils et al. 2014).

7. OPTIMIZATION AND IMPLEMENTATION SCIENCE: WRAPAROUND INTERVENTIONS

7.1. The Multiphase Optimization Strategy and Implementation Science

As applied in the field of clinical psychology, implementation science aims to determine how psychological interventions can fulfill their potential for positive societal impact (Stirman & Beidas 2020)—for example, by studying how to ensure fidelity of intervention delivery or expand an intervention’s reach. MOST intersects with the field of implementation science in at least two ways. First, as discussed above, MOST can be used to develop and evaluate an intervention that has the potential to achieve a positive societal impact because it is designed to meet the demands and constraints of the intended real-world setting from the outset. This can be accomplished by using an optimization objective that includes considerations such as implementation budget, demands on staff time, and so on in decision making. Second, MOST can be used to optimize an implementation strategy to facilitate an intervention’s adoption, implementation, dissemination, and scale-up (Curran et al. 2012, Leeman et al. 2017, Proctor et al. 2013). In this section we highlight this second way in which MOST intersects with implementation science.

7.2. Implementation Strategies as Wraparound Multicomponent Interventions

It is widely recognized that appropriate multicomponent implementation strategies are necessary to achieve desired implementation outcomes (Aarons et al. 2011, Baker et al. 2010, Perry et al. 2019, Powell et al. 2017, Weiner et al. 2012). In fact, more than 70 implementation strategies have been identified, falling into six categories: planning, educating, financing, restructuring, managing quality, and attending to policy (Powell et al. 2012, 2015). Implementation frameworks, such as the Consolidated Framework for Implementation Research (Damschroder et al. 2009, 2022), may be used to identify barriers that implementation strategies must overcome to achieve the goals of adoption, implementation, or dissemination. However, there has yet to be much empirical examination of individual implementation strategies, so it is unknown which implementation strategies are most effective and whether (or how) they may interact.

Optimizing multicomponent implementation strategies is an important next step for the fields of intervention optimization and implementation science. Consider a psychological intervention that has been optimized, evaluated, and deemed ready for implementation. Now, to achieve a high degree of positive societal impact, an effective, affordable, and efficient multicomponent strategy for implementing this intervention must be developed. MOST can be applied to optimize a multicomponent implementation strategy, in much the same way other interventions are optimized, if the implementation strategy is considered an intervention itself—a “wraparound” with respect to the psychological intervention (Guastaferro & Collins 2021).

7.3. Fixed and Adaptive Wraparound Interventions

A wraparound intervention may be fixed or adaptive. For example, Emily Feinberg and colleagues are optimizing a fixed wraparound intervention for the implementation of Family Navigation (R01 MH117123). Family Navigation, an evidence-based care management strategy designed to reduce disparities in care, is a complex intervention comprising multiple strategies (i.e., motivational interviewing, problem solving, education, care coordination) that can be implemented in multiple ways (Broder-Fingert et al. 2019). The investigators are conducting a 2^4 factorial experiment to identify which implementation strategies, independently and in combination, contribute to a family’s achievement of a behavioral health service goal in the shortest period of time (i.e., within 90 days of randomization). At this writing, the project is ongoing.

Kilbourne et al. (2014) used a SMART to design an adaptive wraparound intervention to improve the success of Replicating Effective Programs (REP). REP is an effective strategy designed to support implementation of evidence-based programs in care settings. However, when REP was applied to the implementation of a psychosocial evidence-based program for mood disorders, there was low uptake (<50%) across sites, which led to minimal improvements in clinical outcomes (e.g., depression symptoms). It was hypothesized that REP needed to be enhanced with other implementation strategies in some settings. The SMART enabled investigators to compare the effectiveness of enhanced implementation strategies among clinics not responding to initial implementation support (i.e., REP). Results indicated that patients from clinics that were randomized to receive the more intensive implementation support did not have significantly improved clinical outcomes compared to patients at clinics randomized to receive a lower-intensity implementation support (Smith et al. 2019). For other examples of SMARTs being used to develop adaptive implementation strategies, readers are referred to Smith et al. (2022) and Fernandez et al. (2020).

8. FUTURE DIRECTIONS

In this section we list a number of interesting future directions in this rapidly evolving field.

Much of the literature on intervention optimization has drawn a sharp distinction between fixed and adaptive interventions, but in practice many interventions incorporate elements of both. Moreover, many adaptive interventions involve a variety of timescales of adaptation. More ORCT designs are needed to enable efficient optimization of such interventions. In addition, methods are needed to enable inference from ORCTs conducted in smaller-sample situations. One approach would be to extend *N*-of-1 trial methods to ORCTs. Another would be to make more effective and efficient use of experimental information via Bayesian methods, including sequential analysis and the use of empirical priors based on trials conducted in similar circumstances.

The concept of the optimization trial could be expanded beyond ORCTs. Causal inference procedures could be developed to enable conclusions about individual component effects to be drawn based on nonexperimental data, or data from an ERCT, with increased confidence.

Decision making to select the optimized intervention based on the results of an ORCT is very much an open area. As of the time of writing, DAIVE has been applied only to results obtained using a factorial or fractional factorial ORCT; extensions are needed to the SMART, MRT, and other ORCTs. Determining the best use of the uncertainty estimates that are readily obtainable with a Bayesian approach is another priority. An intriguing possibility for future work involves using DAIVE to optimize based on two or more very different types of outcomes, such as a desirable health outcome and an undesirable side effect, or a health outcome and an implementation outcome.

There are also many open questions regarding how to identify who should be engaged in optimization of an intervention, how to incorporate their preferences into the decision-making process, how to elicit value functions and weights so that they accurately reflect these preferences, and who ultimately selects the optimized intervention.

The concept of intervention *EASE* could be extended to include equitability, where equitability is defined as “the extent to which the health benefits provided by an intervention are distributed evenly, such that all participants have a fair and just opportunity to achieve the desired outcome of the intervention” (Strayhorn et al. 2024a, p. 2). This could enable decision makers to determine, for example, how to design an intervention to reduce health disparities, or what level of resources is needed to maintain effectiveness while achieving equitability.

The barriers that affect implementation may vary across settings. Assuming the results of a particular ORCT can be generalized across settings, it is possible to use its results to arrive at

optimized interventions suitable for different settings by applying different optimization objectives as appropriate (Guastaferro & Collins 2021). For example, if one setting can afford to spend \$300 per participant and another can afford only \$200, two different optimized interventions can be produced. To our knowledge, this has not yet been attempted.

MOST is currently framed as a sequence of discrete phases, but a more dynamic and flexible approach based on a Bayesian perspective on data collection and analysis (Vanness et al. 2023) may be more efficient. The phases of MOST would then be a seamless sequence of steps with more potential for back-and-forth and with the flow among them determined by critical decisions at multiple points. For example, data would be collected for an ORCT only while the information had sufficient value (e.g., until the costs of collecting additional information exceeded the benefits of reducing uncertainty), at which point the immediate next step would be either (*a*) proceeding to an ERCT, (*b*) returning to the preparation phase, or (*c*) if the results of the ORCT were sufficiently strong, implementing the intervention without conducting an ERCT. Such an approach has the potential to speed up the optimization process and increase the efficiency of its use of research resources. Arriving at this more flexible version of MOST will require considerable methodological work as well as development of software to make the framework practical for application by intervention and implementation scientists.

9. CONCLUSIONS

We have proposed that the field of clinical psychology can make considerable progress toward its goals by shifting to a new research paradigm: intervention optimization via MOST. We hope that this article has stimulated readers to consider how MOST can be useful in their research. Readers who wish to keep abreast of developments in this rapidly developing field may be interested in the websites of the Data Science for Dynamic Intervention Decision-Making Center (d³c) at the University of Michigan (<https://d3c.isr.umich.edu>) and the Center for Advancement and Dissemination of Intervention Optimization (cadio) at New York University (<https://cadio.org>).

We predict that as more scholars in the area of clinical psychology adopt the MOST paradigm, and as this paradigm develops, a substantial and coherent body of evidence will accumulate about what intervention approaches work, for whom, and when; interventions will continually become more effective and more readily implementable; and clinical psychology will achieve an increasingly positive societal impact.

SUMMARY POINTS

1. To achieve the goals of psychological intervention science, a new research paradigm is needed: intervention optimization via the multiphase optimization strategy (MOST).
2. In MOST, optimization is defined as the process of empirically selecting components and component levels to achieve intervention *EASE*: a strategic balance of intervention effectiveness, affordability, scalability, and efficiency.
3. Evaluation randomized controlled trials (ERCTs) are experiments designed to be efficient for estimating the effect of an intervention as a package.
4. Optimization randomized controlled trials (ORCTs) are experiments designed to be efficient for assessing the performance of intervention components.
5. There are numerous ORCT designs to choose from, and more are being developed.

6. The logical underpinnings of ORCTs are different from those of ERCTs; for instance, they are powered differently.
7. MOST can be used to optimize fixed or adaptive interventions or interventions that have both fixed and adaptive aspects.
8. Decision making—based on the results of an ORCT and potentially other information—for arriving at the set of components and component levels that will make up the optimized intervention is a critical part of MOST.
9. MOST can be used in implementation science in two ways: to develop interventions that are both effective and implementable and to optimize implementation strategies.
10. Use of the MOST paradigm will enable intervention science to accumulate a coherent base of knowledge, leverage this knowledge to develop high-impact interventions, and continue to increase the positive impact of these interventions.

FUTURE ISSUES

1. Develop additional ORCT designs.
2. Extend optimization trials beyond randomized experiments.
3. Extend decision-making methods and determine how best to incorporate input from interested parties.
4. Extend intervention *EASE* to include equitability as one of the criteria in achieving a strategic balance.
5. Extend MOST to situations in which resources, constraints, and barriers vary across settings.
6. Develop a more dynamic and flexible version of MOST.

DISCLOSURE STATEMENT

L.M.C. and D.J.V. have received compensation for consultation in the areas of intervention optimization and economic evaluation analysis. The other authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

ACKNOWLEDGMENTS

Preparation of this article was partially supported by the National Institute on Drug Abuse of the National Institutes of Health under awards R01DA039901, P50DA054039, P30DA011041, and P50DA039838.

LITERATURE CITED

- Aarons GA, Hurlburt M, Horwitz SM. 2011. Advancing a conceptual model of evidence-based practice implementation in public service sectors. *Adm. Policy Ment. Health* 38(1):4–23

- Almirall D, Chronis-Tuscano A. 2016. Adaptive interventions in child and adolescent mental health. *J. Clin. Child Adolesc. Psychol.* 45(4):383–95
- Almirall D, Nahum-Shani I, Wang L, Kasari C. 2018. Experimental designs for research on adaptive interventions: singly and sequentially randomized trials. See Collins & Kugler 2018, pp. 89–120
- Baker R, Camosso-Stefanovic J, Gillies C, Shaw EJ, Cheater F, et al. 2010. Tailored interventions to overcome identified barriers to change: effects on professional practice and health care outcomes. *Cochrane Database Syst. Rev.* 2010(3):CD005470
- Broder-Fingert S, Kuhn J, Sheldrick RC, Chu A, Fortuna L, et al. 2019. Using the Multiphase Optimization Strategy (MOST) framework to test intervention delivery strategies: a study protocol. *Trials* 20:728
- Caldwell LL, Smith EA, Collins LM, Graham JW, Lai M, et al. 2012. Translational research in South Africa: evaluating implementation quality using a factorial design. *Child Youth Care Forum* 41(2):119–36
- Cohen J. 1988. *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum. 2nd ed.
- Collins LM. 2018. *Optimization of Behavioral, Biobehavioral, and Biomedical Interventions: The Multiphase Optimization Strategy (MOST)*. New York: Springer
- Collins LM, Baker TB, Mermelstein RJ, Piper ME, Jorenby DE, et al. 2011. The multiphase optimization strategy for engineering effective tobacco use interventions. *Ann. Behav. Med.* 41(2):208–26
- Collins LM, Chakraborty B, Murphy SA, Strecher V. 2009. Comparison of a phased experimental approach and a single randomized clinical trial for developing multicomponent behavioral interventions. *Clin. Trials* 6(1):5–15
- Collins LM, Kugler KC, eds. 2018. *Optimization of Behavioral, Biobehavioral, and Biomedical Interventions: Advanced Topics*. Cham, Switz.: Springer
- Collins LM, Murphy SA, Bierman KL. 2004. A conceptual framework for adaptive preventive interventions. *Prev. Sci.* 5(3):185–96
- Collins LM, Strayhorn JC, Vanness DJ. 2021. One view of the next decade of research on behavioral and biobehavioral approaches to cancer prevention and control: intervention optimization. *Transl. Behav. Med.* 11(11):1998–2008
- Collins LM, Trail JB, Kugler KC, Baker TB, Piper ME, Mermelstein RJ. 2014. Evaluating individual intervention components: making decisions based on the results of a factorial screening experiment. *Transl. Behav. Med.* 4(3):238–51
- Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. 2012. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. *Med. Care* 50(3):217–26
- Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. 2009. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implement. Sci.* 4:50
- Damschroder LJ, Reardon CM, Widerquist MAO, Lowery J. 2022. The updated consolidated framework for implementation research based on user feedback. *Implement. Sci.* 17:75
- Downs DS, Savage JS, Rivera DE, Pauley AM, Leonard KS, et al. 2021. Adaptive, behavioral intervention impact on weight gain, physical activity, energy intake, and motivational determinants: results of a feasibility trial in pregnant women with overweight/obesity. *J. Behav. Med.* 44(5):605–21
- Dziak JJ, Almirall D, Dempsey W, Stanger C, Nahum-Shani I. 2024. SMART binary: new sample size planning resources for SMART studies with binary outcome measurements. *Multivar. Behav. Res.* 59(1):1–16
- Dziak JJ, Nahum-Shani I, Collins LM. 2012. Multilevel factorial experiments for developing behavioral interventions: power, sample size, and resource considerations. *Psychol. Methods* 17(2):153–75
- Fernandez ME, Schlechter CR, Del Fiol G, Gibson B, Kawamoto K, et al. 2020. QuitSMART Utah: an implementation study protocol for a cluster-randomized, multi-level sequential multiple assignment randomized trial to increase reach and impact of tobacco cessation treatment in community health centers. *Implement. Sci.* 15:9
- Fisher RA. 1946. *Statistical Methods for Research Workers*. London: Oliver and Boyd. 10th ed.
- Gallis JA, Bennett GG, Steinberg DM, Askew S, Turner EL. 2019. Randomization procedures for multi-component behavioral intervention factorial trials in the multiphase optimization strategy framework: challenges and recommendations. *Transl. Behav. Med.* 9(6):1047–56

- Guastaferro K, Collins LM. 2021. Optimization methods and implementation science: an opportunity for behavioral and biobehavioral interventions. *Implement. Res. Pract.* 2:26334895211054364
- Guo P, Rivera DE, Dong Y, Deshpande S, Savage JS, et al. 2022. Optimizing behavioral interventions to regulate gestational weight gain with sequential decision policies using hybrid model predictive control. *Comput. Chem. Eng.* 160:107721
- Gwadz MV, Collins LM, Cleland CM, Leonard NR, Wilton L, et al. 2017. Using the multiphase optimization strategy (MOST) to optimize an HIV care continuum intervention for vulnerable populations: a study protocol. *BMC Public Health* 17:383
- Kazdin AE. 2000. *Psychotherapy for Children and Adolescents: Directions for Research and Practice*. Oxford, UK/New York: Oxford Univ. Press
- Kidwell KM, Almirall D. 2023. Sequential, multiple assignment, randomized trial designs. *JAMA* 329(4):336–37
- Kidwell KM, Seewald NJ, Tran Q, Kasari C, Almirall D. 2018. Design and analysis considerations for comparing dynamic treatment regimens with binary outcomes from sequential multiple assignment randomized trials. *J. Appl. Stat.* 45:1628–51
- Kilbourne AM, Almirall D, Eisenberg D, Waxmonsky J, Goodrich DE, et al. 2014. Protocol: adaptive implementation of effective programs trial (ADEPT): cluster randomized SMART trial comparing a standard versus enhanced implementation strategy to improve outcomes of a mood disorders program. *Implement. Sci.* 9:132
- Kugler KC, Wyrick DL, Tanner AE, Milroy JJ, Chambers B, et al. 2018. Using the multiphase optimization strategy (MOST) to develop an optimized online STI preventive intervention aimed at college students: description of conceptual model and iterative approach to optimization. See Collins & Kugler 2018, pp. 1–21
- Landoll RR, Vargas SE, Samardzic KB, Clark MF, Guastaferro K. 2022. The preparation phase in the multiphase optimization strategy (MOST): a systematic review and introduction of a reporting checklist. *Transl. Behav. Med.* 12(2):291–303
- Leeman J, Birken SA, Powell BJ, Rohweder C, Shea CM. 2017. Beyond “implementation strategies”: classifying the full range of strategies used in implementation science and practice. *Implement. Sci.* 12:125
- Liao P, Klasnja P, Tewari A, Murphy SA. 2016. Sample size calculations for micro-randomized trials in mHealth. *Stat. Med.* 35(12):1944–71
- Marques MM, Guastaferro K. 2022. The development of effective and tailored digital behavior change interventions: an introduction to the multiphase optimization strategy (MOST). *Eur. Health Psychol.* 22(4):793–800
- McClure JB, Derry H, Riggs KR, Westbrook EW, St John J, et al. 2012. Questions about quitting (Q2): design and methods of a multiphase optimization strategy (MOST) randomized screening experiment for an online, motivational smoking cessation intervention. *Contemp. Clin. Trials* 33(5):1094–102
- Michie S, Atkins L, West R. 2014. *The Behaviour Change Wheel Book: A Guide to Designing Interventions*. London: Silverback
- Montgomery RV, Kwak J, Kosloski K, O’Connell Valuch K. 2011. Effects of the TCARE® intervention on caregiver burden and depressive symptoms: preliminary findings from a randomized controlled study. *J. Gerontol. B* 66(5):640–47
- Murphy SA. 2005. An experimental design for the development of adaptive treatment strategies. *Stat. Med.* 24(10):1455–81
- Murray DM. 1998. *Design and Analysis of Group-Randomized Trials*. Oxford, UK: Oxford Univ. Press
- Nahum-Shani I, Dziak JJ. 2018. Multilevel factorial designs in intervention development. See Collins & Kugler 2018, pp. 47–87
- Nahum-Shani I, Dziak JJ, Collins LM. 2018a. Multilevel factorial designs with experiment-induced clustering. *Psychol. Methods* 23(3):458–79
- Nahum-Shani I, Dziak JJ, Venera H, Pfammatter AF, Spring B, Dempsey W. 2023. Design of experiments with sequential randomizations on multiple timescales: the hybrid experimental design. *Behav. Res. Methods*. <https://doi.org/10.3758/s13428-023-02119-z>
- Nahum-Shani I, Dziak JJ, Walton MA, Dempsey W. 2022a. Hybrid experimental designs for intervention development: what, why, and how. *Adv. Methods Pract. Psychol. Sci.* 5(3):25152459221114280

- Nahum-Shani I, Dziak JJ, Wetter DW. 2022b. MCMTC: a pragmatic framework for selecting an experimental design to inform the development of digital interventions. *Front. Digit. Health* 4:798025
- Nahum-Shani I, Potter LN, Lam CY, Yap J, Moreno A, et al. 2021. The mobile assistance for regulating smoking (MARS) micro-randomized trial design protocol. *Contemp. Clin. Trials* 110:106513
- Nahum-Shani I, Shaw SD, Carpenter SM, Murphy SA, Yoon C. 2022c. Engagement in digital interventions. *Am. Psychol.* 77(7):836–52
- Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, et al. 2018b. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. *Ann. Behav. Med.* 52(6):446–62
- NIH (Nat'l. Inst. Health.). 2017. NIH's definition of a clinical trial. *National Institutes of Health*. <https://grants.nih.gov/policy/clinical-trials/definition.htm>
- O'Hara KL, Knowles LM, Guastaferro K, Lyon AR. 2022. Human-centered design methods to achieve preparation phase goals in the multiphase optimization strategy framework. *Implement. Res. Pract.* 3:26334895221131052
- Papa A, Follette WC. 2015. Dismantling studies of psychotherapy. In *The Encyclopedia of Clinical Psychology*, ed. RL Cautin, SO Lilienfeld. Chichester, UK: Wiley. <https://doi.org/10.1002/9781118625392.wbcep523>
- Perry CK, Damschroder LJ, Hemler JR, Woodson TT, Ono SS, Cohen DJ. 2019. Specifying and comparing implementation strategies across seven large implementation interventions: a practical application of theory. *Implement. Sci.* 14:32
- Piper ME, Schlam TR, Cook JW, Smith SS, Bolt DM, et al. 2017. Toward precision smoking cessation treatment I: moderator results from a factorial experiment. *Drug Alcohol Depend.* 171:59–65
- Powell BJ, Beidas RS, Lewis CC, Aarons GA, McMillen JC, et al. 2017. Methods to improve the selection and tailoring of implementation strategies. *J. Behav. Health Serv. Res.* 44(2):177–94
- Powell BJ, McMillen JC, Proctor EK, Carpenter CR, Griffey RT, et al. 2012. A compilation of strategies for implementing clinical innovations in health and mental health. *Med. Care Res. Rev.* 69(2):123–57
- Powell BJ, Waltz TJ, Chinman MJ, Damschroder LJ, Smith JL, et al. 2015. A refined compilation of implementation strategies: results from the expert recommendations for implementing change (ERIC) project. *Implement. Sci.* 10:21
- Proctor EK, Powell BJ, McMillen JC. 2013. Implementation strategies: recommendations for specifying and reporting. *Implement. Sci.* 8:139
- Qian T, Walton AE, Collins LM, Klasnja P, Lanza ST, et al. 2022. The microrandomized trial for developing digital interventions: experimental design and data analysis considerations. *Psychol. Methods* 27:874–94
- Rivera DE, Hekler EB, Savage JS, Downs DS. 2018. Intensively adaptive interventions using control systems engineering: two illustrative examples. See Collins & Kugler 2018, pp. 121–73
- Shiffman S, Stone AA, Hufford MR. 2008. Ecological momentary assessment. *Annu. Rev. Clin. Psychol.* 4:1–32
- Smith SN, Almirall D, Choi SY, Koschmann E, Rusch A, et al. 2022. Primary aim results of a clustered SMART for developing a school-level, adaptive implementation strategy to support CBT delivery at high schools in Michigan. *Implement. Sci.* 17:42
- Smith SN, Almirall D, Prenovost K, Liebrecht C, Kyle J, et al. 2019. Change in patient outcomes after augmenting a low-level implementation strategy in community practices that are slow to adopt a collaborative chronic care model: a cluster randomized implementation trial. *Med. Care* 57(7):503–11
- Stirman SW, Beidas RS. 2020. Expanding the reach of psychological science through implementation science: introduction to the special issue. *Am. Psychol.* 75(8):1033–37
- Strayhorn JC, Collins LM, Brick TR, Marchese SH, Pfammatter AF, et al. 2022. Using factorial mediation analysis to better understand the effects of interventions. *Transl. Behav. Med.* 12(1):ibab137
- Strayhorn JC, Collins LM, Vanness DJ. 2023. A posterior expected value approach to decision-making in the multiphase optimization strategy for intervention science. *Psychol. Methods*. <https://psycnet.apa.org/doi/10.1037/met0000569>
- Strayhorn JC, Vanness DJ, Collins LM. 2024a. Optimizing interventions for equitability: some initial ideas. *Prev. Sci.* <https://doi.org/10.1007/s11121-024-01644-3>

- Strayhorn JC, Cleland CM, Vanness DJ, Wilton L, Gwadz M, Collins LM. 2024b. Using decision analysis for intervention value efficiency to select optimized interventions in the multiphase optimization strategy. *Health Psychol.* 43(2):89–100
- Tanner AE, Guastaferro KM, Rulison KL, Wyrick DL, Milroy JJ, et al. 2021. A hybrid evaluation-optimization trial to evaluate an intervention targeting the intersection of alcohol and sex in college students and simultaneously test an additional component aimed at preventing sexual violence. *Ann. Behav. Med.* 55(12):1184–87
- Vanness D, Collins LM, Strayhorn JC. 2023. P25 dynamic intervention optimization (DIO): a Bayesian adaptive framework for multicomponent intervention optimization using expected net benefit of sampling. *Value Health* 26(6 Suppl.):S6
- Voils CI, King HA, Maciejewski ML, Allen KD, Yancy WS, Shaffer JA. 2014. Approaches for informing optimal dose of behavioral interventions. *Ann. Behav. Med.* 48(3):392–401
- Walton A, Nahum-Shani I, Crosby L, Klasnja P, Murphy S. 2018. Optimizing digital integrated care via micro-randomized trials. *Clin. Pharmacol. Ther.* 104(1):53–58
- Watkins E, Newbold A, Tester-Jones M, Collins LM, Mostazir M. 2023. Investigation of active ingredients within internet-delivered cognitive behavioral therapy for depression: a randomized optimization trial. *JAMA Psychiatry* 80(9):942–51
- Watkins E, Newbold A, Tester-Jones M, Javaid M, Cadman J, et al. 2016. Implementing multifactorial psychotherapy research in online virtual environments (IMPROVE-2): study protocol for a phase III trial of the MOST randomized component selection method for internet cognitive-behavioural therapy for depression. *BMC Psychiatry* 16:345
- Weiner BJ, Lewis MA, Clauer SB, Stitzenberg KB. 2012. In search of synergy: strategies for combining interventions at multiple levels. *J. Natl. Cancer Inst. Monogr.* 2012(44):34–41
- Wu CFJ, Hamada MS. 2011. *Experiments: Planning, Analysis, and Optimization*. New York: Wiley
- Wyrick DL, Tanner AE, Milroy JJ, Guastaferro K, Bhandari S, et al. 2020. itMatters: optimization of an online intervention to prevent sexually transmitted infections in college students. *J. Am. Coll. Health* 70(4):1212–22
- Yap J, Dziak JJ, Maiti R, Lynch K, McKay JR, et al. 2023. Sample size estimation for comparing dynamic treatment regimens in a SMART: a Monte Carlo-based approach and case study with longitudinal overdispersed count outcomes. *Stat. Methods Med. Res.* 32(7):1267–83

