An Overview of Multiphase Optimization Strategy (MOST) for Behavioral Intervention Development

Liang Li, PhD

Biostatistics @ MD Anderson

Behavioral Science Workshop, October 13, 2014

The Multiphase Optimization Strategy (MOST)

- An increasingly popular research strategy to develop behavioral interventions (not a study design)
- Proposed by Dr. Linda Collins @ Penn State
- NIH is funding proposals specifically for the development and application of MOST strategy
- It involves some statistical techniques, mainly the factorial and fractional factorial design methodology

Multicomponent Behavioral Interventions

Behavioral interventions usually involve multiple components, for example:

- ✓ **Program components (the contents of the message)**
  1. **Outcome expectation messages: address an individual's expectations of outcomes related to quitting; tailored to the individual or not**
  2. **Efficacy expectation messages: address relevant barriers to quitting, high-risk situations, existing skills for successful quitting, and attributions for previous failed attempts at quitting; tailored to the individual or not**
  3. **Message framing: motivates the decision to quit. Positive (e.g., quitting results in more energy) or negative terms (e.g., not quitting increases risk of cancer)**
  4. **Former smokers' testimonials: include or not include**
- ✓ **Delivery components (method of delivery)**
  5. **Exposure schedule: the message is delivered in one large message or several smaller ones (if so, how many)**
  6. **Source of message: health maintenance organization (HMO) or primary care physician (PCP)**

Multicomponent Behavioral Interventions

- Another example for energy balance interventions
  1. Weekly telephone coaching vs. e-mail coaching
  2. Text messages vs. no messages
  3. Social networking vs. no social networking
  4. Daily self-monitoring vs. weekly self monitoring

- Research question: How to build an optimized intervention that involves many components?
  - ✓ Which components to use?
  - ✓ Set the level of components (e.g., dosage)
  - ✓ Study the relative effects of individual components
  - ✓ Weed out inactive components
  - ✓ Any interactions (e.g., synergy) among components
  - ✓ Incorporate cost, compliance, and logistics considerations

Traditional Approach: the Treatment Package

1. Construct an intervention a priori by packaging multiple components together
2. Run a standard two-arm randomized controlled trial (RCT) of the proposed intervention vs. an old one or usual care
3. After RCT, post-hoc analyses are used to explain how the intervention worked (component effects, interactions, etc) or why it did not work
4. Refine the intervention and construct a second generation of it
5. Run a new RCT to evaluate the new intervention.
6. And so on … …

Problem:

- ✓ Post hoc analyses may be subject to bias because they are not based on random assignment
- ✓ May not enable isolation of the individual component effects (if good, not sure which components contribute; if bad, not sure which components are inactive)
- ✓ Gain no knowledge about interaction
- ✓ A very slow process that may involve multiple cycles

Some Other Approaches

- **Individual experiments:**
  - ✓ **Run K independent randomized trials for each of the K candidate components**
- **Single factor experiments: K+1 group randomized trial**

	A	B	C
0	N	N	N
1	Y	N	N
2	N	Y	N
3	N	N	Y

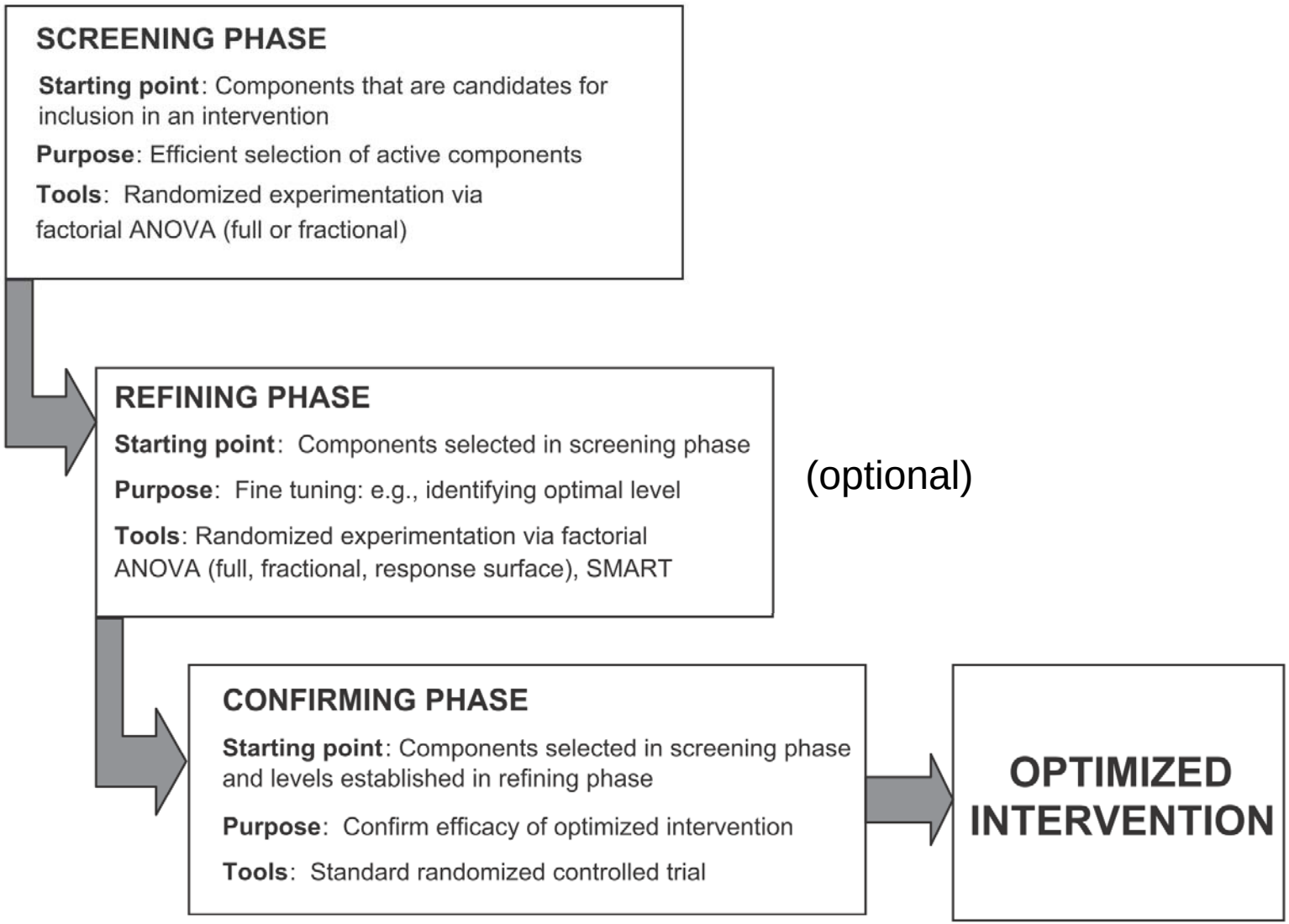
- **Problem:**
  - ✓ **It is unclear how the various components work together**
  - ✓ **Often need large sample size**

The Engineering Perspective

- Manufacture the truck leaf string:
  - ✓ furnace temperature (low, high)
  - ✓ heating time (short, long)
  - ✓ transfer time on conveyer belt (short, long)
  - ✓ hold down time in high pressure (short, long)
  - ✓ quench oil temperature (low, high)

- The key gradients of MOST (factorial design) was developed from engineering applications

- Optimization: the process of finding the best possible solution to a problem … subject to given constraints


SCREENING PHASE
Starting point: Components that are candidates for inclusion in an intervention
Purpose: Efficient selection of active components
Tools: Randomized experimentation via factorial ANOVA (full or fractional)
REFINING PHASE
Starting point: Components selected in screening phase
Purpose: Fine tuning: e.g., identifying optimal level
Tools: Randomized experimentation via factorial ANOVA (full, fractional, response surface), SMART
(optional)
CONFIRMING PHASE
Starting point: Components selected in screening phase and levels established in refining phase
Purpose: Confirm efficacy of optimized intervention
Tools: Standard randomized controlled trial
OPTIMIZED INTERVENTION

The MOST Strategy

- **Features:**
  - ✓ **Indicates which components are active, which are redundant, and which ones work together in synergy**
  - ✓ **Ensures incremental improvement, and therefore is the fastest way to the best intervention in the long run**
  - ✓ **Three-phase approach. Confirmatory RCT in the last phase well justified.**
  - ✓ **Readily incorporates costs/constraints of any kind**

- **Note:**
  - ✓ **There is no "MOST design"; it is a research strategy with factorial design at its core**
  - ✓ **Not for causal effect of individual components (not a RCT), but as a tool for building a behavioral intervention package**
  - ✓ **Exploratory at the beginning, confirmatory at the end**

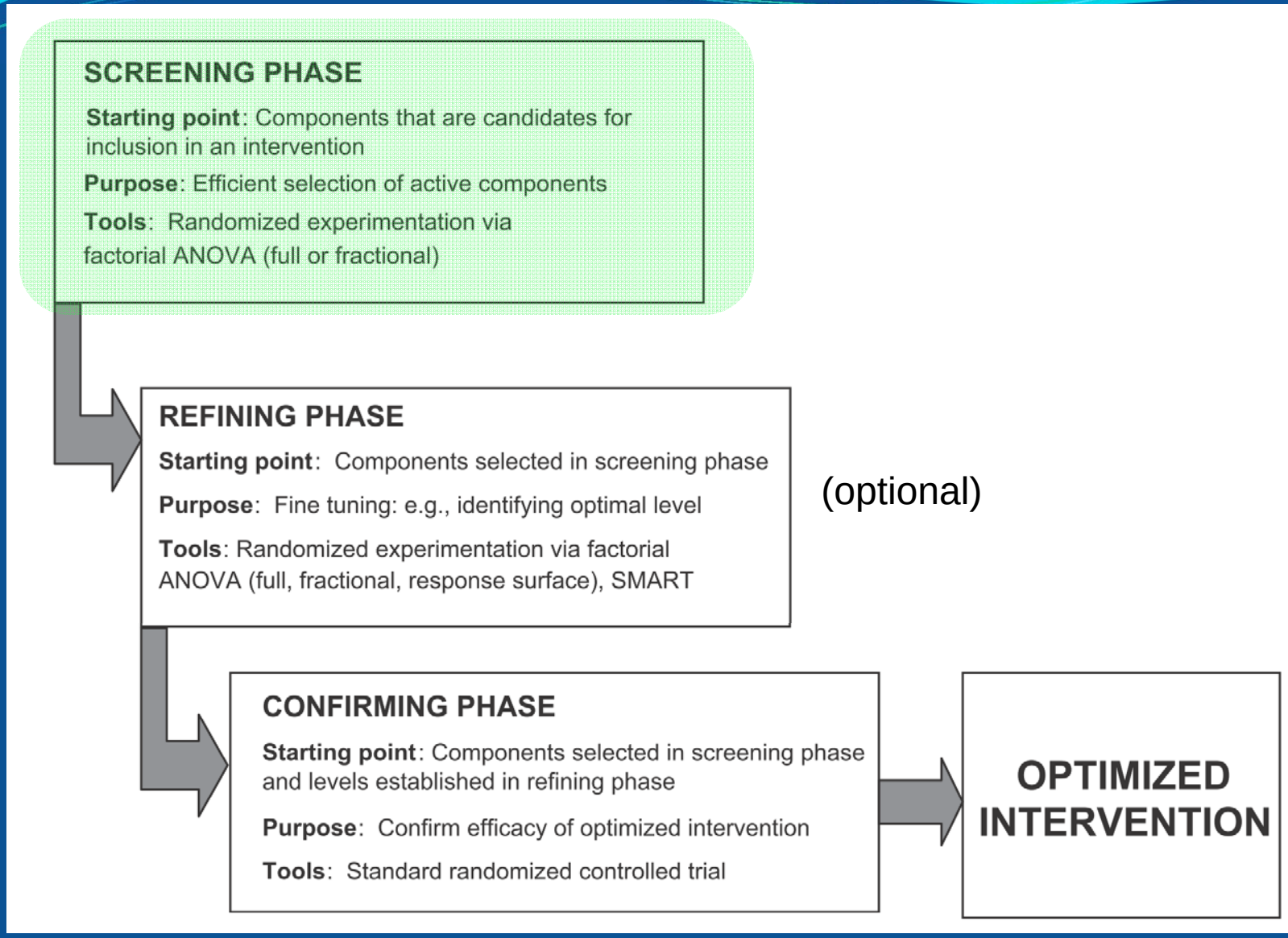
The Two Principles of MOST

- **Resource management principle:**
  - ✓ **Huge (e.g., 64-arm RCT for 6 components) would be definitive, but not feasible**
  - ✓ **Given the resource constraints, what is the most efficient way to achieve the goal**

- **Continuous optimization principle: optimization is a "cyclic process"**

> the best time to design an experiment is after it is finished, the converse … is that the worst time is … the beginning, when least is known. If the entire experiment was designed at the outset, the following would have to be assumed as known: (1) which variables were the most important, (2) over what ranges the variables should be studied. … The experimenter is least able to answer such questions at the outset of an investigation but gradually becomes more able to do so as a program evolves. (p. 303)

George Box et al, 1978 "Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building"


SCREENING PHASE
Starting point: Components that are candidates for inclusion in an intervention
Purpose: Efficient selection of active components
Tools: Randomized experimentation via factorial ANOVA (full or fractional)
REFINING PHASE
Starting point: Components selected in screening phase
Purpose: Fine tuning: e.g., identifying optimal level
Tools: Randomized experimentation via factorial ANOVA (full, fractional, response surface), SMART
(optional)
CONFIRMING PHASE
Starting point: Components selected in screening phase and levels established in refining phase
Purpose: Confirm efficacy of optimized intervention
Tools: Standard randomized controlled trial
OPTIMIZED INTERVENTION

Screening Phase

- **Objective: identify the most promising components and level/dosage**
- **NOT to compare each combination to a control or against each other (not a confirmatory RCT)**
- **Optimization criteria: effectiveness, cost/time and other resource constraints**
  - ✓ **Setting all components to the highest level may be a good idea**

Factorial Design

- The screening phase uses the factorial design (R.A. Fisher, 1926)
- Factorial design: subjects randomized to $m=2^K$ conditions in order to study the main effect and interactions of K intervention components
- Each condition has N/m subjects (balanced assignment)

Condition	A	B	C
1	1	1	1
2	1	1	-1
3	1	-1	1
4	1	-1	-1
5	-1	1	1
6	-1	1	-1
7	-1	-1	1
8	-1	-1	-1

Why Using Factorial Design?

- It enables examination of the effects of individual components
- It requires SMALLER sample sizes than alternative designs (individual experiment, treatment package, single factor experiments)
  - ✓ See next page for an illustration
- But it usually requires more experiment conditions than we may be accustomed to (causing logistic difficulties)
- We do not compare the conditions (they are too many of them!); we estimate the main effect and some interactions of scientific interest
- The estimates are based on all the subjects (efficiently use subjects and reduce sample size)

Factorial Design (cont.)

- NOT to identify single best combination
  - ✓ Multiple combinations of intervention components may lead to similar results
  - ✓ To identify the single best combination, an enormous RCT is the only way and it is often impractical
  - ✓ Find a good combination, if not the best
  - ✓ MOST may identify the best combination in the long run

Individual experiments

condition	A
0	N
1	Y

condition	B
0	N
1	Y

condition	C
0	N
1	Y

Single factor experiments

	A	B	C
0	N	N	N
1	Y	N	N
2	N	Y	N
3	N	N	Y

Factorial design

Condition	A	B	C
1	1	1	1
2	1	1	-1
3	1	-1	1
4	1	-1	-1
5	-1	1	1
6	-1	1	-1
7	-1	-1	1
8	-1	-1	-1

Why are we interested primarily in main effects?

- Effect is likely to be robust if it is obtained as an average across many other factors --- R. A. Fisher

- If theory and prior research specifies and explains an interaction, it must always be dealt with. However,
  - ✓ we know little about interactions
  - ✓ most theories and models are silent on this topic

- Where theory/prior research do not specify whether or not there is an interaction, we rely on these principles:
  - ✓ **Effect sparsity**: there are a lot of effects in a factorial experiment, most are not significant or important
  - ✓ **Hierarchical ordering**: those that are important are likely to be simpler effects, i.e., main effects first, then two-way interactions

Powering the factorial design

- Power for main effects: sample size requirements for a k-factor experiment about the same as for a t-test
- Power the experiment for the smaller effect size
- Adding a factor generally does not increase sample size requirements, unless that factor is expected to have a smaller effect size
- Power the study for the smallest effect size that you would accept for inclusion in the intervention
- Usually not powered for interactions
  - ✓ Little is known about interactions
  - ✓ Effect sizes are probably much smaller than main effects

Powering the factorial design: Example

- **Three intervention components: A, B, C**

Design	Design	n	# conditions	interactions
Individual experiment	A vs. NULL B vs. NULL C vs. NULL	168	6	None can be estimated
Single factor experiment	A vs. B vs. C vs. NULL	112	4	None can be estimated
Complete factorial	Factorial (A, B, C)	56	8	All can be estimated

Some Misconceptions

- **Misconception 1: factorial experimental designs require larger numbers of subjects than available alternative designs**
  - ✓ **Reality: when used to address suitable research questions, balanced factorial designs often require may FEWER subjects than alternative designs**

- **Misconception 2: if you want to add a factor to a balanced factorial design, you will need to increase the sample size dramatically**
  - ✓ **Reality: If the effect size of the added factor is no smaller than the factors already in the experiment, power will be about the same**

- **Misconception 3: the primary motivation for conducting a factorial design is always to test for interactions**
  - ✓ **Reality: even if there is no interaction, you can still conduct a factorial design to make economical use of subjects**

Some Reasons for Not Using a Factorial Design

- Intervention composed of many components with tiny effects, overall effect is cumulative
  - ✓ May be difficult to power the study for tiny effects
  - ✓ May need to sort the components into bundles and study bundles
- Factorial design requires more experimental conditions, which may cause logistics difficulties
  - ✓ May reduce it by using fractional factorial design (FFD)

Fractional Factorial Design (FFD)

- Well established statistical theory & software, applied to behavioral science
- Factorial designs in which only a subset of experimental conditions are run
- FFD requires at most ½ of the cells of a complete factorial design (CFD), often many fewer
- Example: K factors, CFD has 2^K conditions, FFD may have 2^{K-1} or 2^{K-2} conditions

About FFD

- Why run just a subset of conditions?
  - ✓ Economy: K factors, CFD has 2^K conditions; $2^6 = 64$, $2^7 = 128$
  - ✓ Example: FFD may conduct a 2^7 experiment with only 16 conditions
- When you might consider a FFD?
  - ✓ 5 or more factors (FFD exists for 3 or 4 factors, but benefit is small and strong assumptions are needed)
  - ✓ You are primarily interested in main effects and low-order (2-way) interactions
  - ✓ Remaining effects and high order interactions are negligible:
    - ➢ Effect sparsity principle
    - ➢ Hierarchical ordering principle

Cond-ition	A	B	C	A	B	C	A*B	A*C	B*C	A*B*C
1	Off	Off	Off	-1	-1	-1	1	1	1	-1
2	Off	Off	On	-1	-1	1	1	-1	-1	1
3	Off	On	Off	-1	1	-1	-1	1	-1	1
4	Off	On	On	-1	1	1	-1	-1	1	-1
5	On	Off	Off	1	-1	-1	-1	-1	1	1
6	On	Off	On	1	-1	1	-1	1	-1	-1
7	On	On	Off	1	1	-1	1	-1	-1	-1
8	On	On	On	1	1	1	1	1	1	1

Cond-ition	A	B	C	A	B	C	A*B	A*C	B*C	A*B*C
2	Off	Off	On	-1	-1	1	1	-1	-1	1
3	Off	On	Off	-1	1	-1	-1	1	-1	1
5	On	Off	Off	1	-1	-1	-1	-1	1	1
8	On	On	On	1	1	1	1	1	1	1

Statistical Power of FFD

- **FFD and CFD have the same statistical power (using FFD does NOT reduce or increase sample size)**

- **Compared to the corresponding CFD, in a FFD:**
  - ✓ **Each condition will have more subjects than the CFD**
  - ✓ **But each effect estimate based on SAME number of subjects**

Design	Number of subjects needed for power > 0.9	Number of conditions	Interactions
CFD	512	$2^6 = 64$	All can be estimated
FFD	512	8-32 depending on design	Selected subset can be estimated

Notation for FFD

- Suppose 4 factors, each factor has 2 levels.
- CFD: 2^4 (16 conditions/cells)
- An FFD with 8 conditions is represented as 2^{4-1}
  - ✓ $2^{4-1}=2^3=8$
  - ✓ This notation tells you:
    - ➢ The number of conditions in the original CFD
    - ➢ The number of conditions in the FFD
    - ➢ The fraction by which FFD reduces the original: ½
    - ➢ The number of aliases of each estimable effect in FFD: 2

What is Aliasing?

Cond-ition	A	B	C	A	B	C	A*B	A*C	B*C	A*B*C
1	Off	Off	Off	-1	-1	-1	1	1	1	-1
2	Off	Off	On	-1	-1	1	1	-1	-1	1
3	Off	On	Off	-1	1	-1	-1	1	-1	1
4	Off	On	On	-1	1	1	-1	-1	1	-1
5	On	Off	Off	1	-1	-1	-1	-1	1	1
6	On	Off	On	1	-1	1	-1	1	-1	-1
7	On	On	Off	1	1	-1	1	-1	-1	-1
8	On	On	On	1	1	1	1	1	1	1

- **Estimate the effect of A: compare 2, 3, vs. 5, 8; A is aliased with B*C**
- **Estimate the effect of B: compare 2, 5, vs. 3, 8; B is aliased with A*C**
- **Estimate the effect of C: compare 2, 8, vs. 3, 5; C is aliased with A*B**

(A + B*C) equals A if B*C is negligible

Aliasing of Effects

- Consider a 2^4 factorial design
- 4 factors, 16 conditions/cells
- Effects estimated (TOTAL = 16 parameters)
  - ✓ 1 intercept
  - ✓ 4 main effects
  - ✓ 6 two-way interactions
  - ✓ 4 three-way interactions
  - ✓ 1 four-way interactions
- For both CFD and FFD, there are as many estimable effects as the number of conditions
  - ✓ FFD: reduce the number of conditions so that the effects of scientific interest are estimable but negligible effects are not estimated (resource management principle)

Aliasing of Effects

- Now consider a 2^{4-1} fractional factorial design
- 4 factors, 8 conditions/cells, 8 estimable effects
- The original 16 effects are combined into 8 estimable effects (aliased)
- In any FFD, it is known which effects are aliased with which
  - ✓ In a ½ FFD, each effect is aliased with 1 other effect ("bundles" of 2)
  - ✓ In a ¼ FFD, each effect is aliased with 3 other effect ("bundles" of 4)
  - ✓ And so on … …
  - ✓ We choose to bundle the effect of scientific interest (main effects & important interactions) with a few other negligible effects
- For 4 factors, there is only 1 FFD
- As the number of factors increases, there are many FFD for each CFD
  - ✓ Choose the one for the specific scientific study

Resolution of an FFD

- **FFDs are classified according to their resolution**
- **For a given CFD, there may be many FFDs with different resolutions**
- **Resolution is denoted by Roman numbers: III, IV, V, VI**

Resolution	Main effects are NOT aliased with	2-way interactions are NOT aliased with
III	Main effects	
IV	Main effects, 2-way	Main effects
V	Main effects, 2-way, 3-way	Main effects, 2-way
VI	Main effects, 2-way, 3-way, 4-way	Main effects, 2-way, 3-way

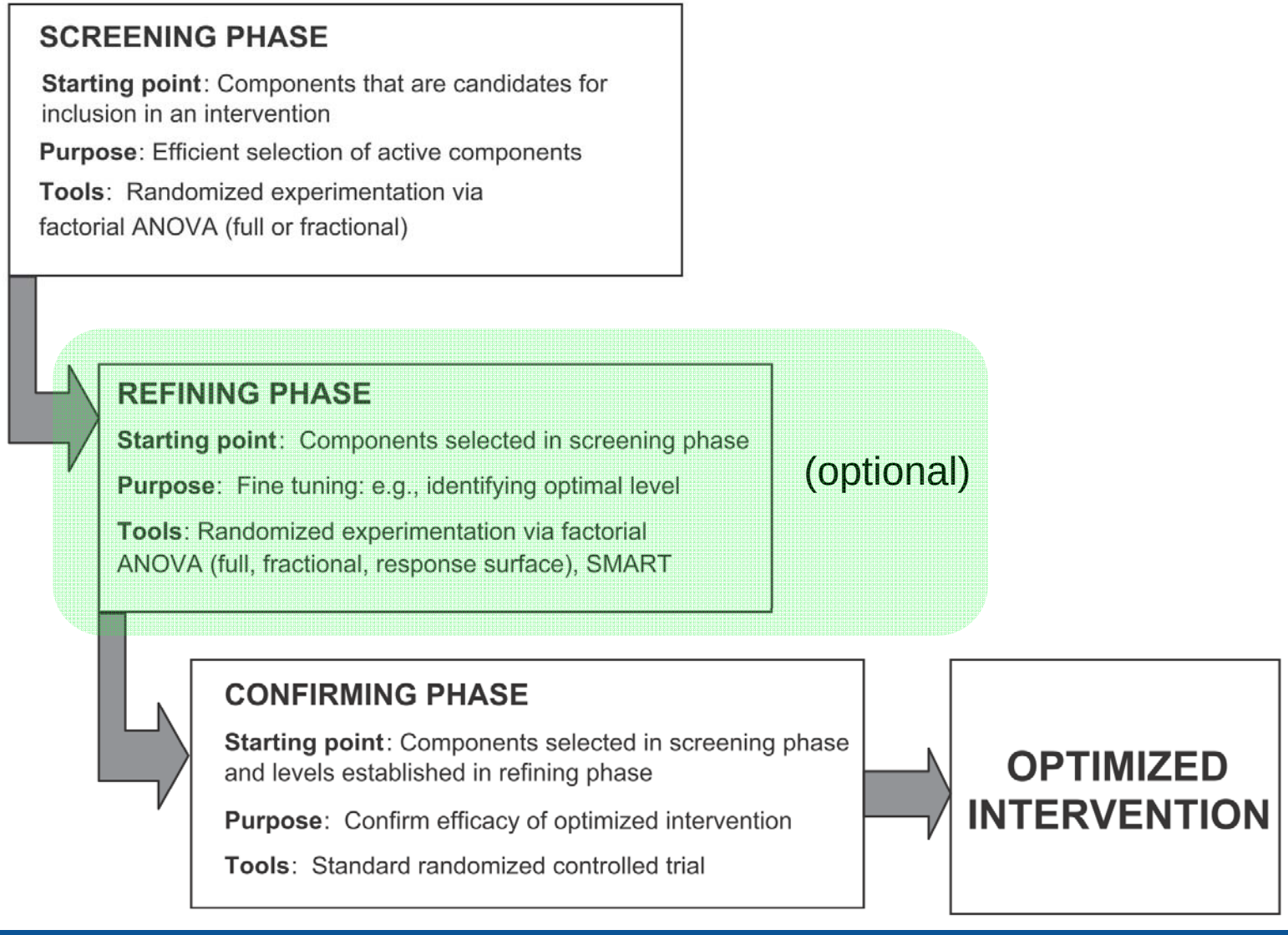
In a resolution R FFD, F-way interactions are aliased ONLY with R-F way interactions or higher order interactions. For example, Resolution V design: 2-ways aliased only with (5-2)=3-ways or higher

Resolution of an FFD: Example

- CFD would be 2^6=64 conditions
- We choose FFD with 2^{6-1}=32 conditions
- Fraction = half; each effect aliased with another effect
- Resolution VI:
  - ✓ Each main effect aliased with a 5-way interaction
  - ✓ Each 2-way interaction aliased with a 4-way interaction

How to Choose a FFD?

- Classify all effects of CFD into 3 categories
  - a) Effects of primary scientific interest: make them estimable
  - b) Effects expected to be 0 or negligible
  - c) Effects not of scientific interest but may be non-negligible
- Alias (a) and (c) with (b); Do not alias (a) with (c)
- More effects are designated negligible → FFD with fewer conditions
- No effect is negligible → CFD is the only choice; FFD does not exist
- Heuristic guiding principles:
  - ✓ Hierarchical ordering: priority be given to lower order effects
  - ✓ Effect sparsity (Pareto principle): number of non-negligible effects is a small fraction of the total number of effects (2^K)
- Higher resolution FFD is better than lower resolution ones, because they alias main effects and 2-way interactions with high order interactions


SCREENING PHASE
Starting point: Components that are candidates for inclusion in an intervention
Purpose: Efficient selection of active components
Tools: Randomized experimentation via factorial ANOVA (full or fractional)
REFINING PHASE
Starting point: Components selected in screening phase
Purpose: Fine tuning: e.g., identifying optimal level
Tools: Randomized experimentation via factorial ANOVA (full, fractional, response surface), SMART
(optional)
CONFIRMING PHASE
Starting point: Components selected in screening phase and levels established in refining phase
Purpose: Confirm efficacy of optimized intervention
Tools: Standard randomized controlled trial
OPTIMIZED INTERVENTION

Refining Phase

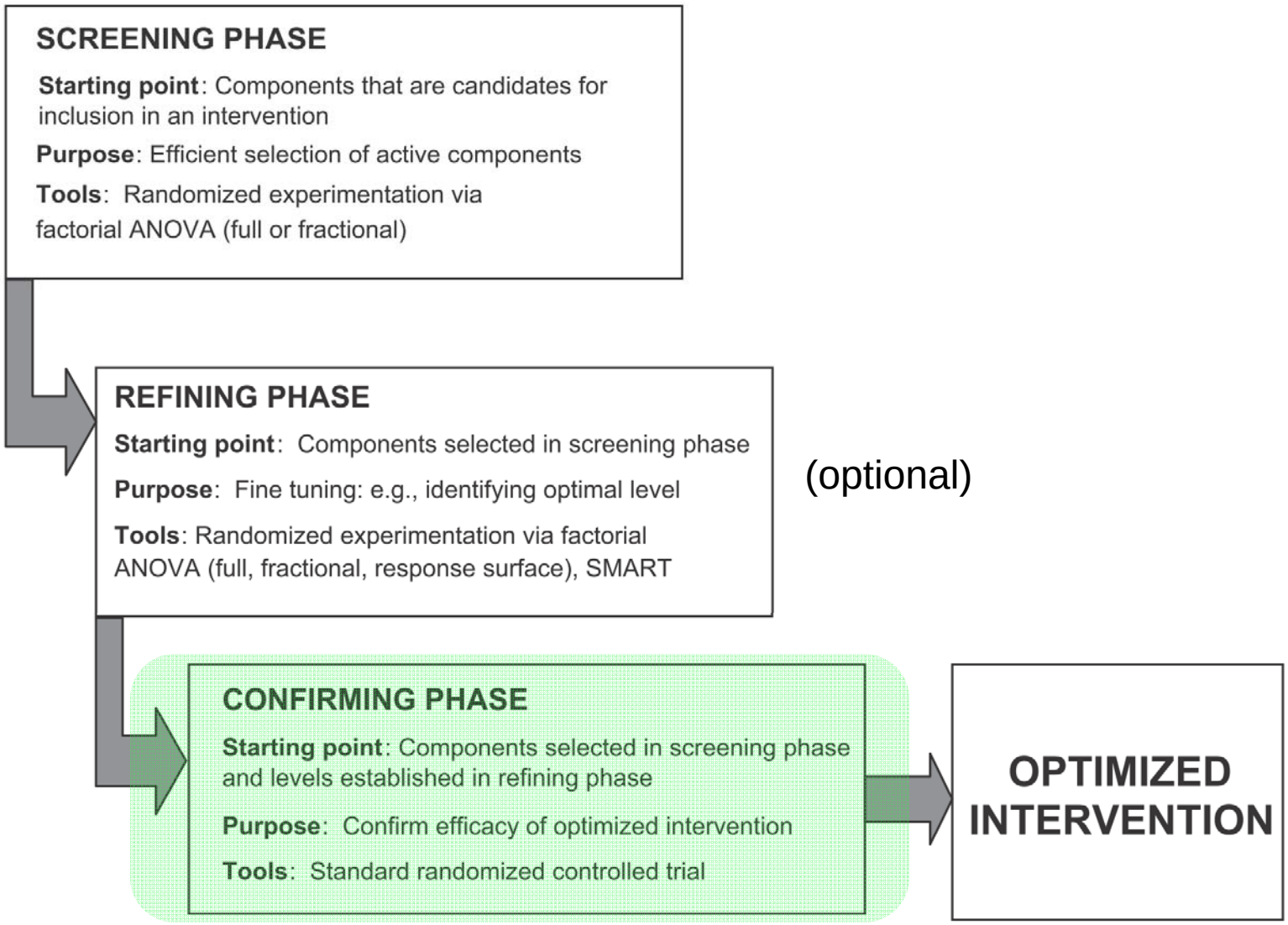
- Used for activities such as obtaining a sense of the best dosages to use, assessing whether key variables moderate important effects, and resolving any lingering questions resulting from the screening phase.

- It can also be used to verify the working assumptions in the screening phase, such as whether the aliased effects are truly negligible.

- Use factorial design or response surface design. Example: a 3 by 3 response surface design for the frequency of outcome and efficacy expectation messages (no message; weekly; daily)

Decision Rule for Screening/Refining Phases

- Random assignment is a cornerstone of all three phases of MOST. However, in the screening and refining phases, traditional hypothesis testing is not (formal hypothesis testing usually plays an important role in the confirming phase).
- In screening and refining phases, a Type II error (i.e., overlooking an active intervention component) is at least as serious as a Type I error (i.e., mistakenly concluding that an inactive component is active).
- We may use a higher Type I error rate than the conventional 0.05. Given a fixed sample size, this increases the statistical power.
- Another approach is to rank components by standardized effect size and select the most important ones rather than examining the statistical significance of each effect.
- May not select the “best” combination, but the “optimized” intervention


SCREENING PHASE
Starting point: Components that are candidates for inclusion in an intervention
Purpose: Efficient selection of active components
Tools: Randomized experimentation via factorial ANOVA (full or fractional)
REFINING PHASE
Starting point: Components selected in screening phase
Purpose: Fine tuning: e.g., identifying optimal level
Tools: Randomized experimentation via factorial ANOVA (full, fractional, response surface), SMART
(optional)
CONFIRMING PHASE
Starting point: Components selected in screening phase and levels established in refining phase
Purpose: Confirm efficacy of optimized intervention
Tools: Standard randomized controlled trial
OPTIMIZED INTERVENTION

Confirming Phase

- When the confirming phase is begun, the screening and refining phases have identified the important components and their appropriate levels or doses. Using this information, the researcher can construct an optimized prototype version of the program, made up of only components determined to be active, at doses determined to be most efficacious.

- A standard randomized two-arm trial to compare the proposed intervention with a control (such as the usual care).

- Type I error is strictly controlled. Statistical power must be adequate.

Summary

- MOST strategy includes three steps:
  - ✓ A screening phase, in which intervention components are efficiently identified for inclusion in an intervention or for rejection, based on their performance
  - ✓ A refining phase, in which the selected components are fine tuned and issues such as optimal levels of each component are investigated
  - ✓ A confirming phase, in which the optimized intervention, consisting of the selected components delivered at optimal levels, is evaluated in a standard randomized controlled trial
- At the core of MOST is the factorial design
- Software: SAS PROC FACTEX can be used to generate FFD
  - ✓ Specify a desired resolution
  - ✓ Specify which effects are of interest, negligible, or non-negligible
  - ✓ Specify constraints, including costs and maximum number of conditions
- More details at http://methodology.psu.edu/

PENNSTATE 1855

The Methodology Center
advancing methods, improving health

Free Software

SAS macro: Calculate power of factorial experiments, more

SAS macro: Chart relative cost of reduced factorial designs

All Center software

Latest News

Linda Collins honored by SBM
April 23, 2014

Using dynamical systems models to treat behavior
April 17, 2014

More MOST news

Recent Articles

An updated list of our most recent publications.

Read more

Optimizing Behavioral Interventions

Behavioral interventions for prevention and treatment are an important part of the fight against drug abuse and related conditions such as HIV/AIDS and Hepatitis C. They are also important in many other areas such as smoking cessation, weight management, and cancer. Among the challenges faced by scientists is how to optimize these interventions in order to achieve the greatest public health benefits.

The Multiphase Optimization Strategy (MOST)

Center researchers are developing the multiphase optimization strategy (MOST) to provide a framework for engineering efficacious and effective behavioral interventions. Conceptually rooted in engineering, MOST emphasizes efficiency and careful management of resources to move intervention science forward systematically. MOST can be used to guide the evaluation of research evidence, develop optimized interventions, and enhance Type I and Type II translation of research.

Researchers

Lead researcher: Linda Collins

Other researchers: Susan Murphy, Daniel Almirall, John Dziak, Inbal "Billie" Nahum-Shani, and Kari Kugler

Resources

- Recommended reading
- Decision making using data from a factorial experiment
- Including MOST in grant proposals
- NIH PAs calling for MOST
- Podcast: MOST in teen risk program
- Video: MOST overview lecture

Center Collaborations

This work began as a collaboration between Linda Collins and Susan Murphy. This led to