

**Comparison of a phased experimental approach and a single randomized clinical trial for
developing multicomponent behavioral interventions**

Linda M. Collins, Ph.D.

The Methodology Center and

Department of Human Development and Family Studies

The Pennsylvania State University, University Park, PA, USA

Bibhas Chakraborty, M.A.

Department of Statistics

Susan A. Murphy, Ph.D.

Department of Statistics and Institute for Social Research

Vijay Nair, Ph.D.

Department of Statistics

Victor Strecher, Ph.D.

Department of Health Behavior and Health Education

University of Michigan, Ann Arbor, MI, USA

Abstract

Background. Many interventions in today's health sciences are multicomponent, and often one or more of the components are behavioral. Two approaches to building behavioral interventions empirically can be identified. The more typically used approach, labeled here the *classical* approach, consists of constructing a likely best intervention a priori, and then evaluating the intervention in a standard randomized controlled trial (RCT). By contrast, the emergent *phased experimental* approach involves programmatic phases of empirical research and discovery aimed at identifying individual intervention component effects and the best combination of components and dosages. **Purpose.** The purpose of this article is to provide a head-to-head comparison between the classical and phased experimental approaches and thereby highlight the relative advantages and disadvantages of these approaches when they are used to select program components and doses so as to arrive at the most potent intervention. **Methods.** A computer simulation was performed. The simulation modeled prior hypotheses of hypothetical scientists as well as hypothetical subject data. Each hypothetical scientist took both the classical and phased experimental approach to intervention development based on the same data. **Results.** The phased experimental approach consistently outperformed the classical approach while using the same number of experimental subjects. However, neither approach achieved maximum intervention potency more than about 30 percent of the time. **Limitations.** The simulation did not model every factor that can have an impact on intervention development. **Conclusions.** The phased experimental approach merits serious consideration, because it has the potential to enable intervention scientists to develop more potent behavioral interventions.

Background

In today's health sciences multicomponent (1) interventions (also called complex (2) and multifaceted (3) interventions) are increasingly common, and it is also increasingly common for one, several, or even all of the components to be behavioral. For example, depression may be treated with a combination of pharmacotherapy and talk therapy (3, 4); cardiovascular disease may be prevented or treated with a combination of medication, exercise, and diet (5); a smoking cessation program may include behavioral and pharmacological components (6).

Multicomponent behavioral interventions are used in prevention and treatment in many other health domains, including HIV/AIDS (7), obesity (8), diabetes (9) and gerontology (1).

In this article we use a novel simulation design to contrast and explicate the relative advantages and disadvantages of two different general approaches for empirically building and evaluating multicomponent behavioral interventions. For purposes of this article, we will label the more established of the two approaches *classical* and the more emergent approach *phased experimental*. The classical approach, which is currently the dominant one in intervention science, consists of constructing a likely best intervention a priori, based primarily on readings of the literature, theory and clinical experience. This intervention is then evaluated in a standard randomized controlled trial (RCT). In the course of the RCT data are collected not only on the outcomes of primary interest but on other variables as well, so that after the evaluation has been concluded quasi-experimental, nonexperimental and post-hoc analyses can be performed in order to shed light on what worked well and what might need improvement. Examples of such analyses include relating outcomes to naturally occurring variation in participation, compliance, or implementation fidelity. Conclusions drawn from the results of these analyses provide the basis for revisions that produce a refined version of the intervention.

The classical approach relies heavily on the RCT, which is the generally accepted method of determining the effectiveness of an intervention. Although the RCT grew out of the need to evaluate single-component interventions, it has been widely applied to multicomponent interventions as well. However, as has been noted by numerous authors (1, 2, 10) multicomponent interventions present some challenges that are outside the scope of assessment of overall treatment efficacy, and therefore are not well addressed by the RCT alone. One challenge is to build the most potent interventions by choosing the best combination of components out of the finite space of possibilities. Another challenge is to build efficient interventions by including only components that “pull their own weight” in terms of time, money, or other resources. Both of these challenges require determining not only that an intervention as a package has an effect, but whether and how much each component under consideration is likely to contribute to an effect.

In response to these challenges, phased experimental approaches to intervention development have begun to emerge. Phased experimental approaches include additional evidentiary steps along with the RCT as part of the process of building and evaluating multicomponent interventions. For example, the Medical Research Council of the United Kingdom (2) outlined an approach that includes programmatic phases of empirical research and discovery leading up to and informing a RCT. Building on this idea, Collins, Murphy, Nair, and Strecher (10) have suggested two evidentiary phases to precede and inform a RCT. The first phase, called screening, consists of randomized experimentation designed to obtain estimates of the effects of individual components and, in some cases, interactions between components. The resulting experimental evidence provides the basis for preliminary decisions about which components to select for inclusion. A second phase of additional experimentation, called

refining, is used to identify the best level or dose of one or more components, to investigate interactions between components, and to resolve any other remaining questions. Frequently these phases of evidence-gathering will employ highly efficient factorial and response surface designs (e.g. (11-14); see the simulation below for one example). Conclusions drawn from the results of these analyses form the basis for specification of an intervention that consists of a set of active components implemented at levels or doses selected to maximize potency. Cost information can be collected in the course of experimentation and included when decisions are made concerning choices of components and/or doses.

Purpose

The purpose of this article is to provide a head-to-head comparison between the classical and phased experimental approaches and thereby highlight the relative advantages and disadvantages of these approaches when they are used to select program components and doses so as to arrive at the most potent intervention. To achieve this goal we employed a novel simulation. The following questions will be addressed: (a) Which approach, the classical or the phased experimental, identified more potent interventions on average? (b) What was the impact of the size of the effects of individual components on the absolute and relative performance of the two approaches? (c) Did the absolute or relative performance of the two approaches differ depending on whether or not the objective was to build a highly efficient intervention?

Methods

Overview of the Simulation

Data sets were generated using a procedure (described below) designed to (a) mimic the kind of behavioral theory that forms the basis of the causal thinking of many behavioral scientists and (b) reflect some of the complexity that can occur in empirical behavioral

intervention data. For each generated data set, the two different approaches for selecting intervention components and dosages were applied. The classical approach consisted of selecting components and dosages a priori and performing a two-group RCT on all available subjects. This was followed by post-hoc analyses. By contrast, the phased experimental approach, as described in Collins, Murphy, Nair, and Strecher (10), began with an initial screening experiment performed on a portion of the sample. Following the screening experiment the remaining portion of the sample was used for a set of refining experiments to select components and dosages. The objective of each approach was to arrive at the most potent behavioral intervention, expressed in terms of an outcome variable Y . Evaluation of each approach in a particular data set was based on the mean value of Y that would be expected if the resulting intervention were applied to all subjects in the population.

Causal and Data Generation Models

Figures 1(a) and 1(b) are schematic causal models representing, respectively, a behavioral scientist's understanding of the relation between intervention components and intervention outcome, and how the relation between intervention components and intervention outcome was simulated in the current study. Figures 1(a) and 1(b) are directed acyclic graphs (15); the presence of an arrow from one variable to another indicates that the former variable *may* have a causal effect on the latter variable. A square represents an observed variable, and a circle represents an unknown, and hence unobserved, variable. The components of the intervention are denoted by $A1$ - $A6$. $A1$ can assume three levels: off, medium, and high (coded 0,1,2); $A2$ - $A6$ can be either off or on (coded 0,1). Higher values of Y indicate a better outcome. The absence of an arrow indicates conditional independence; for example, given the variable $M1$,

Y is independent of $A1$. In Figure 1(b) all relations have a positive (if any) dependency except the relation with the arrow labeled with a minus sign.

Figure 1(a) shows that the behavioral scientist hypothesizes that each intervention component will affect the outcome Y only by first affecting a mediator (16), represented in the figure by $M1$ - $M6$. These mediators in turn are hypothesized to affect Y . The data generation model depicted in Figure 1(b) is more complicated than the behavioral scientist's model, in order to introduce some of the unknown and/or uncontrollable variables that often impact behavioral interventions. First, participants may have received only a fraction of the assigned dose of a component; this is represented in the figure by the adherence variables $Ad1$ - $Ad6$. Second, in the data generation $A5$ can have, depending on the parameter settings, an indirect negative effect on Y through its effect on $Ad4$; in other words, $A5$ can reduce adherence to $A4$. This means that $A4$ and $A5$ can interact negatively, so that all else being equal, an intervention that included both $A4$ and $A5$ would be less potent than one that included $A4$ without $A5$. Third, the generative model includes a variable called *Type*, representing individual characteristics of the experimental subjects that, although unknown to the scientists and unobserved, could impact the level of adherence to the intervention, the mediators and outcome. *Type* was modeled as a binary variable with 1 representing "cooperative subject, good motivation to adhere, likely to register a higher response on Y ," and 0 representing "difficult subject, poor motivation to adhere, likely to register a lower response on Y ." In order to simplify the exposition, *Type* was modeled as a single variable even though in a real-life setting there are likely to be many such variables. To represent the fact that some factors may be ineffective, as shown in Figure 1(b) the mediators $M5$ and $M6$ have no effect on Y , and consequently $A5$ and $A6$ have no effect on Y . In addition, to

maintain simplicity and clarity of exposition no other population heterogeneity was built into the simulation. Thus it was unnecessary to control for pretreatment variables in any of the analyses.

Details of the actual data generating model underlying Figure 1(b) are provided in Appendix A. Averaging over the distribution of *Type*, *Ad1-Ad6*, *M1-M6* produces the marginal linear model

$$E[Y | A_1, \dots, A_6] = c_0 + c_1 A_1 + c_2 A_1^2 + c_3 A_2 + c_4 A_1 A_2 + c_5 A_1^2 A_2 + c_6 A_3 + c_7 A_4 + c_8 A_4 A_5 \quad (1)$$

Furthermore the variance of *Y* given A_1, \dots, A_6 is a function of the components A_1, \dots, A_6 , that is, the variance is nonconstant (see Appendix A).

Experimental conditions

In the simulation there were three conditions corresponding to three different standardized main effect sizes for the active intervention components *A1-A4*. The effect size conditions, following Cohen's (17) guidelines for statistical power in social and behavioral research, were small ($d=.2$), medium ($d=.5$) and large ($d=.8$). We selected true parameters in the models represented by Figure 1(b) (see Appendix A for the exact form of the models) to yield these specific standardized effect sizes for the coefficients in Equation (1). Within a condition standardized main effect sizes for *A1-A4* were the same; all omitted effects were assigned effect size $d=0$.

In real-life empirical settings the choice of how many components to test and, in the case of the phased experimental approach, which fractional factorial design to employ, is governed by the state of knowledge in a particular intervention area. This a priori state of knowledge and its effects on decision making in the classical and experimental approaches was modeled by associating with each of the three effect size conditions a hypothetical research community setting intended to reflect a moderate state of the knowledge in the community concerning the

effects of the intervention components. The moderate state of knowledge in the community was represented by an “a priori” multivariate normal distribution on the parameters in the underlying causal model (see Appendix A for the parameters). Each draw from the a priori distribution represents a hypothetical scientist in the community. The standard deviations in the a priori distribution were set so that the a priori standard deviation of the outcome Y was 2 times the true values of the c_j 's in equation (1) (resulting in a .5 ratio of each c_j to the a priori standard deviation of Y). Furthermore the a priori distribution was centered at the true parameter values so as to model a scenario in which the community, *on average*, knows which parameters are nonzero and which are not, but any particular hypothetical scientist may deviate from this average knowledge – in other words, may be wrong – to some degree.

Data generation in each of the three experimental conditions involved generation of 1000 simulated data sets and 1000 corresponding hypothetical scientists. For example, to generate a data set in the small ($d=.2$) effect size condition, we first derived the true parameter values for the generative model in Figure 1(b) which would yield the small standardized effects sizes in Equation (1), and used these to generate $N=2500$ random experimental subjects. Second, we constructed the hypothetical scientific community by the use of the a priori multivariate normal distribution on the parameters with means equal to the true parameter values and standard deviations equal to two times the true parameter values, and third we drew a vector realization to represent a hypothetical scientist in this community. This vector, which we called a vector of a priori parameter hypotheses, reflected the hypothetical scientist's understanding of the importance of each component and their interactions. We repeated these steps to generate a simulation of 1000 hypothetical scientists (represented by draws of a priori parameter hypotheses). Third, each of the hypothetical scientists was assigned a single data set containing

$N=2500$ subjects drawn independently from the data generation model underlying Figure 1(b).

The entire process was repeated for the moderate and large effect size conditions.

Operationalization of the classical and phased experimental approaches

This section contains a brief overview of the operationalizations of the classical and phased experimental approaches. A detailed description can be found in Appendix A.

Each generated data set was used twice: Once for the classical approach, and once for the experimental approach. The same vector realization of a priori parameter hypotheses informed the classical approach and the experimental approach for any given data set. If these hypotheses indicated that the effect of a component was expected to be positive the component was retained for testing in both approaches; at a minimum the four components expected by the hypothetical scientist to have the highest effects were retained. This resulted in four, five, or six retained components. The two approaches began with the same set of retained components.

The classical approach. The classical approach employed all $N=2500$ experimental subjects in a single RCT with one treatment and one control group. The experimental treatment group was given an intervention consisting of all of the retained components set to “on” (if A1 is retained it is set to “high”) and the remaining components, if any, set to “off.” The control group was given a placebo intervention in which A1-A6 were all set to “off.” An analysis was done to establish whether the treatment-control difference was significant, followed by post-hoc analyses to select active components. These analyses employed both the measures of adherence and the mediators and included dose-response and dose-mediator analyses. Because there may be lack of adherence, dose here is the self-selected, received, dose. For data sets indicating that A1 should be included in the intervention, a regression including a quadratic term was performed. The

results of these analyses were used to determine whether in the final version of the intervention A1 should be set to 0, 1, or 2, and whether A2-A6 should be set to “off” or “on.”

The phased experimental approach. The phased experimental approach employed $N=1600$ in an initial screening phase of experimentation and reserved $N=900$ for a subsequent refining phase. As in the classical approach condition, only the retained components were investigated. If four components were retained then a full factorial design with 16 cells was used in the screening phase. Otherwise, a 16 cell balanced fractional factorial design was employed, with the choice of the particular factorial design determined by the vector of a priori parameter hypotheses (see Appendix A for details). A fractional factorial design was used because in empirical settings it often takes considerable resources to construct and implement many different versions of an intervention. If six components were retained, a full factorial design would require implementing 64 cells (64 different experimental conditions), which would be prohibitively expensive in most areas of intervention science.

In the refining phase, additional small experiments using the remaining $N=900$, such as running a few additional cells, were conducted to shed light on situations in which there are unanticipated interactions, or in which an interaction made it unclear which combination of components was best. Also, if the results of the screening and refining phases suggested that the final intervention should include A1, the refining phase included running a condition with A1=low to determine the optimal dose of this component. Appendix A gives the full details. As in the classical approach, the results indicated the choice of the level settings for each of the 6 components that would comprise the intervention.

Dependent variables – evaluation criteria

This simulation investigated whether the classical or phased experimental approach resulted in more potent interventions. Here “more potent” is defined in terms of larger mean outcome. Using the generative model associated with Figure 1(b) the mean of the outcome Y can be computed for all configurations of levels for the intervention components.

Two different criteria were used to evaluate intervention potency. The Standard criterion corresponded to a scenario in which the presence of inactive or low-performing components is considered neutral and does not detract from the overall outcome. The Standard criterion was simply the mean of Y computed for the chosen set of intervention components and dosages applied to a particular data set. The High-Efficiency criterion corresponded to scenario in which the goal is to maximize the outcome using the fewest number of components by avoiding inactive and low-performing components. The High-Efficiency criterion was the mean of Y/k , where k represents the number of components selected to be included in the intervention. The difference between the two criteria can be illustrated by considering two interventions, one made up exclusively of active intervention components and another including the same active intervention components plus a few inactive components. The Standard criterion would yield the same outcome value for these two interventions. By contrast, the High-Efficiency criterion would yield a larger outcome value for the former intervention, in effect penalizing the latter intervention for any inefficiency due to inclusion of underperforming components.

Results

Table 1 shows mean expected differences between the classical and phased experimental approaches for both criteria, as well as the mean difference and standard error. For reference, the highest achievable mean outcome value is included. Table 1 shows that for both criteria and for all three component effect size conditions, the phased experimental approach produced a larger

criterion value. The difference between the classical and phased experimental approaches is statistically significant in every condition. The difference is smallest when individual component effects sizes are smallest, and when the standard criterion is used. The magnitude of the difference is larger overall when the high-efficiency criterion is used, and the standard error of the difference is smaller.

In no condition did the mean outcome match the highest achievable mean outcome. Across the three effect size conditions, the phased experimental approach attained 71 percent of the highest achievable mean on average for both the standard and high-efficiency criteria. The classical approach attained an average of 64 percent of the highest achievable mean for the standard criterion and 41 percent for the high-efficiency criterion. Table 2 shows the proportion of times each procedure attained maximum intervention potency. As would be expected based on the results in Table 1, this proportion was higher for the phased experimental approach, but it did not exceed 32 percent. Interestingly, even though as Table 1 shows the mean potency was closer to the maximum when the High-efficiency criterion was used, the percent of data sets achieving maximum intervention potency was lower.

Table 3 shows the proportion of data sets in which the classical approach, the phased experimental approach, or neither produced a larger criterion value. For both the standard and high-efficiency criteria the phased experimental approach produced a higher criterion value in more data sets than the classical approach. The two approaches were tied much more often when the standard criterion was used. The most ties occurred in the Standard criterion, small effect size condition. In this condition, the classical approach outperformed or tied the phased experimental approach about 64 percent of the time; the corresponding figure for the phased experimental approach was 85 percent. The risk of arriving at a less potent intervention using

the phased experimental approach was not large; it ranged from about 15 percent in the small effect size condition using the standard criterion to less than one percent.

As discussed above, the state of knowledge in the community concerning the parameters in the underlying model was represented by an a priori multivariate normal distribution over the parameters. The a priori multivariate distribution used above set the standard deviations equal to twice the true parameter value. Although not shown here, simulations were also completed using a priori multivariate distributions representing (a) more variability/uncertainty and (b) less variability/uncertainty (standard deviations equal to 5 and 1.25 times true parameter values, respectively) among scientists. The results from these simulations are consistent with those reported above, reflecting a robustness of the comparison between the phased experimental and classical approaches across a variety of settings in which less (or more) information concerning potential intervention components is available to the scientists making the decisions.

Discussion

The results of the simulation suggest that under the circumstances modeled in this simulation, a phased experimental approach to empirical intervention development outperformed a classical approach. The phased experimental approach produced more potent interventions on average across all effect sizes included in the simulation. This held whether the Standard or High-Efficiency criterion was used. The difference in performance between the two approaches was more pronounced when the High-Efficiency criterion was used.

The smallest difference between the classical and phased experimental approaches was in the Standard criterion, small effect size condition. This was the only condition in which the modal outcome was a tie between the two approaches. The phased experimental approach depends on being able to identify individual component effects. However, some interventions

may be comprised of numerous relatively weak components that when combined have an aggregate effect, even if no one component has an identifiable effect. The phased experimental approach may not be an improved way of building an intervention made up of numerous weak components when the cost of including inactive components is low and therefore efficiency per se is not a goal.

In this simulation, the phased experimental approach outperformed the classical approach by a wide margin whenever the High-efficiency criterion was used. This suggests that a phased experimental approach may be particularly useful when efficiency is an important goal of intervention development. Efficiency was considered here in a simple and relatively crude manner. When using a phased experimental approach it would be possible to take a more nuanced approach to efficiency by including information about resource requirements in relation to component effects when making decisions about inclusion of intervention components and choice of dosages or levels. For example, an intervention scientist might decide that a particularly costly component will be selected for inclusion in an intervention only if its individual effects can be demonstrated to exceed some predetermined level.

Differences in approach and resource requirements

One reason for the difference in performance between the classical approach and the phased experimental approach lies in their differential reliance on randomized experimentation. The post-hoc analyses that formed the basis of intervention refinement in the classical approach were non-experimental, non-randomized comparisons. By contrast, both the initial screening and subsequent refining phases of the phased experimental approach consisted solely of randomized experiments, which generally are subject to fewer alternative explanations and threats to internal validity than nonrandomized comparisons (18). Thus results of experiments

can be expected to be more reliably informative on average, not only for building a single intervention, but for adding incrementally to a cumulative knowledge base that will contribute to future interventions.

One question that arises in considering the phased experimental approach is whether the additional experimentation required by this approach requires additional resources. In this simulation the phased experimental approach outperformed the classical approach using exactly the same number of experimental subjects, suggesting that it is no more demanding with respect to sample size than the classical approach. However, the phased experimental approach may be more costly to implement under some circumstances. The classical approach typically requires implementation of two conditions, a treatment and a control. The phased experimental approach calls for an experimental design that can isolate the effects of individual intervention components. In the screening phase this will usually be some variation of a factorial design, requiring implementation of numerous conditions, each of which represents a different version of the intervention. The refining phase may require follow-up experimentation. Even when in the phased experimental approach the intervention is administered to the same number of people as in a comparable classical approach, there may be additional costs associated with implementing a wider variety of versions of the intervention and conducting follow-up experiments. It may also take more time to implement the phased experimental approach, and may require more training of intervention delivery staff.

Although the logistics of this process and the associated costs are a serious consideration, highly efficient fractional factorial designs offer a way to keep the number of experimental conditions manageable. Moreover, the short-term costs of building and evaluating an intervention must be weighted against long-range costs and benefits. Our results suggest that

the phased experimental approach may help identify more efficient and streamlined interventions. As Allore, Tinetti, Gill, and Peduzzi (1) noted, “Since each component of an intervention adds to the overall cost and complexity, being able to directly estimate component effects could greatly enhance efficiency by reducing the number of components introduced into clinical practice” (p. 14). Our results also suggest that under many circumstances the phased experimental approach may be likely to identify a more potent intervention than the classical approach. Thus, in some applications the long-range gains in terms of increased efficiency and public health benefits expected to result from the phased experimental approach may offset any additional up-front intervention development costs.

Performance in absolute terms

The best performance of either the classical or phased experimental approach in this simulation did not exceed 75 percent of the ideal intervention potency. How important it is to approach 100 percent potency depends on the intervention domain and what 75 percent intervention potency means in public health terms. It is possible that additional measures, or an entirely different approach, could result in interventions that come closer to maximum potency. One promising avenue for intervention refinement may lie in exploring ideas from engineering process control, as discussed in Rivera, Pew, and Collins (19).

Limitations

This simulation was designed to take an initial look at the question of whether a phased experimental approach is a reasonable way to build interventions. It involved only a very small set of conditions out of the infinite number of possibilities that can occur in practice. There are a number of potentially important factors that were not varied in the simulation. A few of these are: the underlying structural model, which could be varied to include more interactions and

higher-order interactions; the number of components under consideration; the number of active vs inactive components; other effect sizes besides the three used here; the impact of measurement noise on the mediator and outcome variables; the effect of complex data structures such as nesting (e.g. individuals within classrooms; patients within clinics); and sophisticated approaches to incorporating cost in the decision. Many other additional factors could be considered. Despite the limitations of this study and the need for additional research, we believe that the results of the simulation show clearly that the phased experimental approach is a promising alternative.

Conclusions

The classical approach is currently the most well-established approach to empirical development of behavioral interventions. However, an emergent strategy, labeled here the phased experimental approach, provides a systematic way of making evidence-based decisions about which components and which doses or levels should comprise an intervention. Comparison of the two approaches in real-world empirical settings is impractical. In the present article a simulation was presented that provides this comparison using a data generation procedure intended to mimic a plausible empirical scenario. The results suggested that the phased experimental approach merits serious consideration, because it has the potential to help intervention scientists to build more potent behavioral interventions. Possible exceptions to this are interventions whose effects comprise the accumulation of many weak components. Although in this simulation the phased experimental approach outperformed the classical approach overall, the resulting interventions still fell short of maximal possible intervention potency. More research is needed on methods to improve intervention potency and efficiency, and thereby increase public health benefits.

References

1. Allore HG, Tinetti ME, Gill TM, Peduzzi PN. Experimental designs for multicomponent interventions among persons with multifactorial geriatric syndromes. *Clinical Trials*. 2005;2(1):13-21.
2. Medical Research Council. A framework for development and evaluation of RCTs for complex interventions to improve health. London: Medical Research Council; 2000 [updated 2000; cited 2007 July 6]; Available from URL: <http://www.mrc.ac.uk/Utilities/Documentrecord/index.htm?d=MRC003372>. Available from.
3. Williams JW, Gerrity M, Holsinger T, Dobscha S, Gaynes B, Dietrich A. Systematic review of multifaceted interventions to improve depression care. *General Hospital Psychiatry*. 2007;29:91-116.
4. Rush AJ, Trivedi MH, Wisniewski SR, Nierenberg A, Stewart JW, Warden D, et al. Acute and longer-term outcomes in depressed outpatients who required one or several treatment steps: A STAR*D report. *American Journal of Psychiatry*. 2006;163(11):1905-17.
5. Cuffe M. The patient with cardiovascular disease: treatment strategies for preventing major events. *Clinical Cardiology*. 2006;29:II4-12.
6. Cofta-Woerpel L, Wright KL, Wetter DW. Smoking cessation 3: multicomponent interventions. *Behavioral Medicine*. 2007;32:135-49.
7. Golin CE, Earp J, Tien HC, Stewart P, Porter C, Howie L. A 2-arm, randomized, controlled trial of a motivational interviewing-based intervention to improve adherence to antiretroviral therapy (ART) among patients failing or initiating ART. *Journal of Acquired Immune Deficiency Syndrome*. 2006;42:42-51.

8. Bluford DA, Sherry B, Scanlon KS. Interventions to prevent or treat obesity in preschool children: A review of evaluated programs. *Obesity*. 2007;15:1356-72.
9. Narayan KM, Kanaya AM, Gregg EW. Lifestyle intervention for the prevention of type 2 diabetes mellitus: putting theory to practice. *Treatments in Endocrinology*. 2003;2:315-20.
10. Collins LM, Murphy SA, Nair VN, Strecher V. A strategy for optimizing and evaluating behavioral interventions. *Annals of Behavioral Medicine*. 2005;30:65-73.
11. Box GEP, Draper NR. *Empirical model-building and response surfaces*. New York: Wiley; 1987.
12. Box GEP, Hunter WG, Hunter JS. *Statistics for experimenters: An introduction to design, data analysis, and model building*. New York: Wiley; 1978.
13. Myers RH, Montgomery DC. *Response surface methodology*. New York: Wiley; 1995.
14. Wu CFJ, Hamada M. *Planning, analysis, and parameter design optimization*. New York: Wiley; 2000.
15. Pearl J. Graphs, causality, and structural equation models. *Sociological Methods and Research*. 1998;27(2):226-84.
16. MacKinnon DP, Fairchild AJ, Fritz MS. Mediation analysis. *Annual Review of Psychology*. 2007;58:593-614.
17. Cohen J. *Statistical power analysis for the behavioral sciences*. Mahwah, NJ: Lawrence Erlbaum Associates; 1988.
18. Shadish WR, Cook TD, Campbell DT. *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin; 2002.

19. Rivera DE, Pew MD, Collins LM. Using engineering control principles to inform the design of adaptive interventions: a conceptual introduction. *Drug and Alcohol Dependence*. 2007;88:S31-S40.

Table 1
 Mean Intervention Outcome under
 Classical and Phased Experimental Approaches

Component Effect	Phased			Maximum Possible
	Classical Approach	Experimental Approach	Difference (standard error)	
Size	Standard Criterion			
Small	6.91	7.23	.32 (.10)	9.65
Medium	20.01	22.82	2.75 (.40)	34.00
Large	37.00	42.30	5.29 (.71)	59.32
	High-Efficiency Criterion			
Small	1.51	2.58	1.01 (.02)	3.60
Medium	4.51	7.93	3.42 (.07)	11.53
Large	8.00	13.62	5.62 (.13)	19.73

Table 2
 Percent of Data Sets in Which Maximum Intervention Potency
 Was Reached

Component Effect Size	Phased Experimental	
	Classical Approach	Approach
	Standard Criterion	
Small	7.2	18.1
Medium	6.1	25.2
Large	6.5	31.8
	High-Efficiency Criterion	
Small	1.0	15.3
Medium	0.6	19.6
Large	1.0	23.7

Table 3

Percent of Data Sets with Higher Criterion Values

Component Effect Size	Phased Experimental		
	Classical Approach	Approach	Neither (tied)
	Standard Criterion		
Small	15.1	36.4	48.5
Medium	1.4	50.9	47.7
Large	0.6	53.7	45.7
	High-Efficiency Criterion		
Small	1.2	93.9	4.9
Medium	0.1	92.1	7.8
Large	0.4	92.8	6.8

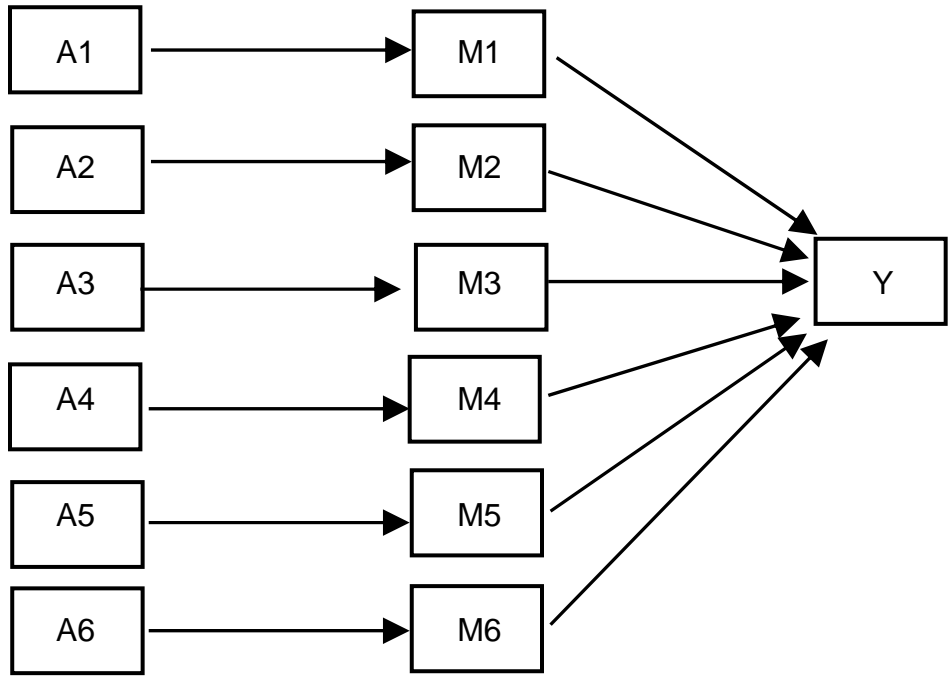


Figure 1(a). Behavioral scientist's causal model.

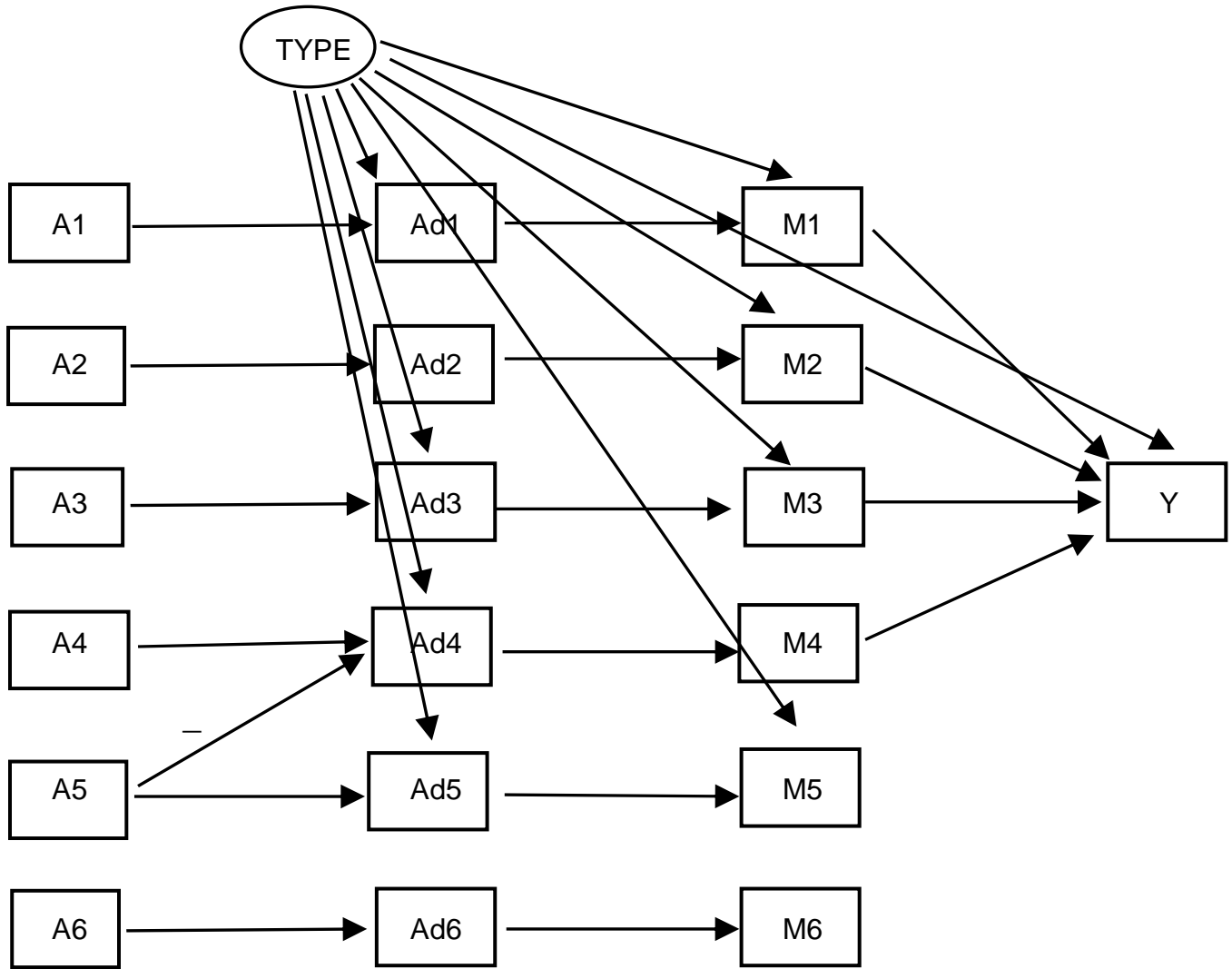


Figure 1(b). Data generation model.