Published in final edited form as:

Am J Prev Med. 2014 October ; 47(4): 498–504. doi:10.1016/j.amepre.2014.06.021.

Factorial Experiments: Efficient Tools for Evaluation of Intervention Components

Linda M. Collins, Ph.D.,

The Methodology Center and Department of Human Development & Family Studies, Penn State, University Park, PA

John J. Dziak, Ph.D.,

The Methodology Center Penn State, University Park, PA

Kari C. Kugler, Ph.D., and

The Methodology Center Penn State, University Park, PA

Jessica B. Trail, M.S.

Department of Statistics and The Methodology Center Penn State, University Park, PA

Abstract

Background—An understanding of the individual and combined effects of a set of intervention components is important for moving the science of preventive medicine interventions forward. This understanding can often be achieved in an efficient and economical way via a factorial experiment, in which two or more independent variables are manipulated. The factorial experiment is a complement to the randomized controlled trial (RCT); the two designs address different research questions.

Purpose—This article offers an introduction to factorial experiments aimed at investigators trained primarily in the RCT.

Method—The factorial experiment is compared and contrasted with other experimental designs used commonly in intervention science to highlight where each is most efficient and appropriate.

Results—Several points are made: factorial experiments make very efficient use of experimental subjects when the data are properly analyzed; a factorial experiment can have excellent statistical power even if it has relatively few subjects per experimental condition; and when conducting research to select components for inclusion in a multicomponent intervention, interactions should be studied rather than avoided.

Conclusions—Investigators in preventive medicine and related areas should begin considering factorial experiments alongside other approaches. Experimental designs should be chosen from a resource management perspective, which states that the best experimental design is the one that provides the greatest scientific benefit without exceeding available resources.

Corresponding author: Linda Collins, The Methodology Center, 204 E. Calder Way, Suite 400, State College, PA, 16801. Phone: 814-865-3253 Fax: 814-863-0000 LMCollins@psu.edu.

Linda M. Collins has no financial disclosures. John J. Dziak has no financial disclosures. Kari C. Kugler has no financial disclosures. Jessica B. Trail has no financial disclosures.

Keywords

Factorial experiment; Multicomponent interventions; Multiphase optimization strategy (MOST)

Moving intervention science toward richer behavioral theory, and more effective, cost-effective, efficient, and sustainable interventions, requires studying the individual and combined effects of sets of intervention components.¹⁻³ It may be surprising to some readers that for this objective, a factorial experiment is often the most efficient and economical alternative.^{4,5} The purpose of this article is to offer an initial introduction to factorial experiments for intervention scientists trained primarily in the randomized controlled trial (RCT).

Factorial experiments are gaining popularity in intervention science; recent applications of factorial experiments to examine individual intervention components include Apfel et al.,⁶ Strecher et al.,⁷ Collins et al.,³ Caldwell et al.,⁸ and McClure et al.⁹ Despite this increase in use, factorial experiments remain relatively rare in preventive medicine research. Of over 600 articles in Volumes 42-45 of this journal, only one featured a factorial experiment.¹⁰ Factorial experiments should not be considered a potential replacement for the RCT; rather, they can be an extremely useful complement to the RCT, because they address a different set of research questions. Thus depending on the research questions at hand, factorial designs may merit serious, informed consideration.

This article breaks no new ground methodologically, nor does it attempt to provide a complete course in experimental design. The objective is to explain, in a readily accessible manner, that under some circumstances factorial experiments analyzed by means of a classical factorial analysis of variance (ANOVA) can be highly efficient in preventive medicine research. To keep the discussion straightforward, it is confined to a widely used simple case of factorial experiment, namely that in which each factor has two levels.

This discussion compares and contrasts the factorial experiment with two other experimental designs that are common in intervention science. The first is referred to as the treatment-package RCT (TPRCT): a two-arm experiment comparing a treatment condition containing all the intervention components packaged together against a control condition (e.g., standard practice or placebo). The second is referred to as the multiple-arm comparative experiment (MACE): a k -arm experiment involving $k-1$ different forms of treatment and a control.

The TPRCT, factorial experiment, and MACE address different research questions

Consider a hypothetical example in which an investigator is developing a multicomponent intervention to produce weight loss in pre-diabetic adults.¹¹ The following components are under consideration: (1) personal counseling sessions; (2) pharmacological therapy (e.g., Orlistat)¹²; and (3) booster counseling session at 9 months. The outcome variable is weight loss at 1 year.

The TPRCT

Suppose an investigator wishes to evaluate a weight-loss intervention containing all three components. When the research question concerns evaluation of an intervention—that is, direct comparison of an intervention package to a control—the TPRCT is ideal. Each participant would be randomly assigned either to a treatment condition, consisting of the treatment package, or to a control condition, such as standard of care. The resulting data would be used to estimate mean weight loss for the treatment and control conditions, which would be directly compared to estimate the effect of the intervention.

The factorial experiment

In contrast, suppose the purpose of the experiment is to obtain information to be used in deciding which out of a set of candidate components is to be included in a streamlined, cost-effective intervention that will be evaluated in a later RCT. Now it becomes necessary both to estimate the individual effects of components and to assess whether one component impacts the effect of another. The factorial experiment is ideal for obtaining this information. In a factorial experiment based on the example, the presence versus absence of each component would be manipulated as an independent variable, and therefore corresponds to a factor in the experimental design. The three factors are designated: *COUN* for personal counseling sessions; *MED* for pharmacological therapy; and *BOOST* for a booster counseling session. The levels of *COUN* are 12 counseling sessions or 24 sessions; of *MED*, active medication or placebo; and of *BOOST*, provide a booster session or do not provide. Because there are three factors and each factor has two levels, this example would be a $2 \times 2 \times 2$, or 2^3 , factorial design.

To conduct this experiment the investigator would randomly assign participants to one of the eight conditions shown in Table 1. A participant assigned to Condition 3, then, would receive 12 counseling sessions, active medication, and no booster session. For notational purposes the factor levels are represented using the symbols “+” (for the “higher” level) and “−” (for the “lower” level).

Effects estimated in the classical factorial ANOVA—The factorial ANOVA can be used to estimate the *main effect* of each factor and the *interactions* between the factors. The main effect is the difference between the mean response at one level of a particular factor and the mean response at the other level, collapsing over the levels of all remaining factors (e.g., Montgomery¹³). Thus the main effect of *COUN* is

$$ME_{COUN} = \bar{\mu}_{+..} - \bar{\mu}_{-..}$$

(e.g., Kirk¹⁴), where $\bar{\mu}_{+..}$ and $\bar{\mu}_{-..}$ represent the mean response for the higher and lower levels, respectively, of *COUN*, averaging over all levels of *MED* and *BOOST* (the dot subscript means “averaged over”). Here $\bar{\mu}$ denotes a mean of condition means, taken across several conditions. In the experiment depicted in Table 1, $\bar{\mu}_{+..}$ is the mean weight loss for Conditions 5-8. $\bar{\mu}_{-..}$ is the mean (weighted by condition sample size if necessary) weight loss for Conditions 1-4. Thus, the main effect of *COUN* represents the difference between

mean weight loss when *COUN* is 24 sessions (high) and when *COUN* is 12 sessions (low), averaging over *MED* and *BOOST*.

Two factors interact if the effect of one is different depending on the level of the other. A two-way interaction is the difference in the effect of a particular factor across the levels of a second factor, averaging over all other factors.¹² Thus the *COUN*×*MED* interaction is defined as

$$INT_{COUN \times MED} = \left(\bar{\mu}_{++} - \bar{\mu}_{-+} \right) - \left(\bar{\mu}_{+-} - \bar{\mu}_{--} \right)$$

(e.g., Kirk¹⁴). Thus the *COUN*×*MED* interaction is the difference between the effect of *COUN* on weight loss when the intervention includes active medication (i.e., the mean of Conditions 7 and 8 minus the mean of Conditions 3 and 4) and when the intervention includes placebo medication (i.e., the mean of Conditions 5 and 6 minus the mean of Conditions 1 and 2), collapsing over the levels of *BOOST*. If the effect of *COUN* on weight loss is the same irrespective of the level of *MED*, this difference is zero, so *COUN* and *MED* do not interact.

Three factors interact when a two-way interaction itself depends on the level of the third factor. In experiments with more than three factors, higher-order interactions are straightforward extensions of the definitions above.

The MACE

Another experimental design option would be a four-arm MACE; for example, comparing *COUN* only, *MED* only, and *BOOST* only conditions to a control condition. Data from this MACE might be analyzed using a one-way ANOVA and pairwise contrasts. The effects estimated would be the simple effects of each component—that is, the effect of each component with the remaining components set to one specific level (see Kirk¹⁴)—not main effects by the classical definition discussed above.⁴ In general the MACE is appropriate if the research question concerns which single component is best, but less useful in determining which subset of components should be selected to include in a multicomponent intervention.

In factorial ANOVA, it is possible to have sufficient power even with small per-condition *ns*

In TPRCTs and MACEs, effects are estimated by directly comparing the means of individual experimental conditions. Therefore, statistical power depends directly on per-condition *n*. Whenever an arm is added to a MACE to make another comparison between treatments, or to a TPRCT to turn it into a MACE, the total *n* must be proportionally increased to provide power for this additional comparison.

The logic of powering factorial experiments is fundamentally different from the logic of powering TPRCTs or MACEs, because factorial experiments are designed to estimate main effects and interactions. Factorial experiments are superficially the same as MACEs, in that

they involve randomization to one of many treatment arms. However, they are analyzed using a factorial approach rather than primarily using pairwise comparisons.

Consider an experiment to test the efficacy of a single intervention component. This could be an RCT involving only two conditions, say 24 sessions of *COUN* and versus 12 sessions of *COUN*. Suppose the effect size (Cohen standardized difference) between these two conditions is expected to be at least .5. A power analysis shows that an overall *N* of 160, or *n*=80 per condition, will provide power >.8 at a Type I error rate $\alpha=.05$.

Now consider a hypothetical experiment with three factors (hence eight conditions). Suppose the standardized main effect of each factor is again at least .5, and the same $\alpha=.05$ is to be used. What if the factorial experiment was conducted with the same overall *N* of 160? Increasing from two to eight experimental conditions would result in a much smaller per-condition *n* of 20. However, this factorial experiment would provide approximately *the same power* to detect the main effect of any factor among the three, all else being equal, as the two-condition RCT. (As a caveat, in the presence of interactions, the effects estimated by the two experiments, i.e., main and simple effects, are not conceptually identical.⁴)

In factorial ANOVA, unlike the TPRCT, main effects are estimated by comparing the means of sets of conditions, not by directly comparing the means of individual conditions. When ME_{COUN} is computed, all conditions, and thus all 160 subjects, are included in the estimation of this effect (four conditions versus four other conditions). Similarly, the main effect of MED would be

$$ME_{MED} = \bar{\mu}_{.+} - \bar{\mu}_{.-},$$

and the main effect of BOOST would be

$$ME_{BOOST} = \bar{\mu}_{..+} - \bar{\mu}_{..-}.$$

Again, all of the conditions and all of the subjects are included in the estimation of each main effect. Note that some subjects who were in the + condition of COUN are in the – condition of MED or BOOST. Because each main effect and each interaction is estimated based on all subjects, a factorial experiment can have satisfactory power even with very small per-condition *ns*, provided the total *N* is sufficiently large. Under many circumstances, more factors could even be added to this experiment without increasing the total sample size to maintain the same approximate power, provided the expected effect sizes of each individual factor remained no smaller than the effects upon which the original power analysis was based.⁴

The meaning of a control group is different in the factorial experiment

In the TPRCT and the MACE there is typically one designated control group. By contrast, in a factorial experiment there is no single experimental condition that always constitutes the control group. Instead, each factor has its own control group, made up of a combination of conditions. In the example in Table 1, the control group for *COUN* consists of Conditions 1

– 4; for *MED* it consists of Conditions 1, 2, 5, and 6; and for *BOOST* it consists of Conditions 1, 3, 5, and 7. Condition 1, in which the participants receive only 12 counseling sessions, the placebo medication, and no booster, most closely resembles the control condition in the corresponding TPRCT. However, although this condition is part of the control group for each factor, it is not “the control group” for the entire experiment.

Thus, whereas in a TPRCT there are “treatment subjects” and “control subjects,” this is not so in factorial experiments. For example, a subject in Experimental Condition 3 in Table 1 is in the control group for *COUN* and *BOOST* but in the treatment group for *MED*.

Coding matters

A common, straightforward way to analyze and interpret the results of a factorial experiment is to fit a regression model to the results. This requires coding the factors (i.e., representing the factors as numbers). Two approaches to coding are widely used: dummy coding and effect coding. For dichotomous factors, the effect-coded levels are -1 and 1 , and the dummy-coded levels are 0 and 1 . Coding is often done automatically by statistical software and may not be apparent to the user. It is important to know which coding has been applied in an analysis, because the resulting regression coefficients represent substantively different quantities.

Effect coding is preferred for factorial ANOVA, for two reasons. First, effect coding produces regression coefficient estimates that correspond directly to the textbook definitions of ANOVA effects described earlier, after a simple rescaling. In contrast, dummy coding usually produces estimates of regression coefficients that do not correspond to the textbook definitions of ANOVA main effects and interactions, and must be interpreted differently. Second, with effect coding all estimated ANOVA effects are uncorrelated when *ns* are equal across experimental conditions, and nearly uncorrelated when *ns* are modestly unequal. With dummy coding, test statistics may be strongly correlated even when the *ns* are equal across experimental conditions, complicating interpretation and (all else being equal) reducing statistical power.¹⁵ Further details about coding appear in Kugler et al.¹⁶

As mentioned above, in an effect-coded ANOVA of a balanced (equal *ns* per condition) factorial experiment, the tests for main effects and interactions are uncorrelated. This is convenient for power analysis: if a particular power level is achieved for the test of the smallest regression coefficient of scientific interest, then it is automatically achieved for any equal or larger regression coefficient. Furthermore, the presence of non-zero interactions does not reduce power for a main effect of a given size. A good initial rough estimate of power for a main effect in a factorial experiment can be obtained by powering an experiment involving a single factor corresponding to the smallest expected main effect.

When equal *ns* across conditions cannot be maintained, as happens in field implementations, the estimated effects in the resulting factorial ANOVA are no longer uncorrelated. The correlations between effects are small unless the *ns* are extremely unbalanced, and can be handled by modern software, with only minor reductions in statistical power, using a regression approach. Even in the extreme case of a completely empty condition, it may be possible to salvage information by collapsing across one factor, as in Strecher et al.⁷

Estimation of interactions between intervention components is critical for understanding multicomponent interventions

Interactions increase interpretational complexity, and main effects must be interpreted thoughtfully when large interactions are present. This should be embraced, not avoided, when conducting an experiment to examine the individual and combined effects of intervention components. An alternative to the factorial experiment, such as separate two-arm TPRCTs for each component, or a MACE comparing each component in isolation to a single control condition, might seem more straightforward to interpret. However, these approaches do not produce certain critical information. For example, consider a MACE with four experimental conditions: (i) a control consisting of 12 counseling sessions, placebo drug, and no booster; (ii) 24 counseling sessions (with placebo drug and no booster); (iii) active drug (with 12 counseling sessions and no booster); and (iv) booster session (with 12 counseling sessions and placebo drug). Comparing each of the treatment conditions to the control condition (and to each other) would efficiently determine which of the three components would be best if used alone. However, because this experimental design does not enable estimation of interactions, it does not provide information about what would happen if some or all components were included together in a multicomponent intervention. Thus the possibility of interactions is a powerful motivation to conduct factorial experiments, not a reason to eschew them.

Different experimental designs are efficient for different research questions

Matched to appropriate research questions, the TPRCT, the MACE, and the factorial experiment use research subjects very efficiently. Applied to the wrong question, any design can become an inefficient choice. The TPRCT is usually the most efficient approach for evaluating the effect of an intervention package, because this requires directly comparing the mean response in a treatment package condition to the mean response in a control condition. However, a TPRCT provides no experimental evidence for evaluating the efficacy of individual components or for examining interactions between components. A MACE is appropriate if there are a limited number of alternatives under consideration (e.g., components can be implemented only one at a time). A factorial experiment is most appropriate when the ultimate objective is building a multicomponent intervention and the components can be combined in any way.

Whether a TPRCT, MACE, or factorial experiment is more economically efficient overall depends not only on sample size requirements but also on overhead costs associated with implementation of each experimental condition.⁴ Typically a factorial experiment will require more experimental conditions than other design alternatives, which in many settings can be an important consideration. When research questions involve interactions between independent variables, factorial experiments are the only option; no other experimental design enables estimation of these effects. Even when interactions are not expected, investigators may still wish to consider a factorial experiment in order to use research subjects efficiently in studying multiple factors.

General discussion

Advantages of factorial experiments

A factorial experiment can be the most efficient way to investigate a set of several intervention components. In many cases the main effects of five, six, or more factors can be studied with adequate power using essentially the same sample size that would be required for a single factor.⁴ This is important when participants are scarce, difficult to recruit, or expensive. Factorial experiments are possible even with cluster randomization.⁵

As discussed above, the classical definition of the main effect of a factor is the average effect of that factor across the levels of all other factors. As Fisher¹⁷ pointed out, a main effect for an intervention component provides indirect evidence for the robustness of that component's effect in the presence or absence of other components. This may be particularly valuable when individual components are investigated for the purpose of choosing an optimal intervention (e.g., Collins et al.³). In an ideal multicomponent intervention, one would wish to include components that are likely to exert some beneficial effect even if another component is later added or removed. A MACE does not address this question well.

Limitations of factorial experiments

A factorial experiment may be more costly than a TP-RCT, because the effects of individual components may well be smaller than those of the combined package, thus requiring more precision to detect.¹⁸ However, a factorial experiment is still much more efficient and feasible than investigating each component separately.⁴

Also, for a given n , a factorial experiment allocates fewer resources to any given condition than does a TP-RCT or MACE, simply because it involves more conditions. Suppose a power analysis indicates that a sample size of 400 is sufficient to detect all main effects and interactions of interest in a 2^5 factorial experiment under consideration. This experiment will have 32 conditions, so if 400 subjects are available there will be only approximately 12 per condition. This is sufficient in that under effect coding each main effect and each interaction is estimated using the entire sample of 400. However, this sample size is unlikely to be sufficient for pairwise comparisons of individual conditions. A comparison of two conditions in isolation would have a total n of about 25 rather than 400. Thus, when there are many factors, factorial experiments are most useful for identifying important main effects and interactions, not comparing each pair of conditions directly. Comparison of conditions, when feasible, is discussed by Kuehl¹⁹ and Taneja & Dudewicz.²⁰

Other varieties of factorial experiments

As previously stated, this brief article has not reviewed all types of factorial designs. Factorial experiments can have more than two levels per factor, and the number of levels can vary across factors. Such designs tend to be more resource-intensive than those with two levels per factor, but they can be useful under some circumstances (see Myers & Montgomery²¹). A more cost-efficient variety of factorial experiment is the fractional

factorial (e.g., Ledolter & Swersey;²² Collins et al.;⁴ Wu & Hamada²³), in which costs are reduced by implementing only a carefully selected subset of conditions.

Choosing an experimental design from a resource management perspective

Collins et al.⁴ described how to choose an experimental approach from a resource management perspective, in which the best experimental design is the one that provides the greatest scientific benefit without exceeding available resources. There are usually tradeoffs to consider. First, every experimental design is most efficient for a particular type of research question and much less efficient for others. No experimental design will be equally efficient for all research questions, so questions must be prioritized in terms of scientific importance before the most efficient design can be identified. Second, sometimes two designs present a choice between economy in terms of the number of experimental subjects and economy in terms of the number of conditions.⁴

Conclusions

The economy of the factorial experiment and the unique scientific information it offers make this design worth consideration in investigating the individual and combined effects of intervention components. Preventive medicine will benefit from continued conversation about how best to manage limited research resources to obtain greatest scientific yield.

Acknowledgments

Discussions with Timothy B. Baker, Susan A. Murphy, and Inbal Nahum-Shani contributed to this work. Preparation of this article was supported by Award Number P50DA010075 from the National Institute on Drug Abuse, Award Number P50CA143188 from the National Cancer Institute, KL2TR000126 from the National Center for Advancing Translational Sciences, and Award Number R01DK097364 from the National Institute of Diabetes and Digestive and Kidney Diseases. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors thank Amanda Applegate for editorial assistance.

References

1. Collins LM, Murphy SA, Nair VN, Strecher VJ. A strategy for optimizing and evaluating behavioral interventions. *Ann Behav Med*. 2005; 30(1):65–73. [PubMed: 16097907]
2. Collins LM, Murphy SA, Strecher VJ. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. *Am J Prev Med*. 2007; 32(5):S112–8. [PubMed: 17466815]
3. Collins LM, Baker TB, Mermelstein RJ, Piper ME, Jorenby DE, Smith SS, et al. The multiphase optimization strategy for engineering effective tobacco use interventions. *Ann Behav Med*. 2011; 41(2):208–26. [PubMed: 21132416]
4. Collins LM, Dziak JJ, Li R. Design of experiments with multiple independent variables: a resource management perspective on complete and reduced factorial designs. *Psychol Methods*. 2009; 14(3): 202–24. [PubMed: 19719358]
5. Dziak JJ, Nahum-Shani I, Collins LM. Multilevel factorial experiments for developing behavioral interventions: power, sample size, and resource considerations. *Psychol Methods*. 2012; 17(2):153–75. [PubMed: 22309956]
6. Apfel CC, Korttila K, Abdalla M, Kerger H, Turan A, Vedder I, et al. A factorial trial of six interventions for the prevention of postoperative nausea and vomiting. *N Engl J Med*. 2004; 350:2441–245. [PubMed: 15190136]

7. Strecher VJ, McClure JB, Alexander GL, Chakraborty B, Nair VN, Konkel JM, et al. Web-based smoking-cessation programs: results of a randomized trial. *Am J Prev Med.* 2008; 34(5):373–81. [PubMed: 18407003]
8. Caldwell LL, Smith RA, Collins LM, Graham JW, Lai MH, Wegner L, et al. Translational research in South Africa: evaluating implementation quality using a factorial design. *Child Youth Care Forum.* 2012; 41(2):119–36. [PubMed: 22707870]
9. McClure JB, Derry H, Riggs KR, Westbrook EW, John JS, Shortreed SM, et al. Questions about quitting (Q(2)): design and methods of a Multiphase Optimization Strategy (MOST) randomized screening experiment for an online, motivational smoking cessation intervention. *Contemp Clin Trials.* 2012; 33(5):1094–102. [PubMed: 22771577]
10. Bonsergent E, Agrinier N, Thilly N, Tessier S, Legrand K, Lecomte E, et al. Overweight and obesity prevention for adolescents: a cluster randomized controlled trial in a school setting. *Am J Prev Med.* 2013; 44(1):30–39. [PubMed: 23253647]
11. LeBlanc ES, O'Connor E, Whitlock EP, Patnode CD, Kapka T. Effectiveness of primary care-relevant treatments for obesity in adults: a systematic evidence review for the U.S. Preventive Services Task Force. *Ann Intern Med.* 2011; 155(7):434–47. [PubMed: 21969342]
12. Gray LJ, Cooper N, Dunkley A, Warren FC, Ara R, Abrams K, et al. A systematic review and mixed treatment comparison of pharmacological interventions for the treatment of obesity. *Obes Rev.* 2012; 13(6):483–98. [PubMed: 22288431]
13. Montgomery, DC. *Design and analysis of experiments.* Wiley; New York, NY: 2008.
14. Kirk, RE. *Experimental design: procedures for the behavioral sciences.* Sage; Los Angeles, CA: 2013.
15. Chakraborty B, Collins LM, Strecher VJ, Murphy SA. Developing multicomponent interventions using fractional factorial designs. *Stat Med.* 2009; 28:2687–2708. [PubMed: 19575485]
16. Kugler, KC.; Trail, JB.; Dziak, JJ.; Collins, LM. *Effect coding versus dummy coding in analysis of data from factorial experiments.* The Methodology Center, The Pennsylvania State University; University Park, PA: 2012.
17. Fisher, RA. *Statistical methods for research workers.* Oliver and Boyd; Edinburgh: 1925.
18. Wolbers M, Heemskerk D, Chau TTH, Yen NTB, Caws M, Farrar J, Day J. Sample size requirements for separating out the effects of combination treatments: Randomised controlled trials of combination therapy vs. standard treatment compared to factorial designs for patients with tuberculous meningitis. *Trials.* 2011; 12:26. [PubMed: 21288326]
19. Kuehl, RO. *Design of experiments: Statistical principles of research design and analysis.* 2nd ed.. Duxbury; Pacific Grove, CA: 2000.
20. Taneja BK, Dudewicz EJ. Selection in factorial experiments with interaction, especially the 2×2 case. *Acta Math Sin.* 1987; 3(3):191–203.
21. Myers, RH.; Montgomery, DC.; Anderson-Cook, CM. *Response surface methodology: process and product optimization using designed experiments.* 3rd ed.. Wiley; Hoboken: 2009.
22. Ledolter J, Swersey AJ. Using a fractional factorial design to increase direct mail response at Mother Jones magazine. *Qual Eng.* 2006; 18(469-475)
23. Wu, CFJ.; Hamada, M. *Experiments: planning, analysis, and parameter design optimization.* Wiley; New York: 2011.

Table 1
Experimental Conditions in the 2³ Factorial Design for the Weight-Loss Intervention

Experimental condition	Factor					
	<i>COUN</i>	symbol	<i>MED</i>	symbol	<i>BOOST</i>	symbol
1	12	–	Placebo	–	Not included	–
2	12	–	Placebo	–	Included	+
3	12	–	Drug	+	Not included	–
4	12	–	Drug	+	Included	+
5	24	+	Placebo	–	Not included	–
6	24	+	Placebo	–	Included	+
7	24	+	Drug	+	Not included	–
8	24	+	Drug	+	Included	+