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

Bibhas Chakraborty Linda Collins Susan Murphy

Appendix A: Data Generating Model

The data generating model is described below in terms of equations involving the intervention components ($A_1 - A_5$), the measures of adherence ($Ad_1 - Ad_5$), an unknown confounder *Type* (T), and the outcome (Y). In this model, $A_2 - A_5$ can take two values: 0 or 1 (absent or present); while A_1 can be 0, 1, or 2 (low, medium, or high). Note that subjects may receive a different dose of a component than that assigned. Measures of adherence (*Ad*'s) simply represent these doses. A multiplicative model is used below to describe the relation between *A*'s and *Ad*'s. The confounder *Type* follows a Bernoulli (1/2) distribution.

$\underline{A \rightarrow Ad}$**:**

$$
\begin{aligned}
Ad_1 &= (\eta_{10} + \eta_{11}\, T + e_1) \cdot A_1 \\
Ad_2 &= A_2 \\
Ad_3 &= (\eta_{30} + \eta_{31}\, T + e_3) \cdot A_3 \\
Ad_4 &= (\eta_{40} + \eta_{41}\, T + \eta_{42} A_5 + e_4) \cdot A_4 \\
Ad_5 &= (\eta_{50} + \eta_{51}\, T + e_5) \cdot A_5
\end{aligned}
$$

where each of e_1, e_3, e_4, and e_5 follows a normal distribution $N(0, \sigma_e^2)$, $\sigma_e = 0.1$. Note that there is no non-adherence to component A_2. The right hand side of the equations in the above display are truncated such that $Ad_j \in [0, A_j], \forall j$. The subsequent equations are only approximate due to this truncation.

$\underline{Ad \rightarrow Y}$**:**

$$Y = \beta_1\, T + \beta_2\, Ad_1 + \beta_3\, Ad_1^2 + \beta_4\, Ad_2 + \beta_5\, Ad_4 + \epsilon_Y; \quad \epsilon_Y \sim N(0, 3).$$

Marginal Form of Y, averaged over Ad's:

$$\begin{aligned} Y &= \beta_1\, T + \beta_2\, (\eta_{10} + \eta_{11}\, T)\, A_1 + \beta_3 \left((\eta_{10} + \eta_{11}\, T)^2 + e_1^2 \right) A_1^2 + \beta_4\, A_2 \\ &\quad + \beta_5\, (\eta_{40} + \eta_{41}\, T)\, A_4 + \beta_5\, \eta_{42}\, A_4\, A_5 \\ &\quad + \left(\epsilon_Y + e_1\, \beta_2\, A_1 + 2 e_1\, \beta_3\, (\eta_{10} + \eta_{11}\, T)\, A_1^2 + e_4\, \beta_5\, A_4 \right) \end{aligned} \tag{1}$$

Let e_T denote the sum of the 4 terms in the last row of the above display. Note that the term e_T has zero mean but heteroscedastic variance because some of the e_j's occur in products with the components. Because of zero mean, e_T functions like an error term. The generated Y will have a mean of the form

$$\begin{aligned} E[Y|A_1, \ldots, A_5] &= \frac{1}{2}\beta_1 + \beta_2\, (\eta_{10} + \frac{1}{2}\eta_{11})\, A_1 + \beta_3\, (\eta_{10}^2 + \frac{1}{2}\eta_{11}^2 + \eta_{10}\, \eta_{11} + \sigma_e^2) A_1^2 \\ &\quad + \beta_4\, A_2 + \beta_5\, (\eta_{40} + \frac{1}{2}\eta_{41})\, A_4 + \beta_5\, \eta_{42}\, A_4\, A_5 \\ &= c_0 + c_1\, A_1 + c_{11}\, A_1^2 + c_2\, A_2 + c_4\, A_4 + c_{45}\, A_4\, A_5 \end{aligned} \tag{2}$$

where each c_j is a function of (η, β, σ_e). In the simulations, we set these parameter values to ensure certain effect sizes as defined in the following section. Equation (2) above corresponds to equation (1) appearing in the article.

Standardized Effect Size

In our simulations, we set the parameter values so that the standardized effect size (Cohen's d) for the two-group comparison of the best treatment combination ($A_1 = A_2 = A_4 = 1$, $A_3 = A_5 = 0$) vs. the control where A_1 is set to its low level and all other components are absent (i.e., $A_i = 0, \forall i$) enjoys Cohen's benchmark values (small=0.2, medium=0.5, large=0.8). Cohen's d in this case is explicitly defined as:

$$\begin{aligned} d &= \frac{E\Big(Y|A_1 = A_2 = A_4 = 1; A_3 = A_5 = 0\Big) - E\Big(Y|A_i = 0, \forall i\Big)}{\sqrt{\frac{1}{2}\Big[\mathrm{Var}\Big(Y|A_1 = A_2 = A_4 = 1; A_3 = A_5 = 0\Big) + \mathrm{Var}\Big(Y|A_i = 0, \forall i\Big)\Big]}} \\ &= \frac{c_1 + c_{11} + c_2 + c_4}{f(\eta, \beta, \sigma_e)}, \end{aligned} \tag{3}$$

where f is some function of η, β, σ_e as can be seen from (1). The numerator follows from (2).

Parameter Values for Specific Standardized Effect Sizes

From now on we denote the parameter values used in a given simulation with a superscript 0. The true parameter values η^0 and β^0 are chosen so that the following conditions are satisfied:

1. The standardized effect size d as defined above attains Cohen's benchmark values (small=0.2, medium=0.5, large=0.8).

2. The active main effects (considered in the screening phase) are roughly equal in magnitude, while the active interaction is half the size of active main effects:

$$2c_1 + 4c_{11} = c_2 = c_4 = c, \quad \text{say} \qquad (4)$$
$$c_{45} = -\frac{c}{2}$$

3. The middle level of A_1 (i.e., $A_1 = 1$) is best, and the main effect of A_1 between levels 1 and 2 (as considered in refining phase) is also equal to the main effects considered in screening phase, i.e.,

$$(c_1 + c_{11}) - (2c_1 + 4c_{11}) = -c_1 - 3c_{11} = c \qquad (5)$$

4. The level of confounding, as quantified by $\beta_1\eta_{11}$ $(= \beta_1\eta_{31} = \beta_1\eta_{41} = \beta_1\eta_{51})$, is made equal to c corresponding to the large effect size $(d = 0.8)$.

The values of the η^0 stay the same across different effect sizes and are set to:

$$\eta^0_{10} = \eta^0_{30} = \eta^0_{40} = \eta^0_{50} = 0.50;$$
$$\eta^0_{11} = \eta^0_{31} = \eta^0_{41} = \eta^0_{51} = 0.25;\ \eta^0_{42} = -0.3125.$$

If we keep the η's and σ_e fixed, then $f(\eta, \beta, \sigma_e)$ appearing in the denominator of (3) can be written as $g(c)$, a function of just c. Also, each β can be expressed as a function of c. From (3), (4), and (5), we get

$$d = \frac{4c}{g(c)} \qquad (6)$$

For all three values of d (=0.2, 0.5, 0.8), we solve (6) for c by recursive method (calculating $g(c)$ by Monte Carlo integration). We get $c = 0.165, 0.415, 0.667$ for small, medium, and large effect size respectively. From the values of c, we can easily obtain the β values.

The β values corresponding to small standardized effect sizes are:

$$\beta^0_1 = 2.6680;\ \beta^0_2 = 0.9240;\ \beta^0_3 = -0.5945;\ \beta^0_4 = 0.1650;\ \beta^0_5 = 0.2640.$$

The β values corresponding to medium standardized effect sizes are:

$$\beta^0_1 = 2.6680;\ \beta^0_2 = 2.3240;\ \beta^0_3 = -1.4953;\ \beta^0_4 = 0.4150;\ \beta^0_5 = 0.6640.$$

The β values corresponding to large standardized effect sizes are:

$$\beta^0_1 = 2.6680;\ \beta^0_2 = 3.7352;\ \beta^0_3 = -2.4033;\ \beta^0_4 = 0.6670;\ \beta^0_5 = 1.0672.$$

Appendix B: Operationalization of the Phased Experimental Approach

As in the classical approach, each scientist following the phased experimental approach studies all the five components. In the screening phase only the two extreme levels (out of three) of A_1 are considered. We restrict the number of cells to 16 in our simulations, so a 16-cell resolution V balanced fractional factorial design with defining word $I = A_1A_2A_3A_4A_5$ is used (see Wu and Hamada, 2000, Ch. 4, for a technical discussion on defining word and resolution). The defining word completely specifies the aliasing pattern in the design. The above choice of design was used in a behavioral study on breast cancer prevention (see Nair et al., 2008). Since this is a resolution V design, the 2-way interactions are not aliased with each other (aliased with three-way interactions, as can be seen from the defining word). In general, the investigators choose the defining word based on prior substantive knowledge concerning the potential strength of higher order interactions relative to the likely noise in the data (e.g. if the size of any three way interaction is likely small compared to the noise level of the data, one can be more confident that the detected two-interaction effect is due to the two-way interaction and not to a three-way interaction). This is in accordance with the *hierarchical ordering principle* (see Wu and Hamada, 2000, p. 112) which states that, absent strong prior knowledge, higher order interactions can be expected to be of smaller size than lower order interactions. Note that in the setting described in the paper ($N = 1200$ subjects), only 800 subjects are used by each scientist in the screening phase of the study.

Screening Phase Analysis

The experiment is simulated using the 16-cell balanced fractional factorial design. A standard analysis of variance (ANOVA) is performed on the outcome Y and the five components. In the screening phase, the following steps are followed:

1. A 10% level of significance is used for testing the main effects and two-way interactions (to have greater power).

2. If the no. of significant effects is less than 3, rank-order the absolute values of the t-statistics corresponding to the main effects only (assuming that main effects are more likely to be significant than two-way interactions) and identify the largest 3. Move to the refining phase with the corresponding effects (treating them as significant).

In step 2 above, we chose the number of components to retain (say, k) to be equal to 3 to ensure that at least 50% of the components always pass the screening phase (3 is the smallest integer greater

than or equal to 5/2, 5 is the total no. of components). Since in general we expect that only a few components are likely to be active, the choice to carry forward the top 3 components to the refining phase is a reasonable one. By doing so, we are being conservative about the hypothesized effect of the components. This is a tuning parameter of the procedure and can vary from one investigator to another. We have conducted simulations for two other choices of this number (e.g., $k = 2$ and 4). A summary of simulation results across all three choices of k (e.g., 2, 3, and 4) can be found in Appendix C.

Moving Towards Refining Phase

As mentioned in the article, a part of the original sample in each simulated data set is reserved for the refining phase. The refining phase may or may not be conducted depending on the results obtained in the screening phase. In general the refining phase is employed in the simulation if (1) the three-level component A_1 is significant in the screening phase, or (2) there is at least one significant interaction involving A_1 (see Algorithm 1 below). The phased experimental approach, just like the classical approach, assumes the prior knowledge that A_1 is a special component having more than two levels. The refining phase uses multi-group experiments; standard analysis of variance, with 5% level of significance is used. In the setting described in the paper ($N = 1200$ subjects), the remaining 400 subjects are used by each scientist in the refining phase of the study.

Algorithm 1

This algorithm is used in the simulation to determine which components should be retained and which should be rejected, based on the results of the screening phase.
Input: set of components, significant effects (both main and interaction effects), estimated effect sizes, and signs of effects.
Output: best (treatment) combination.
Initialize: *best combination* $= [0, 0, 0, 0, 0]$, a 5-component vector. In the following, we will use the notation *best combination*(i) to denote the i-th element of the vector *best combination*.

1. Go through the set of components ($i = 1 : 5$): If main effect and all interactions of component i are insignificant, set *best combination*$(i) = 0$. If main effect of component i is significant, but none of its interactions are, look at the sign. If sign$(i) = +1$, set *best combination*$(i) = 1$. Else, *best combination*$(i) = 0$.

2. Now for any significant interaction, find $[P_1, P_2]$ = parent components of that interaction.

(a) If the main effect of the component P_1 (P_2) is insignificant, initialize sign(P_1) = 0 (sign(P_2) = 0). If P_1 (P_2) is significant, initialize sign(P_1) (sign(P_2)) to the sign of its main effect (either +1 or −1), respectively.

(b) Define *sign vector* as the vector consisting of sign(P_1), sign(P_2), and sign(interaction).

(c) If sign(P_1) = 0 but sign(P_2) $\neq$ 0, set sign(P_1) = sign(P_2) $\times$ sign (interaction). Do a similar operation for P_2.

(d) If both parents are insignificant, i.e., sign(P_1) = sign(P_2) = 0, go to step 4.

3. If sign(P_1) $\times$ sign(P_2) = sign(interaction), set *best combination*(P_1) = (sign(P_1) + 1)/2, and *best combination*(P_2) = (sign(P_2) + 1)/2.

4. For a significant interaction:

   – if *sign vector*= $[0, 0, 1]$, compare the cell means where both P_1 and P_2 are set to +1 (so the interaction P_1P_2 is also set to +1) vs. where both P_1 and P_2 are set to −1 (so P_1P_2 is set to +1).

   – if *sign vector*= $[0, 0, -1]$, compare the cell means where $P_1 = +1, P_2 = -1$ (so $P_1P_2 = -1$) vs. where $P_1 = -1, P_2 = +1$ (so $P_1P_2 = -1$).

   – otherwise, do cell-mean comparison of the following four cells: $P_1 = +1, P_2 = +1$ (so $P_1P_2 = +1$), $P_1 = -1, P_2 = -1$ (so $P_1P_2 = +1$), $P_1 = -1, P_2 = +1$ (so $P_1P_2 = -1$), $P_1 = +1, P_2 = -1$ (so $P_1P_2 = -1$).

   The combination of (P_1, P_2) that gives the highest cell mean is used to determine the best combination. This step implicitly assumes that the four cells do not differ with respect to other active components.

Note that the order in which step 2 considers interactions impacts the results. For example if sign(P_1) = 0, sign(P_2) = −1, sign(P_1P_2) = −1, sign(P_3) = +1, sign(P_1P_3) = −1, then the best combination setting for P_1 will depend on the order in which the interactions are considered. The simulations used the natural ordering, e.g., significant interactions from the ordered list 12, ..., 15, 23, ..., 25, 34, 35, 45 (using the notation ij to denote the interaction A_iA_j).

Refining Phase

1. When A_1 and all its interactions are insignificant, there is no refining phase. The best treatment combination is found by Algorithm 1.

2. When A_1 is significant, but none of its interactions are so, the refining phase uses a two-group follow-up experiment, in which A_1 is varied across the two groups, setting other components at their optimum level (obtained by applying Algorithm 1 on the screening phase results). One group receives the intermediate level of A_1 (not studied in screening phase), the other receives one extreme level depending on the sign of the screening phase estimate of the effect of A_1 (the level is the higher extreme if the sign is a +, lower extreme otherwise). Thus, the best treatment combination is found.

3. If only one interaction involving A_1 is significant, a 6-group follow-up experiment (3 levels of A_1 $\times$ 2 levels of the other component forming the interaction), setting all other components at their optimum levels as found by Algorithm 1, is used.

4. If two interactions involving A_1 are significant, then a 12-group ($3 \times 2^2 = 12$) follow-up experiment is used in the refining phase.

5. For three or more significant interactions involving A_1, conducting follow-up experiment becomes increasingly problematic (constructing many treatment groups), and also results become less reliable (low power for comparing many groups). In our simulations, no refining experiment is conducted in such cases – the best combination is determined by applying Algorithm 1 to the results of screening phase. As mentioned in the previous page, the order in which step 2 of Algorithm 1 considers interactions impacts the results. The simulations used the natural ordering, e.g., significant interactions from the ordered list 12, ..., 15, 23, ..., 25, 34, 35, 45 (using the notation ij to denote the interaction A_iA_j). However, these cases occurred very rarely in our simulation, and hence this step was rarely employed. In real life, investigators can come up with rules to proceed based on additional analysis (see Strecher et al., 2008, for an example).

The abstract discussion of Algorithm 1 and refining phase possibilities are made more concrete below with the help of three simulated examples:

Example 1

Suppose at the screening phase, the significant effects along with their signs are:

$$A_1(+), A_2(+), A_3(-)$$

Before running any follow-up experiment, Algorithm 1 will operate on this as follows:

- Best combination is initialized as $[0, 0, 0, 0, 0]$.

- By Step 1, A_1, A_2, A_3, A_4 and A_5 are set to 1, 1, 0, 0, and 0 respectively. So Best combination becomes $[1, 1, 0, 0, 0]$.

Since only the main effect of A_1 (but none of its interactions) is significant at the screening phase, a 2-group follow-up experiment is conducted at the refining phase, where the 2 groups correspond to the levels 1 and 2 of A_1. In both groups, A_2, A_3, A_4, and A_5 are fixed at levels 1, 0, 0, and 0 respectively – as determined by Algorithm 1 above.

Example 2

Suppose at the screening phase, the significant effects along with their signs are:

$$A_1(+), A_2(+), A_4(+), A_1A_2(+), A_4A_5(-)$$

Before running any follow-up experiment, Algorithm 1 will operate on this as follows:

- Best combination is initialized as $[0, 0, 0, 0, 0]$.
- By Step 1, A_3 is set to 0. Best combination remains same as before.
- By Step 2 (a – c), *sign vector* for the interaction A_1A_2 is $(1, 1, 1)$, and the *sign vector* for the interaction A_4A_5 is $(1, -1, -1)$.
- By Step 3, best combination becomes $[1, 1, 0, 1, 0]$.

Since one interaction involving A_1 is significant at the screening phase, a 6-group follow-up experiment is conducted at the refining phase, where the 6 groups correspond to the combinations (0, 0), (0, 1), (1, 0), (1, 1), (2, 0), and (2, 1) of the components A_1 and A_2. In all 6 groups, A_3, A_4, and A_5 are fixed at levels 0, 1, and 0 respectively – as determined by Algorithm 1 above.

Example 3

Suppose at the screening phase, the significant effects along with their signs are:

$$A_1(+), A_2(+), A_5(+), A_2A_3(-), A_3A_5(+)$$

Before running any follow-up experiment, Algorithm 1 will operate on this as follows:

- Best combination is initialized as $[0, 0, 0, 0, 0]$.
- By Step 1, A_1 is set at 1 and A_4 is set at 0. Best combination becomes $[1, 0, 0, 0, 0]$.
- By Step 2 (a – c), *sign vector* for the interaction A_2A_3 is $(1, -1, -1)$, and the *sign vector* for the interaction A_3A_5 is $(1, 1, 1)$.

- Step 3 applied to the interaction A_2A_3 sets A_2 at 1 and A_3 at 0. Best combination becomes $[1, 1, 0, 0, 0]$.
- Step 3 applied to the interaction A_3A_5 sets A_3 at 1 and A_5 at 1. Thus the best combination becomes $[1, 1, 1, 0, 1]$.

This is an example where the order in which interactions are considered in Algorithm 1 affects the results. Since in our simulations, we considered natural ordering as described in Algorithm 1, A_3A_5 is considered after A_2A_3. We could have ended up with a different best combination (e.g., $[1, 1, 0, 0, 1]$) had we considered A_3A_5 before A_2A_3.

Since only the main effect of A_1 (but none of its interactions) is significant at the screening phase, a 2-group follow-up experiment is conducted at the refining phase, where the 2 groups correspond to the levels 1 and 2 of A_1. In both groups, A_2, A_3, A_4, and A_5 are fixed at levels 1, 1, 0, and 1 respectively – as determined by Algorithm 1 above.

The matlab code for the entire simulation can be found at

`http://www.stat.lsa.umich.edu/~bibhas/MOSTcode.html`

Appendix C: Summary Results across Different Simulation Conditions

We conducted a series of simulations in order to investigate whether the results reported in the paper held across variation along two dimensions: (1) sample size (N), and (2) number of largest main effects retained in the screening phase as a decision rule (k). We investigated three different sample sizes: $N = 600$ (400+200), 1200 (800+400), and 2500 (1600+900); crossed with three different decision rules: $k = 2$, 3, and 4 (i.e., nine simulation settings in total). The following three tables correspond to the three tables in the paper, summarizing across all the nine settings. In general the classical approach tended to produce a larger $E(Y)$ than the phased experimental approach in small effect size, small sample size conditions; and the phased experimental approach tended to produce a larger $E(Y)$ than the classical approach in medium or better effect size conditions, even with small sample sizes ($N = 600$). The phased experimental approach tended to produce larger $E(Y)$ than the classical approach for larger k. In the conditions in which the four largest main effects were retained, the phased experimental approach consistently produced the larger $E(Y)$, even in the small effect size, small sample size ($N = 600$) condition. Details can be seen in the following tables.

Table 1: Whether the Classical (C) or the Phased Experimental (E) approach produced the largest value of $E(Y)$ under a variety of simulation conditions (This is a summary across 9 simulation settings of the entries that correspond to the Table 1 in the paper).

		k		
Sample Size	Effect Size	$k=2$	$k=3$	$k=4$
600	Small	C	C	E
	Medium	C	E	E
	Large	E	E	E
1200	Small	C	C	E
	Medium	E	E	E
	Large	E	E	E
2500	Small	C	E	E
	Medium	E	E	E
	Large	E	E	E

Table 2: Whether the Classical (C) or the Phased Experimental (E) approach produced a higher $E(Y)$ value in more data sets than its counterpart under a variety of simulation conditions (This is a summary across 9 simulation settings of the entries that correspond to the Table 2 in the paper).

		k		
Sample Size	Effect Size	$k=2$	$k=3$	$k=4$
600	Small	C	C	E
	Medium	C	E	E
	Large	E	E	E
1200	Small	C	C	E
	Medium	E	E	E
	Large	E	E	E
2500	Small	C	E	E
	Medium	E	E	E
	Large	E	E	E

Table 3: Whether the Classical (C) or the Phased Experimental (E) approach showed more accuracy in component selection under a variety of simulation conditions (This is a summary across 9 simulation settings of the entries that correspond to the Table 3 in the paper).

Dimension	Sample Size	Effect Size	k		
			$k=2$	$k=3$	$k=4$
Identifying the correct combination of components and levels more frequently	600	Small	C	E	E
		Medium	E	E	E
		Large	E	E	E
	1200	Small	E	E	E
		Medium	E	E	E
		Large	E	E	E
	2500	Small	E	E	E
		Medium	E	E	E
		Large	E	E	E
Identifying all the active components more frequently	600	Small	C	C	C
		Medium	C	C	E
		Large	C	E	E
	1200	Small	C	C	C
		Medium	C	C	E
		Large	E	E	E
	2500	Small	C	C	C
		Medium	E	E	E
		Large	E	E	E
Identifying all the inactive components more frequently	600	Small	E	E	E
		Medium	E	E	E
		Large	E	E	E
	1200	Small	E	E	E
		Medium	E	E	E
		Large	E	E	E
	2500	Small	E	E	E
		Medium	E	E	E
		Large	E	E	E