ADAPTIVE RANDOMIZED TRIAL DESIGNS THAT CANNOT BE DOMINATED BY ANY STANDARD DESIGN AT THE SAME TOTAL SAMPLE SIZE

Michael Rosenblum
Johns Hopkins University, mrosenbl@jhsph.edu
Adaptive Randomized Trial Designs that Cannot be Dominated by Any Standard Design at the Same Total Sample Size

BY MICHAEL ROSENBLUM

Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, U.S.A.
mrosenbl@jhsph.edu

SUMMARY

Prior work has shown that certain types of adaptive designs can always be dominated by a suitably chosen, standard, group sequential design. This applies to adaptive designs with rules for modifying the total sample size. A natural question is whether analogous results hold for other types of adaptive designs. We focus on adaptive enrichment designs, which involve pre-planned rules for modifying enrollment criteria based on accrued data in a randomized trial. Such designs often involve multiple hypotheses, e.g., one for the total population and one for a predefined subpopulation, such as those with high disease severity at baseline. We fix the total sample size, and consider overall power, defined as the probability of rejecting at least one false null hypothesis. We present adaptive enrichment designs whose overall power at two alternatives cannot simultaneously be matched by any standard design. In some scenarios there is a substantial gap between the overall power achieved by these adaptive designs and that of any standard design. We also prove that such gains in overall power come at a cost. To attain overall power above what is achievable by certain standard designs, it is necessary to increase power to reject some hypotheses and reduce power to reject others. We conclude by showing the class of adaptive enrichment designs allows certain power tradeoffs that are not available when restricting to standard designs. We illustrate our results in the context of planning a hypothetical, randomized trial of a new antidepressant, using data distributions from (Kirsch et al., 2008).

Some key words: Adaptive enrichment design; Overall power; Multiple testing procedure

1. INTRODUCTION

Tsiatis and Mehta (2003), in “On the Inefficiency of the Adaptive Design for Monitoring Clinical Trials,” prove the fundamental result that certain adaptive designs can always be dominated by a well chosen, standard, group sequential design. This result applies to the case of a single null hypothesis and designs where the only allowed adaptation is to modify the total sample size. They show for each adaptive design in a certain class, that there exists a standard, group sequential design with equal or greater power and smaller or equal expected sample size at the alternative. Therefore, in their setting adaptive designs do not add new possibilities for improving power or expected sample size at the alternative, compared to what can be achieved with standard, group sequential designs. We answer the question of whether an analogous result holds for a different type of adaptive design.
Adaptive enrichment designs involve preplanned rules for modifying enrollment criteria based on accrued data. They typically involve multiple null hypotheses corresponding to different populations, where there is prior evidence that the treatment is more likely to benefit certain populations. For example, Kirsch et al. (2008) give suggestive evidence that a certain class of antidepressants may only have a clinically meaningful benefit for those with severe depression at baseline. We use this example throughout. Denote those with severe depression at baseline as subpopulation 1, and those with moderate depression at baseline as subpopulation 2. Let $H_0$: be the null hypothesis of no average treatment benefit for the combined population, and let $H_{01}$ be the null hypothesis of no average treatment benefit for those with severe depression at baseline. Analogous null hypotheses for nested populations are considered by Wang et al. (2007), except that they define subpopulations in terms of a biomarker.

We next summarize our main result, for which precise definitions and proofs are given below. Consider trials with a treatment arm and control arm, where each participant has probability $1/2$ of being randomized to each. We focus on two alternatives. In the first, the treatment benefits each subpopulation at a minimum, clinically meaningful level $\Delta_{\text{min}}$; in the second, the treatment benefits subpopulation 1 at level $\Delta_{\text{min}}$, but has zero effect for subpopulation 2. Overall power at a given alternative is defined to be the probability of rejecting at least one false null hypothesis.

Define the class of standard designs with fixed, total sample size $n$ to be those that do not allow any change in enrollment criteria. We allow standard designs to use any multiple testing procedure for $\{H_{0C}, H_{01}\}$ that controls the familywise Type I error rate at a given level $\alpha$, in the strong sense defined by Hochberg and Tamhane (1987). We examine a class of two-stage, adaptive enrichment designs from (Rosenblum and van der Laan, 2011) for testing $\{H_{0C}, H_{01}\}$ with total sample size $n$. For each design in this class, we prove its overall power at the two aforementioned alternatives cannot simultaneously be matched by any standard design with the same total sample size. We also prove there can be substantial gaps between the overall power of certain adaptive designs and that of any standard design, in some scenarios.

Our results do not contradict Tsiatis and Mehta (2003), since their results are for designs that adapt only the total sample size. In contrast, the adaptive enrichment designs below have rules to modify enrollment criteria, but the total sample size is fixed.

Though one can achieve gains in overall power with adaptive enrichment designs compared to standard designs, this can come at a cost. We prove in Section 7 that there is a necessary tradeoff between overall power at one alternative, and power to reject $H_{0C}$ at the other.

2. Related Work

Jennison and Turnbull (2006) consider a similar setting as Tsiatis and Mehta (2003), in which there is one null hypothesis and the only possible adaptation is modifying the sample size. Their results are of a similar type as Tsiatis and Mehta (2003), except they consider a composite null hypothesis and prove additional theoretical guarantees. A caveat is that in the standard, group sequential designs constructed by Tsiatis and Mehta (2003) and by Jennison and Turnbull (2006) to dominate a given adaptive design, a greater number of interim analyses may be required than in the adaptive design.

We focus on adaptive enrichment designs. Examples of related work on adaptive enrichment designs include, e.g., Follmann (1997), Russek-Cohen and Simon (1997), Freidlin and Simon (2005), Jennison and Turnbull (2007), Wang et al. (2007, 2009), Trippa et al. (2012), and Simon and Simon (2013). Though adaptive designs have been compared to specific examples of stan-
Adaptive Enrichment Designs Not Dominated by Any Standard Design

and variance for outcomes \( \{ \mu, \sigma \} \) for subpopulation 1, subpopulation 2, and for the combined population, respectively, as \( \mu, \sigma \) for \( i \). We assume that conditioned on \( \{ (S_i, T_i) \}_{i:W_i=1} \), the stage 1 outcomes \( \{ Y_i \}_{i:W_i=1} \) are mutually independent, with each \( Y_i \) a random draw from a normal distribution with mean \( \mu_{S_iT_i} \) and variance \( \sigma_{S_iT_i}^2 \). Similarly, we assume that conditioned on \( \{ (S_i, T_i) \}_{i:W_i=2} \), the stage 2 outcomes \( \{ Y_i \}_{i:W_i=2} \) are mutually independent, with each \( Y_i \) a random draw from a normal distribution with mean \( \mu_{S_iT_i} \) and variance \( \sigma_{S_iT_i}^2 \).

Let \( \mu = (\mu_{10}, \mu_{11}, \mu_{20}, \mu_{21}) \) and \( \sigma^2 = (\sigma_{10}^2, \sigma_{11}^2, \sigma_{20}^2, \sigma_{21}^2) \). Define the true average treatment effect for subpopulation 1, subpopulation 2, and for the combined population, respectively, as

\[
\Delta_1 = \mu_{11} - \mu_{10}; \quad \Delta_2 = \mu_{21} - \mu_{20}; \quad \Delta_C = p_1 \Delta_1 + p_2 \Delta_2. 
\]

We say there is no (average) benefit of treatment for subpopulation \( s \in \{1, 2\} \) if \( \Delta_s \leq 0 \).

3. Problem Definition

3.1. Data Structure, Assumptions, and Hypotheses

Consider a population partitioned into two subpopulations. We assume the subpopulations are defined before the trial starts, based on attributes measured before randomization. Let \( p_s \) denote the proportion of the population in subpopulation \( s \), for each \( s \in \{1, 2\} \). We focus on trial designs with two stages, where the total sample size in each stage \( w \in \{1, 2\} \) is \( n_w = n_{fw} \), for \( f_1 + f_2 = 1 \). We assume \( 0.05 \leq f_1 \leq 0.95 \), so neither stage contains an inordinate fraction of the total sample size. In each stage \( w \) where the combined population is enrolled, we assume \( p_s n_{fw} \) participants are enrolled from each subpopulation \( s \).

Each participant is randomized to treatment or control with probability 1/2, independent of his/her subpopulation. For each subpopulation enrolled during a given stage, we assume half are randomized to each study arm. This assumption would approximately hold in designs that use block randomization stratified by subpopulation.

The data for each participant \( i \) is \( \{ (W_i, S_i, T_i, Y_i) \}_{i:W_i=1} \), where \( W_i \in \{1, 2\} \) is the stage in which the participant is enrolled, \( S_i \in \{1, 2\} \) is the subpopulation, \( T_i \in \{0, 1\} \) is the treatment arm indicator, and \( Y_i \in \mathbb{R} \) is the outcome. We consider normally distributed outcomes with mean \( \mu_{st} \) and variance \( \sigma_{st}^2 \) for each subpopulation \( s \in \{1, 2\} \) and study arm \( t \in \{0, 1\} \). For each participant \( i \), let \( \mu_{S_iT_i} \) and \( \sigma_{S_iT_i}^2 \) denote \( \mu_{st} \) and \( \sigma_{st}^2 \), respectively, evaluated at \( s = S_i, t = T_i \). Without loss of generality, assume \( W_i = 1 \) for \( i \leq n_1 \) and \( W_i = 2 \) otherwise. Let \( X^{(1)} \) denote the stage 1 data and \( X \) denote the data from both stages, i.e.,

\[
X^{(1)} = \{ (W_i, S_i, T_i, Y_i) \}_{i:W_i=1}; \quad X = \{ (W_i, S_i, T_i, Y_i) \}_{i=1}^n.
\]

We assume that conditioned on \( \{ (S_i, T_i) \}_{i:W_i=1} \), the stage 1 outcomes \( \{ Y_i \}_{i:W_i=1} \) are mutually independent, with each \( Y_i \) a random draw from a normal distribution with mean \( \mu_{S_iT_i} \) and variance \( \sigma_{S_iT_i}^2 \). Similarly, we assume that conditioned on \( \{ (S_i, T_i) \}_{i:W_i=2} \), the stage 2 outcomes \( \{ Y_i \}_{i:W_i=2} \) are mutually independent, with each \( Y_i \) a random draw from a normal distribution with mean \( \mu_{S_iT_i} \) and variance \( \sigma_{S_iT_i}^2 \).

Let \( \mu = (\mu_{10}, \mu_{11}, \mu_{20}, \mu_{21}) \) and \( \sigma^2 = (\sigma_{10}^2, \sigma_{11}^2, \sigma_{20}^2, \sigma_{21}^2) \). Define the true average treatment effect for subpopulation 1, subpopulation 2, and for the combined population, respectively, as

\[
\Delta_1 = \mu_{11} - \mu_{10}; \quad \Delta_2 = \mu_{21} - \mu_{20}; \quad \Delta_C = p_1 \Delta_1 + p_2 \Delta_2. 
\]

We say there is no (average) benefit of treatment for subpopulation \( s \in \{1, 2\} \) if \( \Delta_s \leq 0 \).

As described in Section 1, subpopulation 1 represents those with severe depression at baseline for whom prior data from Kirsch et al. (2008) suggest a greater chance of benefiting from treatment, compared to subpopulation 2. This motivates our focus on multiple testing procedures for the following null hypotheses, also considered by Wang et al. (2007):

\[
H_{0C} = \{ \mu \in \mathbb{R}^4 : \Delta_C \leq 0 \}; \quad H_{01} = \{ \mu \in \mathbb{R}^4 : \Delta_1 \leq 0 \}. 
\]
Let $\Delta_{\text{min}} > 0$ denote the minimum, clinically meaningful treatment effect, which we assume is predefined and the same for each subpopulation. We focus on power at the following simple alternatives, where the first represents an average benefit of $\Delta_{\text{min}}$ for each subpopulation and the second represents an average benefit of $\Delta_{\text{min}}$ only for subpopulation 1:

$$\omega_C : (\mu_{10}, \mu_{11}, \mu_{20}, \mu_{21}) = (0, \Delta_{\text{min}}, 0, \Delta_{\text{min}}); \quad \omega_1 : (\mu_{10}, \mu_{11}, \mu_{20}, \mu_{21}) = (0, \Delta_{\text{min}}, 0, 0).$$

(3)

Then at $\omega_C$ we have $(\Delta_1, \Delta_2) = (\Delta_{\text{min}}, \Delta_{\text{min}})$, and at $\omega_1$ we have $(\Delta_1, \Delta_2) = (\Delta_{\text{min}}, 0)$. At each of these alternatives, both $H_{01}$ and $H_{0C}$ are false, which follows from (1)-(3).

### 3-2. Decision Rules and Multiple Testing Procedures

We consider decision rules for the population to be enrolled in stage 2, which could be either the combined population or only subpopulation 1. In either case, $n_2$ participants are enrolled in stage 2. We assume the decision is made just after stage 1, and that all stage 1 data $X^{(1)}$ is available as input to the decision rule; this requires that each participant’s outcome is observed relatively soon after his/her enrollment.

Liu et al. (2002) describe the importance of measurability of the decision rule in adaptive designs. Measurability in our context is formally defined in Web Appendix A.

Let $D$ denote the class of decision rules defined to be all measurable functions $D$ from the set of possible values of $X^{(1)}$ to $\{C, 1\}$, where $D(X^{(1)}) = C$ indicates enrollment from the combined population in stage 2, and $D(X^{(1)}) = 1$ indicates enrollment from only subpopulation 1 in stage 2. Let $M$ denote the class of multiple testing procedures defined to be all measurable functions $M$ from the data $X$ to the set of null hypotheses $\{H_{0C}, H_{01}\}$, indicating which subset (if any) is rejected. The multiple testing procedure is implemented after all data from both stages have accrued.

A triple $(D, M, f_1)$ defines an adaptive enrichment design. The value $f_1$ is included in this definition since it determines the fraction of the total sample size that contributes data as input to the decision rule.

We require our designs to strongly control the familywise Type I error rate, also called the studywide Type I error rate. Regulators such as the U.S. Food and Drug Administration and the European Medicines Agency generally require studywide Type I error control for confirmatory trials involving multiple hypotheses (FDA and EMEA, 1998). An adaptive enrichment design $A = (D, M, f_1)$ is said to strongly control the familywise Type I error rate at level $\alpha$ if for any $\mu \in \mathbb{R}^4$, the probability of rejecting at least one true null hypothesis is at most $\alpha$, i.e.,

$$\sup_{\mu \in \mathbb{R}^4} P_\mu(A \text{ rejects at least one hypothesis in } H_{\text{TRUE}}(\mu)) \leq \alpha,$$

where $H_{\text{TRUE}}(\mu)$ is the subset of $\{H_{01}, H_{0C}\}$ that are true at $\mu$. $H_{\text{TRUE}}(\mu)$ contains $H_{01}$ if $\mu_{11} - \mu_{10} \leq 0$, and contains $H_{0C}$ if $p_1(\mu_{11} - \mu_{10}) + p_2(\mu_{21} - \mu_{20}) \leq 0$. Throughout, we focus on the case of $\alpha = 0.05$, though some of our results hold for any $\alpha$ as discussed in Section 10.

Define the following class of adaptive enrichment designs

$$A' = \{(D, M, f_1) : D \in D, M \in M, f_1 \in [0.05, 0.95]\},$$

and let $A$ denote the subclass of all $A \in A'$ that strongly control the familywise Type I error rate at level $\alpha = 0.05$.

Define the subclass of standard designs $S \subset A$ to be those that always enroll from the combined population in both stages, i.e., those designs $A = (D, M, f_1) \in A$ for which $D(X^{(1)}) = C$ for all values of $X^{(1)}$. 
The distribution of $X$ is determined by $\mu, \sigma^2, D, n, p_1, f_1$. We assume $p_1$ and $\sigma^2$ are known, and $D, n, f_1$ are set before the trial starts. The vector of unknown parameters is $\mu$, which we put no restrictions on.

3.3. Definition of Overall Power, Power to Reject $H_{0C}$, and Power to Reject $H_{01}$

We focus on three types of power at each alternative $\omega_C$ and $\omega_1$. Overall power is defined as the probability of rejecting at least one false null hypothesis. Power to reject $H_{0C}$ is defined as the probability of rejecting at least $H_{0C}$. Power to reject $H_{01}$ is defined as the probability of rejecting at least $H_{01}$. We prove in Section 7 that there are necessary tradeoffs between these different types of power. For any $A, A' \in A$, we say $A$ dominates $A'$ in overall power if $A$ has equal or greater overall power than $A'$ at $\omega_C$ and at $\omega_1$.

3.4. Statistics

For each $z$-statistic defined below, the first term is a difference between sample means comparing treatment versus control, and the second term standardizes the first. For each subpopulation $s \in \{1, 2\}$, for each stage $w \in \{1, 2\}$ in which subpopulation $s$ is enrolled, define the $z$-statistic comparing participants under treatment $t = 1$ versus control $t = 0$:

$$Z_s^{(w)} = \left\{ \frac{\sum_{i:W_i=w, S_i=s} Y_i T_i}{\sum_{i:W_i=w, S_i=s} T_i} - \frac{\sum_{i:W_i=w, S_i=s} Y_i (1-T_i)}{\sum_{i:W_i=w, S_i=s}(1-T_i)} \right\} \left\{ \frac{2(\sigma_s^2 + \sigma_{s1}^2)}{\sum_{i=1}^n 1[W_i = w, S_i = s]} \right\}^{-1/2},$$

where the indicator variable $1[B]$ equals 1 when $B$ is true and 0 otherwise. If no participants are enrolled from subpopulation $s$ in stage $w$, then the corresponding $z$-statistic $Z_s^{(w)}$ is undefined.

For each $w \in \{1, 2\}$, define the $z$-statistic pooling all participants enrolled in stage $w$:

$$Z^{(w)} = \left\{ \frac{\sum_{i:W_i=w} Y_i T_i}{\sum_{i:W_i=w} T_i} - \frac{\sum_{i:W_i=w} Y_i (1-T_i)}{\sum_{i:W_i=w}(1-T_i)} \right\} \left\{ \frac{1}{n_w^2} \sum_{i=1}^2 \sum_{s=1}^2 2\sigma_s^2 \sum_{i=1}^n 1[W_i = w, S_i = s] \right\}^{-1/2},$$

where we define $\sigma_s^2 = \sigma_{s0}^2 + \sigma_{s1}^2$ for each $s \in \{1, 2\}$. Define the following statistic based on all data from both subpopulations and stages:

$$Z^{\text{TOTAL}} = \left( \frac{n_1}{n_1 + n_2} \right)^{1/2} Z^{(1)} + \left( \frac{n_2}{n_1 + n_2} \right)^{1/2} Z^{(2)}.$$

In Web Appendix C, for the case where $\sigma_s^2$ equals a common value $\sigma^2$ for each $s \in \{1, 2\}$, $t \in \{0, 1\}$, we prove $Z^{\text{TOTAL}}$ reduces to the following $z$-statistic that pools all participants in both stages and compares those assigned treatment $t = 1$ versus control $t = 0$:

$$\left\{ \frac{\sum_{i=1}^n Y_i T_i}{\sum_{i=1}^n T_i} - \frac{\sum_{i=1}^n Y_i (1-T_i)}{\sum_{i=1}^n(1-T_i)} \right\} \left\{ \frac{4\sigma^2}{n} \right\}^{-1/2}. \quad (4)$$

If the combined population is enrolled in stage 2, define the cumulative $z$-statistic across both stages for each subpopulation $s \in \{1, 2\}$ as $Z_s^{\text{TOTAL}} = f_1^{1/2}Z_s^{(1)} + f_2^{1/2}Z_s^{(2)}$. In our notation, whenever there is an index $s$ for the subpopulation, it in the subscript.

4. Upper Bound on Overall Power of Any Standard Design in $S$

In Sections 4-9, we assume the outcome variances $\sigma_s^2$ equal a common value $\sigma^2$ for each $s \in \{1, 2\}, t \in \{0, 1\}$. We discuss the impact of different outcome variances in Section 10.
Let UMP_C denote the standard design that rejects H_{0C} if Z^{TOTAL} > \Phi^{-1}(1 - \alpha), where \Phi is the cumulative distribution function for the standard normal. This is the uniformly most powerful test of H_{0C} at level \alpha, which follows from Proposition 15.2 of (van der Vaart, 1998). Let n^{(\alpha, \beta)} denote the total sample size n for which the power of UMP_C to reject H_{0C} at \omega_C is 1 - \beta. It is straightforward to prove (as shown in Web Appendix C):

\[ n^{(\alpha, \beta)} = 2 \left( \Phi^{-1}(1 - \alpha) + \Phi^{-1}(1 - \beta) \right)^2 \left\{ p_1(\sigma_{10}^2 + \sigma_{11}^2) + p_2(\sigma_{20}^2 + \sigma_{21}^2) \right\} \Delta_{min}^{-2}. \] (5)

Below, we give a theorem bounding the overall power at \omega_1 for any standard design S \in S that matches the overall power of UMP_C at \omega_C. In Section 6, each adaptive enrichment design in a certain subclass of A is shown to exceed this bound.

**Theorem 1.** Let \alpha = 0.05 and consider any \beta \in (0, 1). Let n = n^{(\alpha, \beta)}. Assume \sigma^2_s = \sigma^2 for each s \in \{1, 2\}, t \in \{0, 1\}. Consider any standard design S \in S that has overall power at least 1 - \beta at \omega_C. Then at \omega_1, the overall power of S equals the overall power of UMP_C, which is

\[ \Phi(p_1\Phi^{-1}(1 - \beta) - p_2\Phi^{-1}(1 - \alpha)). \] (6)

The proof of Theorem 1 is given in Web Appendix D. It involves applying the Neyman–Pearson Lemma in Theorem 3.2.1 of (Lehmann and Romano, 2005) to prove that for any S \in S meeting the conditions of Theorem 1, S rejects at least one null hypothesis if and only if Z^{TOTAL} > \Phi^{-1}(1 - \alpha). This implies the overall power of S at \omega_1 equals that of UMP_C, which by straightforward computations in Web Appendix D is shown to equal (6).

5. **Definition of Subclass A_\ast \subset A of Adaptive Enrichment Designs**

In our running example, prior data suggest that subpopulation 2 may be less likely to benefit from the treatment than subpopulation 1. We next define a subclass of the adaptive enrichment designs A. Roughly speaking, each design gives up on subpopulation 2 and enrolls only subpopulation 1 in stage 2, if Z^{(1)}_2 is relatively small.

Consider any \theta \in \mathbb{R}. Define decision rule D_\theta(X^{(1)}) to be C if either Z^{(1)}_2 > Z^{(1)}_1 or Z^{(1)}_2 > \theta, and to be 1 otherwise. Let M_\theta(X) denote the following multiple testing procedure:

i. if D_\theta(X^{(1)}) = C and Z^{TOTAL} > \Phi^{-1}(0.95), reject H_{0C};
ii. if D_\theta(X^{(1)}) = 1 and Z^{TOTAL} > \Phi^{-1}(0.95), reject H_{01}.

Define the adaptive enrichment design A_{\theta, f_1} = (D_\theta, M_\theta, f_1). Also, define the class of designs

\[ A_\ast = \{ A_{\theta, f_1} \}_{\theta \in \mathbb{R}, f_1 \in [0.05, 0.95]}. \]

This is a subclass of designs from (Rosenblum and van der Laan, 2011, Section 3), except we relabeled the subpopulation indices. The proof that each A_{\theta, f_1} \in A_\ast strongly controls the familywise Type I error rate at level 0.05 is challenging, and is given in (Rosenblum and van der Laan, 2011, Section C of Supplementary Materials). The intuition behind these adaptive designs is explained in (Rosenblum and van der Laan, 2011, Section 3). Here, we focus on proving new properties for the class of designs A_\ast.
Adaptive Enrichment Designs Not Dominated by Any Standard Design

6. MAIN THEOREM

THEOREM 2. Let $\alpha = 0.05$ and consider any $\beta \in (0, 1)$. Let $n = n(\alpha, \beta)$ and assume $\sigma_{st}^2 = \sigma^2$ for each $s \in \{1, 2\}, t \in \{0, 1\}$. Consider any $p_1 \in (0, 1)$. Each adaptive enrichment design $A_{\theta, f_1} \in A$, has all of the following properties:

i. It strongly controls the familywise Type I error rate at level $\alpha = 0.05$.

ii. At alternative $\omega_C$, it has overall power equal to that of $\text{UMP}_C$.

iii. At alternative $\omega_1$, it has greater overall power than $\text{UMP}_C$.

iv. No standard design $S \in S$ simultaneously dominates its overall power at $\omega_C$ and at $\omega_1$.

Theorem 2 is proved in Web Appendix E. We next illustrate this theorem, using certain values of $\theta$ and $f_1$ selected to optimize the performance of $A_{\theta, f_1}$. We describe how these values of $\theta$ and $f_1$ were selected in Section 8, where we examine tradeoffs over a range of such values. Below, we round to two decimal places.

COROLLARY 1. Let $\alpha = 0.05$ and $\beta = 0.2$. Let $\sigma_{st}^2 = \sigma^2$ for each $s \in \{1, 2\}, t \in \{0, 1\}$, and let $n = n(\alpha, \beta)$. Consider the case where $p_1 = 1/2$.

(i.) For any standard design $S \in S$ with overall power at least $0.80$ at $\omega_C$, the overall power of $S$ at $\omega_1$ is at most $0.34$.

(ii.) The adaptive enrichment design $A_{-0.19, 0.40}$ has overall power $0.80$ at $\omega_C$, and has overall power $0.46$ at $\omega_1$.

Part (i) of Corollary 1 follows from applying (6) at $\alpha = 0.05$, $\beta = 0.2$, and $p_1 = 1/2$. Part (ii) follows from evaluating the power of $A_{-0.19, 0.40}$ at $\omega_C$ and $\omega_1$ using the multivariate normal distribution function in R with the mvtnorm package.

Corollary 1 implies there is a gap in overall power at $\omega_1$ between the adaptive enrichment design $A_{-0.19, 0.40}$ and any standard design in $S$ matching its overall power at $\omega_C$; specifically, the difference between overall power at $\omega_1$ is at least $0.46 - 0.34 = 0.12$, at $p_1 = 1/2$.

7. UNAVOIDABLE POWER TRADEOFFS FOR ADAPTIVE ENRICHMENT DESIGNS $A$

The theorem below shows a necessary tradeoff between overall power at $\omega_1$ and power to reject $H_{0C}$ at $\omega_C$, for any design in $A$.

THEOREM 3. Let $\alpha = 0.05$ and consider any $n > 0$. Assume $\sigma_{st}^2 = \sigma^2$ for all $s \in \{1, 2\}, t \in \{0, 1\}$. Consider any adaptive design $A \in A$. If $A$ has greater overall power than $\text{UMP}_C$ at $\omega_1$, then $A$ has less power to reject $H_{0C}$ at $\omega_C$ than $\text{UMP}_C$.

We outline the proof of Theorem 3 at the end of this section; full details are in Web Appendix F. The theorem implies that it is not possible for a design in $A$ to simultaneously beat $\text{UMP}_C$ in overall power at $\omega_1$ and to match the power of $\text{UMP}_C$ to reject $H_{0C}$ at $\omega_C$.

Theorems 2 and 3 have important implications for what is possible to achieve by using an adaptive enrichment design in $A$. Theorem 2 showed that every adaptive enrichment design in $A_*$ achieves overall power at $\omega_C$ and at $\omega_1$ that cannot simultaneously be matched by any standard design in $S$. Theorem 3 implies that such improvements in overall power necessarily come at a cost. Specifically, to achieve greater overall power at $\omega_1$ than $\text{UMP}_C$, an adaptive enrichment enrichment design $A \in A_*$ must sacrifice power to reject $H_{0C}$ at $\omega_C$ compared to $\text{UMP}_C$. We characterize this power tradeoff for the designs $A_*$ in Section 8.
Outline of the Proof of Theorem 3: (The full proof is given in Web Appendix F.) The key step is deriving the likelihood ratio test corresponding to any adaptive design $A \in \mathcal{A}$. Consider any $A = (D, M, f_1) \in \mathcal{A}$, and assume the data $X = \{(W_i, S_i, T_i, Y_i)\}_{i=1}^n$ is generated using this adaptive design. Assume the condition of the theorem that each $\sigma^2_{it} = \sigma^2$. Let $\mu^{(1)}$ equal $\omega_C$. Define $\mu^{(0)} = 0.5(\Delta_{\text{min}}, \Delta_{\text{min}}, \Delta_{\text{min}}, \Delta_{\text{min}})$, which is in $H_{01} \cap H_{0C}$. In Web Appendix B, we derive the likelihood ratio test of the null hypothesis $\mu^{(0)}$ versus alternative $\mu^{(1)}$, based on a realization of $X$. We show the likelihood ratio depends only on the following components of the likelihood:

$$
L_1(\mu; \{Y_i\}_i; W_i=1 \mid \{S_i, T_i\}_i; W_i=1) = \prod_{i:W_i=1} (2\pi \sigma^2_{S_i, T_i})^{-1/2} \exp\left\{ -\frac{(Y_i - \mu_{S_i, T_i})^2}{2\sigma^2_{S_i, T_i}} \right\};
$$

$$
L_2(\mu; \{Y_i\}_i; W_i=2 \mid X^{(1)}; \{S_i, T_i\}_i; W_i=2) = \prod_{i:W_i=2} (2\pi \sigma^2_{S_i, T_i})^{-1/2} \exp\left\{ -\frac{(Y_i - \mu_{S_i, T_i})^2}{2\sigma^2_{S_i, T_i}} \right\}.
$$

We prove the log-likelihood ratio can be expressed as follows:

$$
\log \left\{ \frac{L(\mu^{(1)}; X)}{L(\mu^{(0)}; X)} \right\}
= \log \frac{L_1(\mu^{(1)}; \{Y_i\}_i; W_i=1 \mid \{S_i, T_i\}_i; W_i=1)}{L_1(\mu^{(0)}; \{Y_i\}_i; W_i=1 \mid \{S_i, T_i\}_i; W_i=1)} + \log \frac{L_2(\mu^{(1)}; \{Y_i\}_i; W_i=2 \mid X^{(1)}; \{S_i, T_i\}_i; W_i=2)}{L_2(\mu^{(0)}; \{Y_i\}_i; W_i=2 \mid X^{(1)}; \{S_i, T_i\}_i; W_i=2)}
= \left\{ n^{1/2} \Delta_{\text{min}}/(2\sigma) \right\} \frac{Z_{\text{TOTAL}} - n\Delta_{\text{min}}^2/(8\sigma^2)}{}
(7)
$$

This implies the likelihood ratio test at level $\alpha = 0.05$ rejects the null hypothesis $\mu^{(0)}$ if and only if $Z_{\text{TOTAL}} > \Phi^{-1}(1-\alpha)$. Denote this test by $\phi(X)$. Let $1 - \beta$ denote the power of $\phi(X)$ at $\mu^{(1)}$, i.e., the probability that $\phi(X) = 1$ when $X$ is generated under $\mu^{(1)}$. We show that the power of UMP$_C$ to reject $H_{0C}$ at $\omega_C$ equals $1 - \beta$. By (7) and the Neyman–Pearson Lemma in Theorem 3.2.1 of (Lehmann and Romano, 2005), we have $\phi(X)$ is the most powerful test at level $\alpha$ of $\mu^{(0)}$ versus $\mu^{(1)}$, and any test of $\mu^{(0)}$ at level $\alpha$ with the same power (i.e., $1 - \beta$) must reject $\mu^{(0)}$ if and only if $Z_{\text{TOTAL}} > \Phi^{-1}(1-\alpha)$, except on an event with probability 0.

We use the above result to prove Theorem 3. Assume, for the sake of contradiction, that the design $A \in \mathcal{A}$ has greater overall power than UMP$_C$ at $\omega_1$, and greater or equal power as UMP$_C$ to reject $H_{0C}$ at $\omega_C$ (which equals $1 - \beta$). By definition, any $A \in \mathcal{A}$ strongly controls the familywise Type I error rate at level $\alpha = 0.05$. Let $\bar{\phi}(X)$ denote the indicator variable taking value 1 if $A$ rejects $H_{0C}$, and 0 otherwise. Then $\bar{\phi}(X)$ is a valid test at level $\alpha$ of $\mu^{(0)}$ versus the alternative $\mu^{(1)}$, with power at least $1 - \beta$ at $\mu^{(1)}$. By the last sentence in the previous paragraph, $\bar{\phi}(X) = 1$ if and only if $Z_{\text{TOTAL}} > \Phi^{-1}(1-\alpha)$ (except on an event with probability 0, which we ignore below). Therefore $A$ must reject $H_{0C}$ if and only if $Z_{\text{TOTAL}} > \Phi^{-1}(1-\alpha)$. This is used to show the decision rule $D$ satisfies $D(X^{(1)}) = C$ with probability 1; the argument is by contradiction, where we show that otherwise the familywise Type I error rate would exceed $\alpha = 0.05$ for certain $\mu \in H_{0C}$ for which $\Delta_1 > 0, \Delta_2 < 0$. We conclude that $A$ is a standard design, and so by Theorem 1 has overall power at $\omega_1$ equal or less than that of UMP$_C$. This contradicts the assumption above that $A$ has greater overall power than UMP$_C$ at $\omega_1$. This contradiction implies Theorem 3. Full details are given in Web Appendix F.
8. **Power Tradeoffs for the Subclass $A_s$ of $A$**

8.1. **Designs in $A_s$ Optimizing Overall Power at $\omega_1$**

We focus on $\alpha = 0.05$, $\beta = 0.2$, $n = n(t(\alpha, \beta))$, and $\sigma^2_{st}$ equal to a common value $\sigma^2$ for all $s \in \{1, 2\}$, $t \in \{0, 1\}$. By Theorem 3, there is a necessary tradeoff between overall power at $\omega_1$ and power to reject $H_{0C}$ at $\omega_C$. We examine this tradeoff for the class $A_s$ by determining the maximum overall power that can be achieved at $\omega_1$ subject to a constraint on power to reject $H_{0C}$ at $\omega_C$. We solve the following constrained maximization problem at certain values of $q, p_1$:

$$\max_{\theta \in \mathbb{R}, f_1 \in F} P_{\omega_1}(A_{\theta, f_1} \text{ rejects at least one false null hypothesis})$$

under constraint: $P_{\omega_C}(A_{\theta, f_1} \text{ rejects at least } H_{0C}) \geq q$, (8)

where $F = \{0.05, 0.06, \ldots, 0.95\}$; each probability in the above display depends implicitly on $p_1$. For any $q$ and $p_1$, denote the value of $(\theta, f_1)$ at the solution to the above problem by $(\theta(q, p_1), f_1(q, p_1))$. To simplify notation, let $A^{q, p_1}$ represent $A_{\theta(q, p_1), f_1(q, p_1)}$. An algorithm solving the above optimization problem is given in Web Appendix G.

As a special case, consider the solution to the above optimization problem at $(q, p_1) = (0.75, 0.50)$. The pair $(\theta, f_1)$ maximizing (8) under the constraint (9) is $(-0.19, 0.40)$. The corresponding design $A_{-0.19, 0.40}$ was used in part (ii) of Corollary 1 from Section 6. The power of this design to reject $H_{0C}$ at $\omega_C$ is 0.75, and its overall power at $\omega_1$ is 0.46. Compared to UMP$_C$, the design $A_{\theta, f_1}$ sacrifices 0.05 power to reject $H_{0C}$ at $\omega_C$, while gaining 0.46 − 0.34 = 0.12 overall power at $\omega_1$. The overall power of $A_{-0.19, 0.40}$ at $\omega_C$ is 0.80; since this design’s power to reject $H_{0C}$ at $\omega_C$ is 0.75, this implies that some of its overall power at $\omega_C$ is due to rejecting $H_{01}$. We examine the different contributions to overall power in Sections 8.2 and 9.

We next augment certain designs to allow simultaneous rejection of $H_{0C}$ and $H_{01}$ in some cases. Define UMP$_C^+$ to be the standard design that rejects $H_{0C}$ whenever UMP$_C$ does, and that rejects $H_{0C}$ and $H_{01}$ when both $Z_{TOTAL}^\pi$ and $Z_{1TOTAL}^\pi$ exceed $\Phi^{-1}(1 - \alpha)$. It follows from (Maurer et al., 1995) that UMP$_C^+$ strongly controls the familywise Type I error rate at level $\alpha$.

For any $(q, p_1)$, let $A^{q, p_1^+}$ denote the design identical to $A^{q, p_1}$, except that $A^{q, p_1^+}$ rejects $H_{0C}$ and $H_{01}$ when

$$D_{\theta(q, p_1)}(X(1)) = C, \quad Z_{TOTAL}^\pi > \Phi^{-1}(1 - \alpha), \quad \text{and} \quad Z_{1TOTAL}^\pi > \Phi^{-1}(1 - \alpha) + 0.055. \quad (10)$$

In Web Appendix H, we apply the method from Rosenblum and van der Laan (2011) to prove for each design $A^{q, p_1^+}$ considered below, it strongly controls the familywise Type I error rate at level $\alpha = 0.05$. The increase of 0.055 in the threshold in the last inequality in (10) is needed in this proof. Though the above augmentations affect power to reject $H_{01}$, overall power is unaffected; i.e., for any $\mu \in \mathbb{R}^d$, the designs UMP$_C$ and UMP$_C^+$ have equal overall power, and the designs $A^{q, p_1}$ and $A^{q, p_1^+}$ have equal overall power.

8.2. **Power Tradeoffs**

Define the set of designs

$$A_s^+ = \{A^{q, p_1^+} : q \in Q, p_1 \in \Pi\} \cup \{\text{UMP}_C^+\},$$

for $Q = \{0.7, 0.75, 0.78\}$ and $\Pi = \{0.01, 0.02, \ldots, 0.99\}$. For any $p_1 \in \Pi$, the designs $A^{0.7, p_1^+}$, $A^{0.75, p_1^+}$, $A^{0.78, p_1^+}$, UMP$_C^+$ have power 0.7, 0.75, 0.78, 0.80, respectively, to reject $H_{0C}$ at $\omega_C$. In other words, the designs $A^{0.7, p_1^+}$, $A^{0.75, p_1^+}$, $A^{0.78, p_1^+}$ sacrifice 10%, 5%, 2% power, respectively, to reject $H_{0C}$ at $\omega_C$, compared to UMP$_C^+$. (Here and below, each difference given as a percent represents the absolute difference, e.g., 10% represents the difference $0.10 = 0.80 - 0.70$.)
In Figure 1a, we plot overall power at \( \omega_1 \) versus \( p_1 \) for each design in \( A^*_S \). For each of \( A^{0.7,p_1+}, A^{0.75,p_1+}, A^{0.78,p_1+}, \text{UMP}_C^+ \), respectively, the corresponding points as \( p_1 \) varies from 0.01 to 0.99 are connected with a line. For each \( q \in Q \), the difference in overall power at \( \omega_1 \) between \( A^{q,p_1+} \) and \( \text{UMP}_C^+ \) is a decreasing function of \( p_1 \). This makes sense intuitively, since at smaller values of \( p_1 \), there is greater potential impact from switching to enroll only subpopulation 1. As an extreme case, for \( p_1 = 0.01 \), the overall power at \( \omega_1 \) of \( \text{UMP}_C^+ \) is close to \( \alpha = 0.05 \), while that of \( A^{0.7,p_1+} \) is close to 0.30. By Theorem 1, for any standard design \( S \in S \) that has overall power at least 0.8 at \( \omega_C \), the corresponding curve in Figure 1a would coincide with that of \( \text{UMP}_C^+ \).

When \( \omega_1 \) is true, there may be special interest in rejecting \( H_{01} \), since this corresponds precisely to the subpopulation benefiting from treatment. Figure 1b plots the probability of rejecting (at least) \( H_{01} \), for the same designs compared in Figure 1a. The differences between the curves are nearly identical to the differences between the corresponding curves in Figure 1a. This reflects that at \( \omega_1 \), the increased overall power of \( A^{q,p_1+} \) compared to \( \text{UMP}_C^+ \) is essentially all due to increased power to reject \( H_{01} \).

We next consider the special case of \( p_1 = 1/2 \), and focus on power to reject \( H_{01} \) at \( \omega_1 \). The adaptive designs \( A^{0.7,0.5+}, A^{0.75,0.5+}, A^{0.78,0.5+} \) have 17%, 12%, 7% greater power to reject \( H_{01} \) at \( \omega_1 \) than \( \text{UMP}_C^+ \), respectively. The tradeoff is that \( A^{0.7,0.5+}, A^{0.75,0.5+}, A^{0.78,0.5+} \) have 10%, 5%, 2% less power to reject \( H_{01C} \) at \( \omega_C \), respectively. All of these designs have overall power 0.8 at \( \omega_C \).

9. **Example: Designing a Randomized Trial of a New Antidepressant**

As introduced in Section 1, Kirsch et al. (2008) give suggestive evidence that a certain class of antidepressants may only have a clinically meaningful benefit for those with severe depression at baseline. We consider the problem of planning a randomized trial for a hypothetical, new antidepressant in this class. This motivating example was used in (Rosenblum and van der Laan,
Adaptive Enrichment Designs Not Dominated by Any Standard Design

Here we focus on different data generating distributions and different adaptive designs. Subpopulation 1 refers to those with severe depression at baseline, and subpopulation 2 refers to those with moderate depression at baseline. Define the outcome $Y_i$ for each participant to be the difference between his/her Hamilton Rating Scale of Depression (HRSD) score at baseline and at 6 weeks after enrollment.

In our simulations, we set the distribution of each $Y_i$ given $(S_i, T_i)$ to mimic features from the meta-analysis of Kirsch et al. (2008), which pools across 35 randomized trials. The mean improvement comparing baseline versus 6 weeks was 7.80 HRSD points for those assigned to the control (placebo) arm. We set $\mu_{st} = 7.80$ for each $s \in \{1, 2\}$. The minimum, clinically meaningful average treatment effect established by the United Kingdom’s National Institute for Clinical Excellence (NICE) in their guidelines for treating depression, is 3 HRSD points comparing drug versus placebo (NICE, 2004). We set $\Delta_{min} = 3$.

Instead of generating $Y$ according to a normal distribution, we set the conditional distribution of $Y$ given $S = s, T = t$ to be uniformly distributed over the integers in the interval $[\mu_{st} - 13.9, \mu_{st} + 13.9]$, for each $s \in \{1, 2\}, t \in \{0, 1\}$. Using a uniform rather than a normal distribution respects the fact that change in HRSD score is bounded, and also allows us to examine finite sample performance for a non-normal distribution. The value 13.9 was chosen so the standard deviation of $Y$ given $S = s, T = t$ is approximately 8, which approximately matches the estimated standard deviation under treatment and under control in (Kirsch et al., 2008).

For the above values, and for $\alpha = 0.05, \beta = 0.2$, the minimum sample size for UMP$_C^+$ to have power 0.8 to reject $H_{0C}$ at $\omega_C$, is $n = n(\alpha, \beta) = 176$, which follows from (5). We consider each $p_1 \in \{1/4, 1/2, 3/4\}$.

We compare the adaptive design $A^{0.75, p_1 +}$ to UMP$_C^+$ in scenarios corresponding to $\omega_C$ and $\omega_1$. For each, we simulated $10^6$ randomized trials, and report the proportion of trials in which different sets of null hypotheses are rejected in Table 1. In our simulations, the $z$-statistics in Section 3.4 were computed using sample variances in place of true variances; these were computed using the data available at the corresponding analysis, e.g., the sample variances used in place of true variances in $Z^{(1)}$ were computed using only stage 1 data.

The values in Table 1 match (to the nearest percent) the theoretical values computed assuming normally distributed outcomes and using the multivariate normal distribution function. This makes sense intuitively, since the joint distribution of $z$-statistics used in the above designs converges to a multivariate normal distribution, as we discuss in Section 10. For each $p_1$ considered, the familywise Type I error rate in our simulations at $(\Delta_1, \Delta_2) = (0, 0)$ was always at most 0.05 for $A^{0.75, p_1 +}$ and for UMP$_C^+$.

Consider the case of $p_1 = 1/2$. The benefit in using $A^{0.75, p_1 +}$ rather than UMP$_C^+$ is 12% increased power to reject $H_{01}$ (and 12% increased overall power) when the new antidepressant only benefits subpopulation 1, i.e., those with severe depression at baseline; the cost is 5% decreased power to reject $H_{0C}$ when the new antidepressant benefits both subpopulations (though overall power is the same for $A^{0.75, p_1 +}$ and UMP$_C^+$). Compared to the case of $p_1 = 1/2$, the difference between $A^{0.75, p_1 +}$ and UMP$_C^+$ in power to reject $H_{01}$ at $\omega_1$ is larger (18%) at $p_1 = 1/4$, and smaller (4%) at $p_1 = 3/4$. For each $p_1 \in \{1/4, 1/2, 3/4\}$, the power to reject $H_{0C}$ at $\omega_C$ is 5% less for $A^{0.75, p_1 +}$ than for UMP$_C^+$.

Another tradeoff involves the probability of rejecting both $H_{01}$ and $H_{0C}$, which is lower for $A^{0.75, p_1 +}$ than UMP$_C^+$ in all cases in Table 1. However, at $\omega_1$, it may be of primary importance to reject $H_{01}$; $A^{0.75, p_1 +}$ has greater power to reject $H_{01}$ at $\omega_1$ in all cases in Table 1.

Whether a clinical investigator finds the above tradeoffs useful depends on the relative value of rejecting different hypotheses and on prior evidence.
Table 1. Power Comparison of adaptive enrichment design $A^{0.75,p_1,+}$ and standard design $UMP_{C}^{+}$. Values are given as percents, rounded to the nearest percent. “Only $H_{01}$” is the probability of rejecting only $H_{01}$; “Only $H_{0C}$” is the probability of rejecting only $H_{0C}$; “Both” is the probability of rejecting both $H_{0C}, H_{01}$.

<table>
<thead>
<tr>
<th>$p_1$ Design</th>
<th>Overall</th>
<th>Only $H_{01}$</th>
<th>Only $H_{0C}$</th>
<th>Both</th>
<th>Overall</th>
<th>Only $H_{01}$</th>
<th>Only $H_{0C}$</th>
<th>Both</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1/4$ $A^{0.75,p_1,+}$</td>
<td>80</td>
<td>5</td>
<td>47</td>
<td>28</td>
<td>32</td>
<td>20</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>$1/4$ $UMP_{C}^{+}$</td>
<td>80</td>
<td>0</td>
<td>48</td>
<td>32</td>
<td>15</td>
<td>0</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>$1/2$ $A^{0.75,p_1,+}$</td>
<td>80</td>
<td>5</td>
<td>29</td>
<td>46</td>
<td>46</td>
<td>20</td>
<td>5</td>
<td>22</td>
</tr>
<tr>
<td>$1/2$ $UMP_{C}^{+}$</td>
<td>80</td>
<td>0</td>
<td>28</td>
<td>52</td>
<td>34</td>
<td>0</td>
<td>5</td>
<td>30</td>
</tr>
<tr>
<td>$3/4$ $A^{0.75,p_1,+}$</td>
<td>80</td>
<td>5</td>
<td>14</td>
<td>61</td>
<td>63</td>
<td>16</td>
<td>4</td>
<td>43</td>
</tr>
<tr>
<td>$3/4$ $UMP_{C}^{+}$</td>
<td>80</td>
<td>0</td>
<td>13</td>
<td>67</td>
<td>59</td>
<td>0</td>
<td>3</td>
<td>55</td>
</tr>
</tbody>
</table>

10. DISCUSSION, LIMITATIONS, AND GENERALIZATIONS

We characterized tradeoffs in using certain adaptive enrichment designs, and showed adaptive enrichment designs cannot be discarded on the grounds that they add no new possibilities compared to standard designs.

Though we assumed outcomes are normally distributed with known variances, we conjecture that our results may have implications more generally. The reason is that the designs $A_s$ use the data $X$ only through the z-statistics from Section 3.4. Under certain assumptions, the multivariate central limit theorem implies that these z-statistics converge to a multivariate normal distribution, as $n$ goes to infinity. We consider the setting of local alternatives in Web Appendix J, where we give a conjecture on the asymptotic properties of the designs $A_s$ under non-normal distributions. It is an area of future work to explore the impact of different data generating distributions on the above results, and to extend the results to a wider class of distributions.

We focused on the case of $\alpha = 0.05$. Theorems 1 and 3 hold for any $\alpha : 0 < \alpha < 1$. We made the simplifying assumption in Sections 4-9 that each outcome variance $\sigma_{st}^2$ equals a common value $\sigma^2$ for all $s \in \{1, 2\}, t \in \{0, 1\}$. Some of the above results hold under certain relaxations of this assumption. The adaptive enrichment designs $A_s$ strongly control the familywise Type I error rate at level 0.05 for any values of $\sigma_{st}^2 > 0$. We prove in Web Appendix E that parts (i)-(iii) of Theorem 2 hold if we replace the equal variance condition by $\sigma_{11}^2 + \sigma_{10}^2 \leq \sigma_{21}^2 + \sigma_{20}^2$.

The adaptive enrichment designs $A$ may require longer duration than a standard design using the same total sample size. This could occur if only subpopulation 1 is enrolled in the second stage, and enrollment is slower due to the more restrictive enrollment criteria.

Other limitations of the designs $A$ are that they require outcomes to be observed soon after enrollment, and they assume that the data generating distribution does not change over time. These limitations apply to many adaptive enrichment designs.

A limitation of the designs $A_s$ is they only allow two choices for stage 2 enrollment. An area of future work is to allow additional choices, such as increasing the proportion enrolled from one subpopulation and decreasing (but not setting to 0) the proportion enrolled from the other.

We focused on designs whose overall power at $\omega_C$ matches $UMP_C$, and explored tradeoffs in power at $\omega_1$. It is an area of future work to consider designs that sacrifice overall power at
Adaptive Enrichment Designs Not Dominated by Any Standard Design

\( \omega_C \) compared to UMP\(_C \). Also, all of our designs required that the total sample size be fixed in advance. It is an area of future work to prove results as above for designs with preplanned rules for modifying enrollment criteria and the total sample size.

SUPPLEMENTARY MATERIALS

The Web Appendices are provided on pages 14-36.

REFERENCES


A. DEFINITION OF MEASURABILITY OF DECISION RULES AND MULTIPLE TESTING PROCEDURES

Denote the sample space by \( \Omega_n = \{(1,2) \times (1,2) \times \{0,1\} \times \mathbb{R} \}^n \), which includes all possible realizations of \( X = \{(W_i, S_i, T_i, Y_i)\}_{i=1}^n \). Let \( \mathcal{F}_n \) denote the product \( \sigma \)-algebra generated from \( \{\mathcal{P}(\{1,2\}) \times \mathcal{P}(\{1,2\}) \times \mathcal{P}(\{0,1\}) \times \mathcal{B}\}^n \), where \( \mathcal{B} \) is the Borel \( \sigma \)-algebra on \( \mathbb{R} \), and for any set \( A, \mathcal{P}(A) \) denotes the power set of \( A \). Let \( \Omega' = \mathcal{P}(\{H_{01}, H_{02}\}) \), and let \( \mathcal{F}' = \mathcal{P}(\Omega') \). The class of multiple testing procedures \( \mathcal{M} \) is defined as all measurable functions \( M \) from \( (\Omega_n, \mathcal{F}_n) \) to \( (\Omega', \mathcal{F}') \).

The stage 1 sample space \( \Omega_{n1} = \{(1,2) \times (1,2) \times \{0,1\} \times \mathbb{R} \}^{n1} \) contains all possible realizations of \( X^{(1)} = \{(W_i, S_i, T_i, Y_i)\}_{i=1}^{n1} \). Let \( \mathcal{F}_{n1} \) denote the product \( \sigma \)-algebra generated from \( \{\mathcal{P}(\{1,2\}) \times \mathcal{P}(\{1,2\}) \times \mathcal{P}(\{0,1\}) \times \mathcal{B}\}^{n1} \). Let \( \Omega'' = \mathcal{P}(\{C,1\}) \), and let \( \mathcal{F}'' = \mathcal{P}(\Omega'') \). The class of decision rules \( \mathcal{D} \) is defined as all measurable functions \( D \) from \( (\Omega_{n1}, \mathcal{F}_{n1}) \) to \( (\Omega'', \mathcal{F}'') \).

B. LIKELIHOOD RATIO BASED ON DATA \( X \)

The distribution of \( X \) is determined by \( \Theta = \{\mu, \sigma^2, D, n, p_1, f_1\} \). Specifically, the distribution of \( \{(S_i, T_i)\}_{i:W_i=1} \) is uniform over all possible settings such that for each \( s \in \{1,2\}, t \in \{0,1\} \), exactly \( np_fj_1/2 \) stage 1 participants have \( (S_i, T_i) = (s, t) \). This follows from the assumption in Section 3-1 that in each stage where the combined population is enrolled, the proportion from subpopulation \( s \) is \( p_s \), and exactly half in each subpopulation are assigned to each arm. Conditioned on \( \{(S_i, T_i)\}_{i:W_i=1} \), the probability density function of \( \{Y_i\}_{i:W_i=1} \) is the following expression evaluated at \( w = 1 \):

\[
\prod_{i:W_i=w} \left(2\pi\sigma_{S_i,T_i}^2\right)^{-1/2} \exp \left\{ -\frac{(Y_i - \mu_{S_i,T_i})^2}{2\sigma_{S_i,T_i}^2} \right\}.
\]

The above display follows from the corresponding assumption in Section 3-1.

The distribution of the stage 2 data given \( X^{(1)} \) depends on the decision \( D(X^{(1)}) \). Conditioned on \( (X^{(1)}, D(X^{(1)}) = C) \), we have \( \{(S_i, T_i)\}_{i:W_i=2} \) is uniform over all possible settings such
Adaptive Enrichment Designs Not Dominated by Any Standard Design

distribution. We sometimes include other components from 
\[ \Theta = \{ \mu, \sigma^2, D, n, p_1, f_1 \} \] is \( \Theta \). For any \( \mu \in \mathbb{R}^4 \), 
let \( P_\mu \) denote the corresponding probability distribution of \( X \), and \( E_\mu \) denote expectation under this distribution. We sometimes include other components from \( \Theta \) in the subscript when they are not clear from context, e.g., writing \( P_{D, \mu} \) to denote the probability distribution of \( X \) under decision rule \( D \) and under \( \mu \).

Consider any decision rule \( D \in \mathcal{D} \). We construct the likelihood ratio test of \( \mu = \mu^{(0)} = (\mu^{(0)}_1, \mu^{(0)}_2, \mu^{(0)}_3, \mu^{(0)}_4) \) versus \( \mu = \mu^{(1)} = (\mu^{(1)}_1, \mu^{(1)}_2, \mu^{(1)}_3, \mu^{(1)}_4) \), based on the data \( X \), generated as described above. Recall \( 1[B] \) is the indicator variable taking value 1 if \( B \) is true and 0 otherwise. Since we assumed in Section 3 that \( W_i = 1 \) for each \( i \leq n_1 \) and \( W_i = 2 \) otherwise, we have that \( \{W_i\}_{i=1}^n \) is deterministic; these values are only used in the likelihood below to indicate the sets of participants in each stage.

Consider the likelihood \( L(\mu; X) \), which we partition into four components:

\[
L(\mu; X) = L(\mu; \{S_i, T_i, Y_i\}_{i=1}^n) = L_1(\mu; \{S_i, T_i\}_{i:W_i=1}) L_1(\mu; \{S_i, T_i\}_{i:W_i=2}) \\
\times L_2(\mu; Y_i_{i:W_i=1} \mid X(1)) L_2(\mu; Y_i_{i:W_i=2} \mid \{S_i, T_i\}_{i:W_i=2}, X(1)),
\]

where for constants \( c_1, c_2, c'_2 \) that do not depend on the parameters \( \mu \), we have

\[
\tilde{L}_1(\mu; \{S_i, T_i\}_{i:W_i=1}) = c_1 \times 1 \left[ \begin{array}{c} 1 \\
\text{for each } s \in \{1, 2\}, \sum_{i:W_i=1, S_i=s} \sum_{i:W_i=1, S_i=s} T_i = \sum_{i:W_i=1, S_i=s} (1 - T_i) = np_sf_{1s}/2 \end{array} \right] ;
\]

\[
\tilde{L}_2(\mu; \{S_i, T_i\}_{i:W_i=1} \mid X(1)) = \left\{ c_2 \left[ \begin{array}{c} 1 \\
\text{for each } s \in \{1, 2\}, \sum_{i:W_i=2, S_i=s} \sum_{i:W_i=2, S_i=s} T_i = \sum_{i:W_i=2, S_i=s} (1 - T_i) = np_sf_{2s}/2 \end{array} \right] \right\} \frac{1[D(X(1))=C]}{1[D(X(1))=1]}
\]

\[
\times \left\{ c'_2 \left[ \begin{array}{c} 1 \\
\sum_{i:W_i=2} \sum_{i:W_i=2} (1 - T_i) = nf_{2}/2 \end{array} \right] \right\} ;
\]

\[
L_2(\mu; Y_i_{i:W_i=2} \mid \{S_i, T_i\}_{i:W_i=2}, X(1)) = \prod_{i:W_i=2} \left( 2\pi \sigma_{S_iT_i}^2 \right)^{-1/2} \exp \left\{ -\frac{(Y_i - \mu_{S_iT_i})^2}{2\sigma_{S_iT_i}^2} \right\} .
\]

The constant \( c_1 \) is the probability of any given partition of the indices \( \{1, \ldots, n_1\} \) for the first stage participants into the four categories defined by \( s \in \{1, 2\}, t \in \{0, 1\} \), where there are exactly \( np_sf_{1s} \) participants in each subpopulation \( s \), exactly half of whom are assigned to each arm \( t \). The constant \( c_2 \) is the analogous probability for the second stage participants. The
constant $c_2$, which corresponds to the case where only subpopulation 1 is enrolled in stage 2, is the probability of any given partition of the indices \( \{n_1 + 1, \ldots, n\} \) into treatment \( (t = 1) \) and control \( (t = 0) \) with exactly half the indices in each category.

In constructing the likelihood above for the experiment generating the data \( X \), we took into account the adaptive nature of the design. Specifically, this was taken into account in the term \( L_2 \).

The components \( L_1 \) and \( L_2 \) of the likelihood do not depend on the parameters \( \mu \). Therefore, the log-likelihood ratio comparing any parameter vector \( \mu = \mu^{(1)} \) to \( \mu = \mu^{(0)} \) is

\[
\log \frac{L(\mu^{(1)}; X)}{L(\mu^{(0)}; X)} = \log \frac{L_1(\mu^{(1)}; \{Y_i\}_{i: W_i=1} | \{S_i, T_i\}_{i: W_i=1})}{L_1(\mu^{(0)}; \{Y_i\}_{i: W_i=1} | \{S_i, T_i\}_{i: W_i=1})} + \log \frac{L_2(\mu^{(1)}; \{Y_i\}_{i: W_i=2} | \{S_i, T_i\}_{i: W_i=2}, X^{(1)})}{L_2(\mu^{(0)}; \{Y_i\}_{i: W_i=2} | \{S_i, T_i\}_{i: W_i=2}, X^{(1)})}
\]

\[
= \sum_{i=1}^{n} \frac{1}{2\sigma^2_{S_iT_i}} \left\{ \left( Y_i - \mu^{(0)}_{S_iT_i} \right)^2 - \left( Y_i - \mu^{(1)}_{S_iT_i} \right)^2 \right\}
\]

\[
= \sum_{i=1}^{n} \frac{1}{2\sigma^2_{S_iT_i}} \left\{ 2Y_i(\mu^{(1)}_{S_iT_i} - \mu^{(0)}_{S_iT_i}) - \left( \mu^{(1)}_{S_iT_i} \right)^2 + \left( \mu^{(0)}_{S_iT_i} \right)^2 \right\}
\]

\[
= \sum_{i=1}^{n} \left\{ \frac{Y_i}{\sigma^2_{S_iT_i}} - \frac{\mu^{(1)}_{S_iT_i}}{\sigma^2_{S_iT_i}} - \frac{\left( \mu^{(1)}_{S_iT_i} \right)^2}{2\sigma^2_{S_iT_i}} \right\}.
\] (13)

The above derivation implies the following lemma:

**Lemma 1.** Consider any decision rule \( D \in \mathcal{D} \), and the corresponding experiment generating \( X \). The log-likelihood ratio comparing any parameter vector \( \mu = \mu^{(1)} \) to \( \mu = \mu^{(0)} \) is (13).

**C. PROOFS OF RESULTS FROM SECTION 3 AND LEMMAS USED IN PROOFS OF THEOREMS 1-3**

We first prove a lemma used in the Proof of Theorem 3.

**Lemma 2.** For any \( n > 0 \), for any \( \mu^{(1)}, \mu^{(2)} \in \mathbb{R}^4 \), for any event \( E \in \mathcal{F}_n \), if \( E \) has probability zero under \( \mu = \mu^{(1)} \) then \( E \) has probability zero under \( \mu = \mu^{(2)} \).

Proof: Consider the space \( (\Omega_n, \mathcal{F}_n) \) defined in Section A. We will define a measure on \( (\Omega_n, \mathcal{F}_n) \) that dominates each measure \( P_\mu \) for all \( \mu \in \mathbb{R}^4 \). We then show that the measure \( P_{\mu^{(1)}} \) is absolutely continuous with respect to \( P_{\mu^{(2)}} \), for any \( \mu^{(1)}, \mu^{(2)} \in \mathbb{R}^4 \).

Let \( \lambda \) denote Lebesgue measure on \( \mathbb{R} \), let \( \zeta_1 \) denote counting measure on \( \{1, 2\} \) (i.e., assigning measure 1 to each element of \( \{1, 2\} \)), and let \( \zeta_2 \) denote counting measure on \( \{0, 1\} \). Define \( \eta \) to be the product measure \( \zeta_1 \times \zeta_1 \times \zeta_2 \times \lambda \). Then the triple \( (\Omega_n, \mathcal{F}_n, \eta) \) is a measure space.

Consider any \( \mu \in \mathbb{R}^4 \). The density of \( P_\mu \) with respect to the dominating measure to be the product measure \( \zeta_1 \times \zeta_1 \times \zeta_2 \times \lambda \). Then the triple \( (\Omega_n, \mathcal{F}_n, \eta) \) is a measure space.

Consider any \( \mu \in \mathbb{R}^4 \). The density of \( P_\mu \) with respect to the dominating measure is the following, for any \( x = \{(w_i, s_i, t_i, y_i)\}_{i=1}^{n} \in \Omega_n : \)

\[
g_\mu(x) = 1 \text{ if for all } i \leq n_1, w_i = 1, \text{ and for all } i : n_1 + 1 \leq i \leq n, w_i = 2] L(\mu; x),
\]

where \( L(\mu; x) \) is defined in (12). This follows from the characterization of the distribution of \( X \) from Section B, and the assumption from Section 3 that \( W_i = 1 \) for each \( i \leq n_1 \) and equals 2 otherwise. The support of the density \( g_\mu(x) \) is identical for any value of \( \mu \in \mathbb{R}^4 \). This follows
since the components $\tilde{L}_1$ and $\tilde{L}_2$ of $L(\mu; x)$ do not depend on $\mu$, and the remaining components $L_1$ and $L_2$ are strictly positive for any $\mu \in \mathbb{R}^4$. Therefore, by part (iii) of Lemma 6.2 of (van der Vaart, 1998), for any $\mu^{(1)}, \mu^{(2)} \in \mathbb{R}^4$, the measure $P_{\mu^{(1)}}$ is absolutely continuous with respect to $P_{\mu^{(2)}}$. This proves the lemma, by the definition of absolute continuity.

For each $s \in \{1, 2\}$, define

$$\rho_s = \left\{ \frac{p_s(\sigma^2_s + \sigma^2_{s1})}{\sum_{s' = 1}^{2} p_{s'}(\sigma^2_{s'} + \sigma^2_{s'1})} \right\}^{1/2}. \quad (14)$$

**Lemma 3.** Consider any decision rule $D \in \mathcal{D}$ and any $\mu \in \mathbb{R}^4$. Then the following hold:

i. $Z_1^{(1)}$, $Z_2^{(1)}$ are independent, and each is normally distributed with unit variance.

ii. The mean of $Z_s^{(1)}$ is $\Delta_s \left[ n_1 p_s / \left\{ 2(\sigma^2_{s0} + \sigma^2_{s1}) \right\} \right]^{1/2}$ for each $s \in \{1, 2\}$.

iii. $Z^{(1)} = \rho_1 Z_1^{(1)} + \rho_2 Z_2^{(1)}$.

iv. Conditioned on $D(\mu^{(1)}) = C$, the following hold:

1. $(Z_1^{(2)}, Z_2^{(2)})$ is independent of $X^{(1)}$;
2. $Z_1^{(2)}, Z_2^{(2)}$ are independent;
3. $Z_s^{(2)}$ is normally distributed with unit variance and mean $\Delta_s \left[ n_2 p_s / \left\{ 2(\sigma^2_{s0} + \sigma^2_{s1}) \right\} \right]^{1/2}$ for each $s \in \{1, 2\}$;
4. $Z^{(2)} = \rho_1 Z_1^{(2)} + \rho_2 Z_2^{(2)}$;
5. $E_{\mu,D} \{ Z^{(2)} | D(\mu^{(1)}) = C \} = \sum_{s = 1}^{2} \rho_s \Delta_s \left[ n_2 p_s / \left\{ 2(\sigma^2_{s0} + \sigma^2_{s1}) \right\} \right]^{1/2}$.

v. Conditioned on $D(\mu^{(1)}) = 1$, we have $Z^{(2)} = Z_1^{(2)}$, which is independent of $X^{(1)}$ and normally distributed with unit variance and mean $\Delta_1 \left[ n_2 / \left\{ 2(\sigma^2_{10} + \sigma^2_{11}) \right\} \right]^{1/2}$.

vi. $\rho_1^2 + \rho_2^2 = 1$.

vii. For any $\mu \in \mathbb{R}^4$, the distribution of $X^{(1)}$ under $P_{D,\mu}$ is the same for any $D \in \mathcal{D}$.

Proof: Consider any $s \in \{1, 2\}$. Then we have, for each $t \in \{0, 1\}$,

$$\sum_{i:W_i = 1, S_i = s} 1[T_i = t] = \sum_{i=1}^{n} 1[W_i = 1, S_i = s] / 2 = p_sn_1/2, \quad (15)$$

which follows from the assumptions in Section 3. We have

$$Z_s^{(1)} = \left\{ \frac{\sum_{i:W_i = 1, S_i = s} Y_i T_i}{\sum_{i:W_i = 1, S_i = s} T_i} - \frac{\sum_{i:W_i = 1, S_i = s} Y_i (1 - T_i)}{\sum_{i:W_i = 1, S_i = s} (1 - T_i)} \right\} \left\{ \frac{2(\sigma^2_{s0} + \sigma^2_{s1})}{\sum_{i=1}^{n} 1[W_i = 1, S_i = s]} \right\}^{-1/2}, \quad (16)$$

where (16) follows from (15). Therefore, conditioned on $\{(S_i, T_i)\}_{i:W_i = 1}$, we have $Z_1^{(1)}$ and $Z_2^{(1)}$ are independent, each being normally distributed, which follows from the set of outcomes $Y_i$ involved in each of these statistics being disjoint, and the assumption from Section 3 that conditional on $\{(S_i, T_i)\}_{i:W_i = 1}$ the stage 1 outcomes are mutually independent.
By (16), the mean of $Z_s^{(1)}$ conditional on \{$(S_i, T_i)$\};$W_i=1$ equals
\[
\left\{ \frac{(p_s n_1/2) \mu_{s1}}{p_s n_1/2} - \frac{(p_s n_1/2) \mu_{s0}}{p_s n_1/2} \right\} \left\{ \frac{\sigma^2_{s0} + \sigma^2_{s1}}{p_s n_1/2} \right\}^{-1/2} = \Delta_s \left\{ \frac{p_s n_1}{2(\sigma^2_{s0} + \sigma^2_{s1})} \right\}^{1/2}.
\] (17)

Since this does not depend on \{$(S_i, T_i)$\};$W_i=1$, the (unconditional) mean of $Z_s^{(1)}$ equals $\Delta_s \left[ n_1 p_s/ \{ 2(\sigma^2_{s0} + \sigma^2_{s1}) \} \right]^{1/2}$. By (16), the variance of $Z_s^{(1)}$ conditional on \{$(S_i, T_i)$\};$W_i=1$ equals
\[
\left\{ \frac{(p_s n_1/2)^2 \sigma^2_{s1}}{(p_s n_1/2)^2} + \frac{(p_s n_1/2)(2\sigma^2_{s0} + \sigma^2_{s1})}{(p_s n_1/2)^2} \right\} \left\{ \frac{\sigma^2_{s0} + \sigma^2_{s1}}{p_s n_1/2} \right\}^{-1} = 1.
\] (18)

Since, as shown above, the conditional mean of $Z_s^{(1)}$ given \{$(S_i, T_i)$\};$W_i=1$ is a constant, we have the (unconditional) variance of $Z_s^{(1)}$ equals 1. We have shown that conditioned on \{$(S_i, T_i)$\};$W_i=1$, $Z_1^{(1)}$ and $Z_2^{(1)}$ are independent, each being normally distributed with mean and variance not depending on \{$(S_i, T_i)$\};$W_i=1$. Therefore, $(Z_1^{(1)}, Z_2^{(1)})$ is independent of \{$(S_i, T_i)$\};$W_i=1$, and so the conditional distribution of $(Z_1^{(1)}, Z_2^{(1)})$ given \{$(S_i, T_i)$\};$W_i=1$ equals the unconditional distribution of $(Z_1^{(1)}, Z_2^{(1)})$. The above argument implies $Z_1^{(1)}$ and $Z_2^{(1)}$ are independent, each being normally distributed with means and variances as above. We have shown (i) and (ii).

To show (iii), we have
\[
\rho_1 Z_1^{(1)} + \rho_2 Z_2^{(1)}
\]
\[
= \sum_{s=1}^2 \rho_s \left\{ \sum_{i:W_i=1,S_i=s} Y_i T_i - \frac{\sum_{i:W_i=1,S_i=s} Y_i (1 - T_i)}{\sum_{i:W_i=1,S_i=s} (1 - T_i)} \right\} \left\{ \frac{2(\sigma^2_{s0} + \sigma^2_{s1})}{\sum_{i=1}^n \mathbb{1}[W_i = 1, S_i = s]} \right\}^{-1/2}
\]
\[
= \sum_{s=1}^2 \rho_s \left\{ \frac{\sum_{i:W_i=1,S_i=s} Y_i T_i}{p_s \sum_{i:W_i=1} T_i} - \frac{\sum_{i:W_i=1,S_i=s} Y_i (1 - T_i)}{p_s \sum_{i:W_i=1} (1 - T_i)} \right\} \left( \frac{p_s(\sigma^2_{s0} + \sigma^2_{s1})}{\rho^2_s(n_1/2)} \right)^{-1/2}
\]
\[
= \sum_{s=1}^2 \left\{ \frac{\sum_{i:W_i=1,S_i=s} Y_i T_i}{\sum_{i:W_i=1} T_i} - \frac{\sum_{i:W_i=1,S_i=s} Y_i (1 - T_i)}{\sum_{i:W_i=1} (1 - T_i)} \right\} \left( \frac{\sum_{s'=1}^2 p_{s'}(\sigma^2_{s'0} + \sigma^2_{s'1})}{n_1/2} \right)^{-1/2}
\]
\[
= \left\{ \frac{\sum_{i:W_i=1} Y_i T_i}{\sum_{i:W_i=1} T_i} - \frac{\sum_{i:W_i=1} Y_i (1 - T_i)}{\sum_{i:W_i=1} (1 - T_i)} \right\} \left\{ \frac{1}{n_1^2} \sum_{s'=1}^2 2\sigma^2_{s'} \sum_{i=1}^n \mathbb{1}[W_i = 1, S_i = s'] \right\}^{-1/2}
\]
\[
= Z^{(1)},
\]
where (19) follows from (15) and $\sum_{i:W_i=1,S_i=s} \mathbb{1}[T_i = t] = p_s \sum_{i:W_i=1} \mathbb{1}[T_i = t]$ for each $s \in \{1, 2\}, t \in \{0, 1\}$; (20) follows from the definition (14) of $\rho_s$; and (21) follows from $\sum_{i=1}^n \mathbb{1}[W_i = 1, S_i = s] = n_1 p_s$, and $\sigma^2 = \sigma^2_{s0} + \sigma^2_{s1}$ by definition (from Section 3.4). We have shown (iii).
Part (iv) of Lemma 3 follows by analogous arguments as for parts (i)-(iii), except conditioned on the stage 1 data \(X^{(1)}\). To show part (v), for the remainder of the proof of this part of the lemma, we condition on \(X^{(1)}\) and \(D(X^{(1)}) = 1\). Then \(S_i = 1\) for each \(i : W_i = 2\), which implies 

\[
\sum_{i=1}^n 1[W_i = 2, S_i = 1] = \sum_{i=1}^n 1[W_i = 2] \quad \text{and} \quad \sum_{i=1}^n 1[W_i = 2, S_i = 2] = 0.
\]

Therefore, 

\[
\frac{1}{n^2} \sum_{s=1}^2 2\sigma_s^2 \sum_{i=1}^n 1[W_i = 2, S_i = s] = \frac{2\sigma_1^2}{n^2} \sum_{i=1}^n 1[W_i = 2, S_i = 1] = 2(\sigma_{10}^2 + \sigma_{11}^2).
\]

(22) 

\[
\frac{2(\sigma_{10}^2 + \sigma_{11}^2)}{n^2},
\]

(23)

where (22) follows from \(\sum_{i=1}^n 1[W_i = 2, S_i = 2] = 0\); (23) follows from 

\[
\sum_{i=1}^n 1[W_i = 2, S_i = 1] = n_2, \quad \text{and} \quad \sigma_s^2 = \sigma_{s0}^2 + \sigma_{s1}^2 \text{ (by definition)}.
\]

We then have 

\[
Z^{(2)}_1 = \left\{ \frac{\sum_{i:W_i=2,S_i=1} Y_i T_i}{\sum_{i:W_i=2,S_i=1} T_i} - \frac{\sum_{i:W_i=2,S_i=1} Y_i (1 - T_i)}{\sum_{i:W_i=2,S_i=1} (1 - T_i)} \right\} \frac{2(\sigma_{10}^2 + \sigma_{11}^2)}{n_2} \frac{1}{\sum_{i=1}^n 1[W_i = 2, S_i = 1]} - 1/2
\]

(24)

\[
\frac{\sum_{i:W_i=2,S_i=1} Y_i T_i}{\sum_{i:W_i=2,S_i=1} T_i} - \frac{\sum_{i:W_i=2,S_i=1} Y_i (1 - T_i)}{\sum_{i:W_i=2,S_i=1} (1 - T_i)} \right\} \frac{2(\sigma_{10}^2 + \sigma_{11}^2)}{n_2}
\]

(25)

\[
= Z^{(2)},
\]

where (24) follows from \(S_i = 1\) for each \(i : W_i = 2\) on the event \(D(X^{(1)}) = 1\); (25) follows from the equality of (22) and (23) shown above. This shows \(Z^{(2)} = Z^{(2)}_1\) on the event \(D(X^{(1)}) = 1\). It follows by analogous arguments as for parts (i)-(iii) that conditional on \(X^{(1)}\) and \(D(X^{(1)}) = 1\), we have \(Z^{(2)}_1\) is normally distributed with unit variance and mean 

\[
\left\{ \frac{(n_2/2)\mu_{11}}{n_2/2} - \frac{(n_2/2)\mu_{10}}{n_2/2} \right\} \frac{2(\sigma_{10}^2 + \sigma_{11}^2)}{n_2} \frac{1}{\sum_{i=1}^n 1[W_i = 2, S_i = 1]} - 1/2 = \Delta_1 [n_2/\{2(\sigma_{10}^2 + \sigma_{11}^2)\}]^{1/2},
\]

(26)

which follows from (24) and \(\sum_{i:W_i=2} 1[T_i = t] = \sum_{i=1}^n 1[W_i = 2]/2 = n_2/2\) for each \(t \in \{0, 1\}\). This shows part (v). Part (vi) follows directly from the definition of \(\rho_s\). Part (vii) follows from the assumptions in Section 3, which imply the distribution of the stage 1 data does not depend on which decision rule \(D \in \mathcal{D}\) is used (since the decision rule only affects the stage 2 data). This completes the proof of Lemma 3.

\[\Box\]

**Lemma 4.** When the combined population is enrolled in both stages, \(Z^{\text{TOTAL}}\) reduces to the following \(z\)-statistic that pools all participants in both stages and compares those assigned treatment \(t = 1\) versus control \(t = 0\):

\[
\left\{ \frac{\sum_{i=1}^n Y_i T_i}{\sum_{i=1}^n T_i} - \frac{\sum_{i=1}^n Y_i (1 - T_i)}{\sum_{i=1}^n (1 - T_i)} \right\} \left\{ \frac{2}{n} \sum_{s=1}^2 p_s (\sigma_{s0}^2 + \sigma_{s1}^2) \right\}^{1/2}
\]

(27)

Proof: Recall \(f_w = n_w/n, n_1 + n_2 = n, \text{ and } \sigma_s^2 = \sigma_{s0}^2 + \sigma_{s1}^2\). The condition of the lemma implies that \(\sum_{i=1}^n 1[W_i = w] = n_w \text{ and } \sum_{i=1}^n 1[W_i = w, S_i = s] = p_s n_w = p_s f_w n_w\) for each

Hosted by The Berkeley Electronic Press
\[ s \in \{1, 2\}, \, w \in \{1, 2\}. \text{ It follows that for each } w \in \{1, 2\}, \]
\[ \frac{1}{n_{w}^2} \sum_{s=1}^{2} 2\sigma_{s}^2 \sum_{i=1}^{n} 1[W_i = w, S_i = s] = 2 \sum_{s=1}^{2} p_s(\sigma_{s0}^2 + \sigma_{s1}^2)/(f_w n). \] (28)

We also have for each \( w \in \{1, 2\}, \)
\[ \sum_{i:W_i=w} T_i = \sum_{i:W_i=w} (1 - T_i) = f_w n/2 = f_w \sum_{i=1}^{n} T_i = f_w \sum_{i=1}^{n} (1 - T_i), \] (29)

which holds by the assumption in Section 3 that in each stage exactly half of participants enrolled in that stage are randomized to each study arm. We then have

\[
Z_{\text{TOTAL}} = f_{1}^{1/2} Z^{(1)} + f_{2}^{1/2} Z^{(2)}
\]
\[
= \sum_{w=1}^{2} f_{w}^{1/2} \left\{ \sum_{i:W_i=w} Y_i T_i - \sum_{i:W_i=w} Y_i (1 - T_i) \right\} \left( \frac{2}{f_{w} n} \sum_{s=1}^{2} p_s(\sigma_{s0}^2 + \sigma_{s1}^2) \right)^{-1/2}
\]
\[
= \left\{ \sum_{w=1}^{2} \sum_{i:W_i=w} Y_i T_i - \sum_{w=1}^{2} \sum_{i:W_i=w} Y_i (1 - T_i) \right\} \left( \frac{2}{n} \sum_{s=1}^{2} p_s(\sigma_{s0}^2 + \sigma_{s1}^2) \right)^{-1/2}
\]
\[
= \left\{ \sum_{i=1}^{n} Y_i T_i - \sum_{i=1}^{n} Y_i (1 - T_i) \right\} \left( \frac{2 n}{\sum_{s=1}^{2} p_s(\sigma_{s0}^2 + \sigma_{s1}^2)} \right)^{-1/2}, \]

where (30) follows from (28) and (29). This shows that when that the combined population is enrolled in both stages, \( Z_{\text{TOTAL}} \) equals (27). This proves Lemma 4.

We next prove that when the outcome variances equal a common value, the statistic \( Z_{\text{TOTAL}} \) reduces to a simple form.

**Lemma 5.** Consider any decision rule \( D \in \mathcal{D} \), and the data \( X \) generated in any adaptive design that follows rule \( D \). Assume \( \sigma_{st}^2 \) equals a common value \( \sigma^2 \) for each \( s \in \{1, 2\}, \, t \in \{0, 1\}. \) Then the following hold:

i. \( Z_{\text{TOTAL}} \) has a normal distribution with unit variance and mean \((p_1 \Delta_1 + p_2 \Delta_2)(f_1 n)^{1/2}/(2\sigma)\).

ii. \( Z^{(1)} \) has a normal distribution with conditional variance 1. The conditional mean \( E(Z^{(2)} \mid X^{(1)}) \) equals
Adaptive Enrichment Designs Not Dominated by Any Standard Design

\( (p_1 \Delta_1 + p_2 \Delta_2) (f_2 n)^{1/2} / (2 \sigma) \) if \( D(X^{(1)}) = C \), and equals \( \Delta_1 (f_2 n)^{1/2} / (2 \sigma) \) if \( D(X^{(1)}) = 1 \).

Proof: We have for each stage \( w \in \{1, 2\} \),

\[
\frac{1}{n_w^2} \sum_{s=1}^{2} \sum_{i=1}^{n} 1[W_i = w, S_i = s] = \frac{4\sigma^2}{n_w^2} \sum_{s=1}^{2} \sum_{i=1}^{n} 1[W_i = w, S_i = s] = 4\sigma^2 \frac{\sum_{i} 1[W_i = w]}{n_w} = 4\sigma^2 \frac{n_w}{n_w} = 4\sigma^2 \frac{n_w}{n_w},
\]

where (31) follows from \( \sigma^2 = \sigma^2_s + \sigma^2_t \) and the assumption in the lemma that \( \sigma^2_s = \sigma^2 \) for each \( s \in \{1, 2\} \), \( t \in \{0, 1\} \); (32) follows from

\[
\sum_{s=1}^{2} \sum_{i=1}^{n} 1[W_i = w, S_i = s] = \sum_{i=1}^{n} 1[W_i = w] = n_w = f_w n,
\]

since there are a total of \( f_w n \) participants in each stage \( w \in \{1, 2\} \), regardless of the decision \( D \) for stage 2 enrollment.

By the equality of (31) and (32), it follows that for each stage \( w \in \{1, 2\} \),

\[
Z(w) = \left\{ \frac{\sum_{i: W_i = w} Y_i T_i}{\sum_{i: W_i = w} T_i} - \frac{\sum_{i: W_i = w} Y_i (1 - T_i)}{\sum_{i: W_i = w} (1 - T_i)} \right\} \left\{ \frac{1}{n_w^2} \sum_{s=1}^{2} \sum_{i=1}^{n} 1[W_i = w, S_i = s] \right\}^{-1/2}
\]

\[
= \left\{ \frac{\sum_{i: W_i = w} Y_i T_i}{\sum_{i: W_i = w} T_i} - \frac{\sum_{i: W_i = w} Y_i (1 - T_i)}{\sum_{i: W_i = w} (1 - T_i)} \right\} \left\{ \frac{4\sigma^2}{n_w} \right\}^{-1/2}.
\]

We then have

\[
Z^{TOTAL} = f_1^{1/2} Z^{(1)} + f_2^{1/2} Z^{(2)}
\]

\[
= \sum_{w=1}^{2} f_w^{1/2} \left\{ \frac{\sum_{i: W_i = w} Y_i T_i}{\sum_{i: W_i = w} T_i} - \frac{\sum_{i: W_i = w} Y_i (1 - T_i)}{\sum_{i: W_i = w} (1 - T_i)} \right\} \left\{ \frac{4\sigma^2}{n_w} \right\}^{-1/2}
\]

\[
= \sum_{w=1}^{2} f_w^{1/2} \left\{ \frac{\sum_{i: W_i = w} Y_i T_i}{f_w \sum_{i=1}^{n} T_i} - \frac{\sum_{i: W_i = w} Y_i (1 - T_i)}{f_w \sum_{i=1}^{n} (1 - T_i)} \right\} \left\{ \frac{4\sigma^2}{n_f}\right\}^{-1/2}
\]

\[
= \sum_{w=1}^{2} \left\{ \frac{\sum_{i=1}^{n} Y_i T_i}{\sum_{i=1}^{n} T_i} - \frac{\sum_{i=1}^{n} Y_i (1 - T_i)}{\sum_{i=1}^{n} (1 - T_i)} \right\} \left\{ \frac{n^{1/2}}{2\sigma} \right\},
\]

where (34) follows from (33); and (35) follows from the fact that for each \( w \in \{1, 2\} \), \( t \in \{0, 1\} \),

\[
\sum_{i: W_i = w} 1[T_i = t] = f_w n/2 = f_w \sum_{i=1}^{n} 1[T_i = t].
\]

This proves claim (i) in Lemma 5.
Parts (ii) and (iii) of the lemma follow from Lemma 3 and using that \( \sigma^2_{st} = \sigma^2 \) implies \( \rho_s = p_s^{1/2} \) for each \( s \in \{1, 2\} \). This completes the proof of Lemma 5.

We next prove several relationships between \( Z^{\text{TOTAL}} \) and \( Z^s \), for each \( s \in \{1, 2\} \).

**Lemma 6.** When the combined population is enrolled in both stages,

\[
Z^{\text{TOTAL}} = p_1 Z_1^{\text{TOTAL}} + p_2 Z_2^{\text{TOTAL}}.
\]

**Proof:** Consider any standard design \( S \in \mathcal{S} \), where (38) follows from parts (ii) and (iv) of Lemma 3; (39) follows from (37) follows from parts (iii) and (iv) of Lemma 3; the last line above follows from the definition of \( Z^{\text{TOTAL}} \) in Section 3.4. This proves Lemma 6.

**Lemma 7.** For each \( s \in \{1, 2\} \), define \( \delta_s = \Delta_s \left[ np_s / \left\{ 2(\sigma_{s0}^2 + \sigma_{s1}^2) \right\} \right]^{1/2} \). For any standard design \( S \in \mathcal{S} \),

i. \( E \left[ Z_s^{\text{TOTAL}} \right] = \delta_s \);

ii. The correlation between \( Z^{\text{TOTAL}} \) and \( Z^s \) is \( \rho_s \) as defined in (14), for each \( s \in \{1, 2\} \).

iii. \( Z^{\text{TOTAL}} \) is normally distributed with unit variance and mean \( \rho_1 \delta_1 + \rho_2 \delta_2 \).

iv. \( E[Z^{(2)} | X^{(1)}] = \sum_{s=1}^{2} \rho_s \Delta_s \left[ np_s / \left\{ 2(\sigma_{s0}^2 + \sigma_{s1}^2) \right\} \right]^{1/2} \) with probability 1.

**Proof:** Consider any standard design \( S \in \mathcal{S} \), and any subpopulation \( s \in \{1, 2\} \). Since by assumption the combined population is always enrolled in both stages, we have

\[
E \left[ Z_s^{\text{TOTAL}} \right] = E f_1^{1/2} Z_1^{(1)} + E f_2^{1/2} Z_2^{(2)}
\]

\[
= f_1^{1/2} \Delta_s \left[ np_s / \left\{ 2(\sigma_{s0}^2 + \sigma_{s1}^2) \right\} \right]^{1/2} + f_2^{1/2} \Delta_s \left[ np_s / \left\{ 2(\sigma_{s0}^2 + \sigma_{s1}^2) \right\} \right]^{1/2}
\]

\[
= f_1 \Delta_s \left[ np_s / \left\{ 2(\sigma_{s0}^2 + \sigma_{s1}^2) \right\} \right]^{1/2} + f_2 \Delta_s \left[ np_s / \left\{ 2(\sigma_{s0}^2 + \sigma_{s1}^2) \right\} \right]^{1/2}
\]

\[
= \Delta_s \left[ np_s / \left\{ 2(\sigma_{s0}^2 + \sigma_{s1}^2) \right\} \right]^{1/2}, \tag{38}
\]

where (38) follows from parts (ii) and (iv) of Lemma 3; (39) follows from \( n_w = f_w n \) for each \( w \in \{1, 2\} \); (40) follows from \( f_1 + f_2 = 1 \). This shows part (i) of the lemma.

We next show \( Z_1^{\text{TOTAL}} \) and \( Z_2^{\text{TOTAL}} \) are independent. It was proved in part (i) of Lemma 3 that \( Z_1^{(1)} \) and \( Z_2^{(1)} \) are independent, each being normally distributed with unit variance. By part (iv) of Lemma 3 and using that \( D(X^{(1)}) = C \) with probability 1 for a standard design, conditioned on \( X^{(1)} \) we have \( Z_1^{(2)} \) and \( Z_2^{(2)} \) are independent, each being normally distributed with unit variance and mean a constant not depending on \( X^{(1)} \). Therefore, \( \left( Z_1^{(2)}, Z_2^{(2)} \right) \) is independent of \( X^{(1)} \). The previous two sentences imply \( Z_1^{(2)} \) and \( Z_2^{(2)} \) are (unconditionally) independent, each
normally distributed with unit variance. It follows that \((Z_1^{(1)}, Z_2^{(1)})\) (which is a function of \(X^{(1)}\)) is independent of \((Z_1^{(2)}, Z_2^{(2)})\).

The above arguments imply the variables \(\{Z_1^{(1)}, Z_2^{(1)}, Z_1^{(2)}, Z_2^{(2)}\}\) are mutually independent, each with unit variance. Therefore, \(Z_1^{\text{TOTAL}} = f_1^{1/2} Z_1^{(1)} + f_2^{1/2} Z_2^{(1)}\) and \(Z_2^{\text{TOTAL}} = f_1^{1/2} Z_1^{(2)} + f_2^{1/2} Z_2^{(2)}\) are independent, each being normally distributed with unit variance. By Lemma 6, the covariance of \(Z_1^{\text{TOTAL}}\) and \(Z_2^{\text{TOTAL}}\) equals \(\rho_s\), for each \(s \in \{1, 2\}\).

The independence of \(Z_1^{\text{TOTAL}}\) and \(Z_2^{\text{TOTAL}}\), combined with Lemma 6, implies \(Z_1^{\text{TOTAL}}\) is normally distributed with variance \(\rho_1^2 \text{VAR}(Z_1^{\text{TOTAL}}) + \rho_2^2 \text{VAR}(Z_2^{\text{TOTAL}})\). Since \(Z_1^{\text{TOTAL}}\) and \(Z_2^{\text{TOTAL}}\) each have unit variance, and since \(\rho_1^2 + \rho_2^2 = 1\) by part (vi) of Lemma 3, we have the variance of \(Z^{\text{TOTAL}}\) is 1. This proves part (ii) of Lemma 7. Part (iii) of the lemma follows by Lemma 6 and part (i) of Lemma 7.

We next show part (iv). Since \(S\) is assumed to be a standard design, we have the corresponding decision rule \(D\) satisfies \(D(X^{(1)}) = C\) for all possible values of \(X^{(1)}\). Thus, \(D(X^{(1)}) = C\) with probability 1. Therefore, by part (iv) of Lemma 3, we have that with probability 1,

\[
E[Z^{(2)}|X^{(1)}] = E[Z^{(2)}|X^{(1)}, D(X^{(1)}) = C] \\
= E[Z^{(2)}|D(X^{(1)}) = C] \\
= \sum_{s=1}^{2} \rho_s \Delta_s \left[ \frac{n_2 p_s}{2} \right]^{1/2} \left( \frac{\sigma_{s0}^2 + \sigma_{s1}^2}{np_s/2} \right)^{-1/2},
\]

where (41) follows from \(D(X^{(1)}) = C\) with probability 1; (42) follows from \(Z^{(2)}\) being independent of \(X^{(1)}\) conditioned on \(D(X^{(1)}) = C\), as shown in part (iv) of Lemma 3; (43) follows from the equality in part (iv.5) of Lemma 3. This shows part (iv), and completes the proof of Lemma 7.

**Lemma 8.** For \(n^{(\alpha, \beta)}\) as defined in Section 4, we have \(n^{(\alpha, \beta)}\) equals the expression (5).

Proof: Consider total sample size \(n\) equal to the expression (5). We will show the power of UMP\(_C\) to reject \(H_{0C}\) at \(\omega_C\) is \(1 - \beta\). Under the standard design UMP\(_C\), by Lemmas 6 and 7, we have

\[
E_{\omega_C} [Z_1^{\text{TOTAL}}] = \sum_{s=1}^{2} \rho_s E_{\omega_C} [Z_s^{\text{TOTAL}}] \\
= \sum_{s=1}^{2} \rho_s \Delta_{\text{min}} \left( \frac{\sigma_{s0}^2 + \sigma_{s1}^2}{np_s/2} \right)^{-1/2} \\
= \sum_{s=1}^{2} \rho_s \Delta_{\text{min}} \left( \frac{\Phi^{-1}(1 - \alpha) + \Phi^{-1}(1 - \beta)}{\sum_{s'=1}^{2} p_{s'} \sum_{t=0}^{1} 2\sigma_{st}^2} \right)^{1/2} \left( \frac{\sigma_{s0}^2 + \sigma_{s1}^2}{np_s/2} \right)^{-1/2} \\
= \sum_{s=1}^{2} \rho_s \left( \Phi^{-1}(1 - \alpha) + \Phi^{-1}(1 - \beta) \right) \left( \frac{\sigma_{s0}^2 + \sigma_{s1}^2}{np_s} \right)^{-1/2} \left( \frac{\sigma_{s0}^2 + \sigma_{s1}^2}{\sum_{s'=1}^{2} p_{s'} \sum_{t=0}^{1} \sigma_{st}^2} \right)^{-1/2}.
\]
\[\begin{align*} 
&= \sum_{s=1}^{2} \rho_s \left\{ \Phi^{-1}(1-\alpha) + \Phi^{-1}(1-\beta) \right\} \rho_s \left\{ \frac{p_s \left( \sigma_{s0}^2 + \sigma_{s1}^2 \right)}{\sum_{s'=1}^{2} p_{s'} \sum_{t=0}^{1} \sigma_{s't}} \right\}^{-1/2} \\
&= \sum_{s=1}^{2} \rho_s \left\{ \Phi^{-1}(1-\alpha) + \Phi^{-1}(1-\beta) \right\} \rho_s \left\{ \rho_s^2 \right\}^{-1/2} \\
&= \sum_{s=1}^{2} \rho_s \left\{ \Phi^{-1}(1-\alpha) + \Phi^{-1}(1-\beta) \right\} = \Phi^{-1}(1-\alpha) + \Phi^{-1}(1-\beta), 
\end{align*}\]

where (45) follows from the definition (14) of \(\rho_s\). Combining the above argument with part (iii) of Lemma 7, we have that under UMP\(_C\) and at the alternative \(\omega_C\), \(Z_{\text{TOTAL}}\) has normal distribution with mean (46) and unit variance. Therefore, the power of UMP\(_C\) to reject \(H_{0C}\) at \(\omega_C\) is

\[
P_{\omega_C} \left\{ Z_{\text{TOTAL}} > \Phi^{-1}(1-\alpha) \right\} = P_{\omega_C} \left\{ Z_{\text{TOTAL}} - E_{\omega_C} \left[ Z_{\text{TOTAL}} \right] > \Phi^{-1}(1-\alpha) - E_{\omega_C} \left[ Z_{\text{TOTAL}} \right] \right\} = \Phi \left( E_{\omega_C} \left[ Z_{\text{TOTAL}} \right] - \Phi^{-1}(1-\alpha) \right) = \Phi \left( \Phi^{-1}(1-\alpha) + \Phi^{-1}(1-\beta) - \Phi^{-1}(1-\alpha) \right) = 1 - \beta, 
\]

where (47) follows from the equality of (44) and (46). Therefore, at total sample size \(n\) equal to the expression (5), the power of UMP\(_C\) to reject \(H_{0C}\) at \(\omega_C\) is \(1 - \beta\). This proves Lemma 8. \(\square\)

We next show that in a special case, the log-likelihood ratio can be expressed as a simple function of \(Z_{\text{TOTAL}}\).

**Lemma 9.** Consider any \(D \in \mathcal{D}\), and assume the following hold:

a. \(\sigma_{st}^2\) equals a common value \(\sigma^2\) for each \(s \in \{1, 2\}, t \in \{0, 1\}\).
b. \(\mu^{(0)} = (\mu_{10}, \mu_{11}, \mu_{20}, \mu_{21}) = \Delta_{\text{min}}(1/2, 1/2, 1/2, 1/2)\) (which is in \(H_{0C} \cap H_{01}\)).
c. \(\mu^{(1)} = (\mu_{10}', \mu_{11}', \mu_{20}', \mu_{21}') = \Delta_{\text{min}}(0, 1, 0, 1)\) (which is the alternative \(\omega_C\)).

Then the log-likelihood ratio (13) equals \(\left\{ \frac{n^{1/2} \Delta_{\text{min}}/(2\sigma)}{Z_{\text{TOTAL}} - n \Delta_{\text{min}}^2/(8\sigma^2)} \right\} \). The most powerful test of the null hypothesis \(\mu^{(0)}\) versus \(\mu^{(1)}\) at level \(\alpha\) rejects \(\mu^{(0)}\) if and only if \(Z_{\text{TOTAL}} > \Phi^{-1}(1-\alpha)\), except on an event with probability 0. Furthermore, consider any test \(\phi\) of \(\mu^{(0)}\) at level \(\alpha\) with power at \(\mu^{(1)}\) equal to the most powerful test; then \(\phi\) rejects \(\mu^{(0)}\) if and only if \(Z_{\text{TOTAL}} > \Phi^{-1}(1-\alpha)\), except on an event with probability 0.

Proof: By (a) and the representation of \(Z_{\text{TOTAL}}\) proved in Lemma 5, we have

\[
Z_{\text{TOTAL}} = \left\{ \frac{\sum_{i=1}^{n} Y_i T_i}{\sum_{i=1}^{n} T_i} - \frac{\sum_{i=1}^{n} Y_i (1-T_i)}{\sum_{i=1}^{n} (1-T_i)} \right\} \left\{ \frac{n^{1/2}}{2\sigma} \right\}. 
\]

By assumptions (b) and (c), for each \(s \in \{1, 2\}, t \in \{0, 1\}\), we have

\[
\mu_{st}^{(1)} - \mu_{st}^{(0)} = (2t-1)(\Delta_{\text{min}}/2) = \{t - (1-t)\}(\Delta_{\text{min}}/2), 
\]
and \((\mu_{st}(1))^2 - (\mu_{st}(0))^2 = (t - 1/4)A_{\min}^2\). By (13), the log-likelihood ratio comparing \(\mu^{(1)}\) to \(\mu^{(0)}\) is

\[
\log \frac{L(\mu^{(1)}; X)}{L(\mu^{(0)}; X)} = \sum_{i=1}^{n} \left\{ \frac{Y_i \mu^{(1)}_{S_i T_i} - \mu^{(0)}_{S_i T_i}}{\sigma^2_{S_i T_i}} - \frac{(\mu^{(1)}_{S_i T_i} - \mu^{(0)}_{S_i T_i})^2}{2\sigma^2_{S_i T_i}} \right\}
\]

\[
= \frac{1}{\sigma^2} \sum_{i=1}^{n} \left\{ \{Y_i T_i - Y_i(1 - T_i)\} \frac{\Delta_{\min}}{2} - (T_i - 1/4)A_{\min}^2 \right\}
\]

\[
= \frac{\Delta_{\min}}{2\sigma^2} \left\{ \sum_{i=1}^{n} Y_i T_i - \sum_{i=1}^{n} Y_i(1 - T_i) \right\} - \left\{ \sum_{i=1}^{n} (T_i - 1/4) \right\} \frac{A_{\min}^2}{2\sigma^2}
\]

\[
= \frac{n^{1/2}\Delta_{\min}}{2\sigma} \left\{ \sum_{i=1}^{n} Y_i T_i - \sum_{i=1}^{n} Y_i(1 - T_i) \right\} - \left\{ \sum_{i=1}^{n} Y_i(1 - T_i) \right\} \left\{ \frac{n^{1/2}}{2\sigma} \right\} - \frac{nA_{\min}^2}{8\sigma^2}
\]

(48)

where (48) follows from \(\sum_{i=1}^{n} T_i = n/2\), and (49) follows from part (i) of Lemma 5. Recall we assumed \(n\), \(\Delta_{\min}\) and \(\sigma\) are known. By (49) and the Neyman–Pearson Lemma in Theorem 3.2.1 of (Lehmann and Romano, 2005), the most powerful test at level \(\alpha\) of the null hypothesis \(\mu^{(0)}\) versus the alternative \(\mu^{(1)}\) rejects \(\mu^{(0)}\) if and only if \(Z_{\text{TOTAL}} > k\), except on an event with probability 0, where \(k\) is a constant for which the Type I error at \(\mu^{(0)}\) equals \(\alpha\); furthermore, any test of \(\mu^{(0)}\) at level \(\alpha\) with power equal to the most powerful test must reject \(\mu^{(0)}\) if and only if \(Z_{\text{TOTAL}} > k\), except on an event with probability 0. Since \(Z_{\text{TOTAL}}\) has a standard normal distribution under \(\mu^{(0)}\), we must have \(k = \Phi^{-1}(1 - \alpha)\). This proves Lemma 9. \(\square\)

### D. Proof of Theorem 1

Proof: By the condition in the theorem that \(n = n^{(\alpha, \beta)}\), the power of UMP\(C\) to reject \(H_{0C}\) at \(\omega_C\) equals \(1 - \beta\). Consider any standard design \(S \in S\) satisfying the condition of the theorem. We show \(S\) must reject at least one null hypothesis if and only if \(Z_{\text{TOTAL}} > \Phi^{-1}(1 - \alpha)\), except on an event with probability zero. Let \(\phi(X)\) denote the indicator variable taking value 1 if \(S\) rejects at least one null hypothesis and 0 otherwise. By the condition of the theorem that \(S\) strongly controls the familywise Type I error rate at level \(\alpha\), we have \(\phi(X)\) is a valid test of the null hypothesis \(H_{0C} \cap H_{01}\) at level \(\alpha\). This follows since under any \(\mu \in (H_{0C} \cap H_{01})\), the probability that \(S\) rejects one or more null hypotheses is at most \(\alpha\), and so \(\phi(X) = 1\) with probability at most \(\alpha\). By the condition of the theorem that \(S\) has overall power at least \(1 - \beta\) at \(\omega_C\), the probability of \(\phi(X) = 1\) at \(\omega_C\) is at least \(1 - \beta\).

Consider the simple null hypothesis:

\[
\mu^{(0)} = (\mu_{01}, \mu_{10}, \mu_{20}, \mu_{21}) = \Delta_{\min}(1/2, 1/2, 1/2, 1/2),
\]

which is in \(H_{0C} \cap H_{01}\). Consider the simple alternative

\[
\mu^{(1)} = (\mu_{01}^{(1)}, \mu_{10}^{(1)}, \mu_{20}^{(1)}, \mu_{21}^{(1)}) = \Delta_{\min}(0, 1, 0, 1),
\]
which equals \( \omega_C \). By Lemma 9, the most powerful test at level \( \alpha \) of the null hypothesis \( \mu(0) \) versus the alternative \( \mu(1) \) rejects \( \mu(0) \) if and only if \( Z^{\text{TOTAL}} > \Phi^{-1}(1 - \alpha) \), except on an event with probability 0; furthermore, any test of \( \mu(0) \) at level \( \alpha \) with power at \( \mu(1) \) equal to the most powerful test must reject \( \mu(0) \) if and only if \( Z^{\text{TOTAL}} > \Phi^{-1}(1 - \alpha) \), except on an event with probability 0. By the condition in the theorem that \( n = n^{(\alpha, \beta)} \), the most powerful test at level \( \alpha \) has power \( 1 - \beta \) at \( \omega_C \). Since \( \phi(X) \) is a valid test at level \( \alpha \) of \( \mu(0) \), and has power at least \( 1 - \beta \) at \( \mu(1) \), by the above argument it must reject \( \mu(0) \) if and only if \( Z^{\text{TOTAL}} > \Phi^{-1}(1 - \alpha) \), except on an event with probability zero. This shows \( S \) rejects at least one null hypothesis if and only if \( Z^{\text{TOTAL}} > \Phi^{-1}(1 - \alpha) \), except on an event with probability zero.

We next show that under \( \omega_1 \), the probability of \( Z^{\text{TOTAL}} > \Phi^{-1}(1 - \alpha) \) equals

\[
\Phi(p_1 \Phi^{-1}(1 - \beta) - p_2 \Phi^{-1}(1 - \alpha)).
\]

By part (iii) of Lemma 7, \( Z^{\text{TOTAL}} \) has a normal distribution with unit variance. It remains to compute its mean.

Define \( \delta_{s, \min} = \Delta_{\min} \left[ np_s / \left\{ 2(\sigma_{s0}^2 + \sigma_{s1}^2) \right\} \right]^{1/2} \) for each \( s \in \{1, 2\} \). Since \( n = n^{(\alpha, \beta)} \), we have

\[
\delta_{s, \min} = \Delta_{\min} \left[ n^{(\alpha, \beta)} p_s / \left\{ 2(\sigma_{s0}^2 + \sigma_{s1}^2) \right\} \right]^{1/2}
\]

(50)

\[
= \Delta_{\min} \left\{ \Phi^{-1}(1 - \alpha) + \Phi^{-1}(1 - \beta) \right\}^2 \left\{ \sum_{s' = 1}^2 p_{s'} \sum_{t = 0}^1 2\sigma_{s't}^2 \right\} \left( \frac{\Delta_{\min}^2 p_s}{2(\sigma_{s0}^2 + \sigma_{s1}^2)} \right)^{1/2}
\]

(51)

\[
= \left\{ \Phi^{-1}(1 - \alpha) + \Phi^{-1}(1 - \beta) \right\} p_s \left\{ \sum_{s' = 1}^2 p_{s'} \sum_{t = 0}^1 \sigma_{s't}^2 \right\} \left( \frac{1}{p_s(\sigma_{s0}^2 + \sigma_{s1}^2)} \right)^{1/2}
\]

(52)

where (51) follows from (5) and (52) follows from (14). By Lemma 7, we have \( E_{\omega_1} Z_1^{\text{TOTAL}} = \delta_{1, \min} \) and \( E_{\omega_1} Z_2^{\text{TOTAL}} = 0 \). Then by Lemma 6, we have

\[
E_{\omega_1} Z^{\text{TOTAL}} = E_{\omega_1} \left( \rho_1 Z_1^{\text{TOTAL}} + \rho_2 Z_2^{\text{TOTAL}} \right) = \rho_1 \delta_{1, \min} = \rho_1 \left\{ \Phi^{-1}(1 - \alpha) + \Phi^{-1}(1 - \beta) \right\}
\]

where the last equality follows from (52). Then the probability that \( S \) rejects at least one null hypothesis under \( \omega_1 \) equals

\[
P_{\omega_1} \left\{ Z^{\text{TOTAL}} > \Phi^{-1}(1 - \alpha) \right\} = P_{\omega_1} \left\{ Z^{\text{TOTAL}} - E_{\omega_1} Z^{\text{TOTAL}} > \Phi^{-1}(1 - \alpha) - E_{\omega_1} Z^{\text{TOTAL}} \right\}
\]

\[
= \Phi \left( E_{\omega_1} Z^{\text{TOTAL}} - \Phi^{-1}(1 - \alpha) \right)
\]

\[
= \Phi \left( \rho_1 \left\{ \Phi^{-1}(1 - \alpha) + \Phi^{-1}(1 - \beta) \right\} - \Phi^{-1}(1 - \alpha) \right)
\]

\[
= \Phi \left( \rho_1 \Phi^{-1}(1 - \beta) - (1 - \rho_1) \Phi^{-1}(1 - \alpha) \right)
\]

(53)

We have shown that the standard design \( S \) has overall power (53) at \( \omega_1 \). The above chain of equalities also shows that at \( \omega_1 \), the overall power of UMP \( C \) is (53). This proves Theorem 1. □

E. PROOF OF THEOREM 2

We prove a slightly more general version of Theorem 2, where we replace the assumption \( \sigma_{s1}^2 = \sigma^2 \) for each \( s \in \{1, 2\}, t \in \{0, 1\} \) by the weaker assumption \( \sigma_{s1}^2 + \sigma_{s0}^2 \leq \sigma_{s1}^2 + \sigma_{s0}^2 \).
Adaptive Enrichment Designs Not Dominated by Any Standard Design

**Theorem 4.** Let $\alpha = 0.05$ and consider any $\beta \in (0, 1)$. Assume $\sigma_{11}^2 + \sigma_{10}^2 \leq \sigma_{21}^2 + \sigma_{20}^2$, and let $n = n^{(\alpha, \beta)}$. Consider any $p_1 \in (0, 1)$. Each adaptive enrichment design $A_{\theta, f_1} \in A_*$ has all of the following properties:

i. It strongly controls the familywise Type I error rate at level $\alpha = 0.05$.

ii. At alternative $\omega_C$, it has overall power equal to that of $UMP_C$.

iii. At alternative $\omega_1$, it has greater overall power than $UMP_C$.

iv. If $\sigma_s^2 = \sigma^2$ for each $s \in \{1, 2\}, t \in \{0, 1\}$, then no standard design $S \in S$ simultaneously dominates its overall power at $\omega_C$ and at $\omega_1$.

Proof: Let $\alpha = 0.05$ and $n = n^{(\alpha_1, \beta_1)}$. Consider any $p_1 \in (0, 1), \theta \in \mathbb{R}$, and $f_1 \in [0.05, 0.95]$; we will show the design $A_{\theta, f_1} \in A_*$ has properties (i)-(iv). It was proved in Section C of the Supplementary Materials of (Rosenblum and van der Laan, 2011) that for each adaptive design in a class that includes $A_*$, the familywise Type I error rate is strongly controlled at level $\alpha = 0.05$. This shows part (i).

Let $D_\theta$ denote the decision rule (defined in Section 5) corresponding to $A_{\theta, f_1}$. Let $D_C$ denote the decision rule corresponding to the standard design $UMP_C$, which always enrolls the combined population in stage 2. We use the notation from Section A, where for example $P_{D, \omega_C}$ and $E_{D, \omega_C}$ denote probability and expectation, respectively, where $X$ is generated under decision rule $D$ at alternative $\omega_C$. We suppress $f_1$ in our notation.

Let $F_C$ denote the event $D_\theta(X^{(1)}) = C$ and let $F_1$ denote the event $D_\theta(X^{(1)}) = 1$. By part (vii) of Lemma 3, for any decision rules $D_1, D_2 \in D$ and any alternative $\omega \in \{\omega_C, \omega_1\}$,

$$P_{D_1, \omega} \{F_C\} = P_{D_2, \omega} \{F_C\} \text{ and } P_{D_1, \omega} \{F_1\} = P_{D_2, \omega} \{F_1\}. \quad (54)$$

We next prove the following relationships between the conditional mean of the second stage statistic $Z^{(2)}$ when $X$ is generated under $D_\theta$ versus when $X$ is generated under $D_C$:

**Lemma 10.** For any $\theta \in \mathbb{R}$,

$$E_{D_\theta, \omega_C}(Z^{(2)} | F_C) = E_{D_C, \omega_C}(Z^{(2)} | F_C); \quad (55)$$

$$E_{D_\theta, \omega_C}(Z^{(2)} | F_1) \geq E_{D_C, \omega_C}(Z^{(2)} | F_1); \quad (56)$$

$$E_{D_\theta, \omega_1}(Z^{(2)} | F_C) = E_{D_C, \omega_1}(Z^{(2)} | F_C); \quad (57)$$

$$E_{D_\theta, \omega_1}(Z^{(2)} | F_1) > E_{D_C, \omega_1}(Z^{(2)} | F_1). \quad (58)$$

Proof of Lemma 10: By Lemma 3 we have

$$E_{D_\theta, \omega_C}(Z^{(2)} | F_C) = \sum_{s=1}^{2} p_s \Delta_{\min} \left[ n_2 p_s / \left\{ 2(\sigma_s^2 + \sigma_{s1}^2) \right\} \right]^{1/2}; \quad (59)$$

$$E_{D_\theta, \omega_C}(Z^{(2)} | F_1) = \Delta_{\min} \left[ n_2 / \left\{ 2(\sigma_{10}^2 + \sigma_{11}^2) \right\} \right]^{1/2}. \quad (60)$$
By part (iv) of Lemma 7, we have

$$E_{D_{C, \omega_C}}(Z^{(2)} \mid X^{(1)}) = \sum_{s=1}^{2} \rho_s \Delta_{\min} \left[ n_2 p_s / \left\{ 2(\sigma_{s0}^2 + \sigma_{s1}^2) \right\} \right]^{1/2}$$

(61)

$$= \sum_{s=1}^{2} \Delta_{\min} n_2 / \left\{ 2(p_1(\sigma_{10}^2 + \sigma_{11}^2) + p_2(\sigma_{20}^2 + \sigma_{21}^2)) \right\}^{1/2}$$

(62)

$$= \Delta_{\min} [n_2 / \left\{ 2(p_1(\sigma_{10}^2 + \sigma_{11}^2) + p_2(\sigma_{20}^2 + \sigma_{21}^2)) \right\}]^{1/2}$$

(63)

$$\leq \Delta_{\min} [n_2 / \left\{ 2(\sigma_{10}^2 + \sigma_{11}^2) \right\}]^{1/2}.$$  

(64)

where (62) follows from (14); (63) follows from $p_1 + p_2 = 1$; (64) follows from the assumption above that $\sigma_{11}^2 + \sigma_{10}^2 \leq \sigma_{21}^2 + \sigma_{20}^2$. We have shown $E_{D_{C, \omega_C}}(Z^{(2)} \mid X^{(1)})$ equals (63), and therefore does not depend on $X^{(1)}$. It then follows from (61)-(64) that

$$E_{D_{C, \omega_C}}(Z^{(2)} \mid F_C) = E_{D_{C, \omega_C}}(Z^{(2)} \mid F_1)$$

(65)

$$= \sum_{s=1}^{2} \rho_s \Delta_{\min} \left[ n_2 p_s / \left\{ 2(\sigma_{s0}^2 + \sigma_{s1}^2) \right\} \right]^{1/2}$$

(66)

$$\leq \Delta_{\min} [n_2 / \left\{ 2(\sigma_{10}^2 + \sigma_{11}^2) \right\}]^{1/2}.$$  

(67)

Equation (55) then follows from (59) and the equality of (65) and (66). Inequality (56) follows from (60) and (65)-(67).

We next show (57) and (58). By Lemma 3, we have

$$E_{D_{\theta, \omega_1}}(Z^{(2)} \mid F_C) = \rho_1 \Delta_{\min} [n_2 p_1 / \left\{ 2(\sigma_{10}^2 + \sigma_{11}^2) \right\}]^{1/2};$$

(68)

$$E_{D_{\theta, \omega_1}}(Z^{(2)} \mid F_1) = \Delta_{\min} [n_2 / \left\{ 2(\sigma_{10}^2 + \sigma_{11}^2) \right\}]^{1/2}.$$  

(69)

By part (iv) of Lemma 7 we have

$$E_{D_{C, \omega_1}}(Z^{(2)} \mid X^{(1)}) = \rho_1 \Delta_{\min} [n_2 p_1 / \left\{ 2(\sigma_{10}^2 + \sigma_{11}^2) \right\}]^{1/2}$$

(70)

$$< \Delta_{\min} [n_2 / \left\{ 2(\sigma_{10}^2 + \sigma_{11}^2) \right\}]^{1/2},$$  

(71)

where (71) follows from $0 < p_1 < 1$ and $0 < p_1 < 1$ which implies $\rho_1(p_1^{1/2}) < 1$. It follows from the above display that

$$E_{D_{C, \omega_1}}(Z^{(2)} \mid F_C) = E_{D_{C, \omega_1}}(Z^{(2)} \mid F_1)$$

(72)

$$\rho_1 \Delta_{\min} [n_2 p_1 / \left\{ 2(\sigma_{10}^2 + \sigma_{11}^2) \right\}]^{1/2}$$

(73)

$$< \Delta_{\min} [n_2 / \left\{ 2(\sigma_{10}^2 + \sigma_{11}^2) \right\}]^{1/2}.$$  

(74)

Equation (57) follows from (68) and the equality of (72) and (73). Inequality (58) follows from (69) and (72)-(74). This proves Lemma 10.

We next use Lemma 10 to show the following relationships, where we let $z_{1-\alpha} = \Phi^{-1}(1 - \alpha)$:
LEMMA 11. For any \( \theta \in \mathbb{R} \),

\[
P_{D_{\theta},\omega_C} \{ Z_{\text{TOTAL}} > z_{1-\alpha} \mid F_C \} = P_{D_{C},\omega_C} \{ Z_{\text{TOTAL}} > z_{1-\alpha} \mid F_C \}; \quad (75)
\]

\[
P_{D_{\theta},\omega_C} \{ Z_{\text{TOTAL}} > z_{1-\alpha} \mid F_1 \} \geq P_{D_{C},\omega_C} \{ Z_{\text{TOTAL}} > z_{1-\alpha} \mid F_1 \}; \quad (76)
\]

\[
P_{D_{\theta},\omega_1} \{ Z_{\text{TOTAL}} > z_{1-\alpha} \mid F_C \} = P_{D_{C},\omega_1} \{ Z_{\text{TOTAL}} > z_{1-\alpha} \mid F_C \}; \quad (77)
\]

\[
P_{D_{\theta},\omega_1} \{ Z_{\text{TOTAL}} > z_{1-\alpha} \mid F_1 \} > P_{D_{C},\omega_1} \{ Z_{\text{TOTAL}} > z_{1-\alpha} \mid F_1 \}. \quad (78)
\]

Proof of Lemma 11: Let \( U \) denote a standard normal random variable independent of \( X \). We have (where in lines (82) and (83) the expressions that differ from the previous line are in boldface)

\[
P_{D_{\theta},\omega_C} \{ Z_{\text{TOTAL}} > z_{1-\alpha} \mid F_C \} \quad (79)
\]

\[
P_{D_{\theta},\omega_C} \{ f_1^{1/2} Z(1) + f_2^{1/2} Z(2) > z_{1-\alpha} \mid F_C \} \quad (80)
\]

\[
P_{D_{\theta},\omega_C} \{ f_1^{1/2} Z(1) + f_2^{1/2} \left[ U + E_{D_{\theta},\omega_C}(Z(2) \mid F_C) \right] > z_{1-\alpha} \mid F_C \} \quad (81)
\]

\[
P_{D_{\theta},\omega_C} \{ f_1^{1/2} Z(1) + f_2^{1/2} \left[ U + E_{D_{\theta},\omega_C}(Z(2) \mid F_C) \right] > z_{1-\alpha} \mid F_C \} \quad (82)
\]

\[
P_{D_{C},\omega_C} \{ f_1^{1/2} Z(1) + f_2^{1/2} \left[ U + E_{D_{C},\omega_C}(Z(2) \mid F_C) \right] > z_{1-\alpha} \mid F_C \} \quad (83)
\]

\[
P_{D_{C},\omega_C} \{ f_1^{1/2} Z(1) + f_2^{1/2} Z(2) > z_{1-\alpha} \mid F_C \} \quad (84)
\]

\[
P_{D_{C},\omega_C} \{ Z_{\text{TOTAL}} > z_{1-\alpha} \mid F_C \}; \quad (85)
\]

where (80) follows from the definition of \( Z_{\text{TOTAL}} \); (81) follows from part (iv) of Lemma 3, which implies that under \( P_{D_{\theta},\omega_C} \), conditioned on \( F_C \) the statistic \( Z(2) \) is independent of \( X(1) \) (and so independent of \( Z(1) \)) and has normal distribution with mean \( E_{D_{\theta},\omega_C}(Z(2) \mid F_C) \); (82) follows from (55); (83) follows by part (vii) of Lemma 3 and that \( U \) is independent of \( X \), and \( U \) has the same distribution (i.e., standard normal) under \( P_{D_{\theta},\omega_C} \) for any \( D \in \mathcal{D} \); (84) follows from part (iv) of Lemma 3, which implies that under \( P_{D_{C},\omega_C} \), conditioned on \( F_C \) the statistic \( Z(2) \) is independent of \( X(1) \) (and so independent of \( Z(1) \)) and has normal distribution with mean \( E_{D_{C},\omega_C}(Z(2) \mid F_C) \); (85) follows from the definition of \( Z_{\text{TOTAL}} \). The equality of (79) and (85) proves (75). The equality (77) follows by an analogous argument as above, where the only differences are that \( \omega_C \) is replaced by \( \omega_1 \) in (79)-(85).

To show (76), we have

\[
P_{D_{\theta},\omega_C} \{ Z_{\text{TOTAL}} > z_{1-\alpha} \mid F_1 \} \quad (86)
\]

\[
P_{D_{\theta},\omega_C} \{ f_1^{1/2} Z(1) + f_2^{1/2} Z(2) > z_{1-\alpha} \mid F_1 \} \quad (87)
\]

\[
P_{D_{\theta},\omega_C} \{ f_1^{1/2} Z(1) + f_2^{1/2} \left[ U + E_{D_{\theta},\omega_C}(Z(2) \mid F_1) \right] > z_{1-\alpha} \mid F_1 \} \quad (88)
\]

\[
P_{D_{\theta},\omega_C} \{ f_1^{1/2} Z(1) + f_2^{1/2} \left[ U + E_{D_{\theta},\omega_C}(Z(2) \mid F_1) \right] > z_{1-\alpha} \mid F_1 \} \quad (89)
\]

\[
P_{D_{C},\omega_C} \{ f_1^{1/2} Z(1) + f_2^{1/2} \left[ U + E_{D_{C},\omega_C}(Z(2) \mid F_1) \right] > z_{1-\alpha} \mid F_1 \} \quad (90)
\]

\[
P_{D_{C},\omega_C} \{ f_1^{1/2} Z(1) + f_2^{1/2} Z(2) > z_{1-\alpha} \mid F_1 \} \quad (91)
\]

\[
P_{D_{C},\omega_C} \{ Z_{\text{TOTAL}} > z_{1-\alpha} \mid F_1 \}; \quad (92)
\]

where (87) follows from the definition of \( Z_{\text{TOTAL}} \); (88) follows from part (v) of Lemma 3, which implies that conditioned on \( F_1 \), the statistic \( Z(2) \) is independent of \( X(1) \) (and so independent
of $Z^{(1)}$ and has normal distribution with mean $ED_\theta,\omega_C(Z^{(2)} \mid F_1)$; (89) follows from (56); (90) follows by part (vii) of Lemma 3 and that $U$ is independent of $X$, and $U$ has the same distribution (i.e., standard normal) under $P_{D,\omega_C}$ for any $D \in \mathcal{D}$; (91) follows from part (v) of Lemma 3; (92) follows from the definition of $Z^{\text{TOTAL}}$. The inequality shown above regarding (86) and (92) proves (76). The inequality (78) follows by an analogous argument as above, where the only differences are that $\omega_C$ is replaced by $\omega_1$ in (86)-(92), and the $\geq$ is replaced by $>$ in (89) and is justified by the strict inequality (58). This completes the proof of Lemma 11.

We next compute the overall power of $A_{\theta,f_1}$ at $\omega_C$. Since both null hypotheses $H_{0C}$ and $H_{01}$ are false at $\omega_C$, the overall power at $\omega_C$ equals the probability of rejecting one or more of these null hypotheses. This occurs for $A_{\theta,f_1}$ if and only if $Z^{\text{TOTAL}} > z_{1-\alpha}$. We have

$$P_{D_\theta,\omega_C} \{ Z^{\text{TOTAL}} > z_{1-\alpha} \} = P_{D_\theta,\omega_C} \{ Z^{\text{TOTAL}} > z_{1-\alpha} \mid F_C \} P_{D_\theta,\omega_C} \{ F_C \}$$

$$+ P_{D_\theta,\omega_C} \{ Z^{\text{TOTAL}} > z_{1-\alpha} \mid F_1 \} P_{D_\theta,\omega_C} \{ F_1 \}$$

$$\geq P_{D_C,\omega_C} \{ Z^{\text{TOTAL}} > z_{1-\alpha} \mid F_C \} P_{D_C,\omega_C} \{ F_C \}$$

$$+ P_{D_C,\omega_C} \{ Z^{\text{TOTAL}} > z_{1-\alpha} \mid F_1 \} P_{D_C,\omega_C} \{ F_1 \}$$

$$= P_{D_C,\omega_C} \{ Z^{\text{TOTAL}} > z_{1-\alpha} \},$$

where (94) follows from (54), (75), and (76). This proves part (ii) of Theorem 4.

To prove part (iii) of Theorem 4, we compute the overall power of $A_{\theta,f_1}$ at $\omega_1$. Since both null hypotheses $H_{0C}$ and $H_{01}$ are false at $\omega_1$, the overall power at $\omega_1$ equals the probability of rejecting one or more of these null hypotheses. This occurs for $A_{\theta,f_1}$ if and only if $Z^{\text{TOTAL}} > z_{1-\alpha}$. We have

$$P_{D_\theta,\omega_1} \{ Z^{\text{TOTAL}} > z_{1-\alpha} \} = P_{D_\theta,\omega_1} \{ Z^{\text{TOTAL}} > z_{1-\alpha} \mid F_C \} P_{D_\theta,\omega_1} \{ F_C \}$$

$$+ P_{D_\theta,\omega_1} \{ Z^{\text{TOTAL}} > z_{1-\alpha} \mid F_1 \} P_{D_\theta,\omega_1} \{ F_1 \}$$

$$> P_{D_C,\omega_1} \{ Z^{\text{TOTAL}} > z_{1-\alpha} \mid F_C \} P_{D_C,\omega_1} \{ F_C \}$$

$$+ P_{D_C,\omega_1} \{ Z^{\text{TOTAL}} > z_{1-\alpha} \mid F_1 \} P_{D_C,\omega_1} \{ F_1 \}$$

$$= P_{D_C,\omega_1} \{ Z^{\text{TOTAL}} > z_{1-\alpha} \},$$

where (97) follows from (54), (77), (78), and $P_{D_\theta,\omega_1} \{ F_1 \} > 0$. This proves part (iii) of Theorem 4.

Part (iv) follows by parts (i)-(iii) combined with Theorem 1. This completes the proof of Theorem 4, which implies Theorem 2.

\[ \square \]

**F. PROOF OF THEOREM 3**

Proof: Consider any $n > 0$. The key step is deriving the likelihood ratio test corresponding to any adaptive design $A \in \mathcal{A}$, which was done in Lemmas 1 and 9.

Consider any $A = (D, M, f_1) \in \mathcal{A}$, and the corresponding experiment where $X$ is generated using this design. Assume the condition of the theorem that each $\sigma^2 = \sigma^2$. Let $\mu^{(1)}$ equal $\omega_C$.

Define $\mu^{(0)} = 0.5(\Delta_{\text{min}}, \Delta_{\text{min}}, \Delta_{\text{min}}, \Delta_{\text{min}})$, which is in $H_{01} \cap H_{0C}$. Then by Lemma 9, the most powerful test of the null hypothesis $\mu^{(0)}$ versus $\mu^{(1)}$ at level $\alpha$ rejects $\mu^{(0)}$ if and only if $Z^{\text{TOTAL}} > \Phi^{-1}(1-\alpha)$, except on an event with probability 0. Denote the corresponding power at $\mu^{(1)}$ by $1 - \beta$. Consider any test $\phi$ of $\mu^{(0)}$ at level $\alpha$ with power at $\mu^{(1)}$ equal to the most powerful test. Then Lemma 9 implies that $\phi$ rejects $\mu^{(0)}$ if and only if $Z^{\text{TOTAL}} > \Phi^{-1}(1-\alpha)$, except on an event with probability 0.
We next show the power of UMP_C to reject \( H_{0C} \) at \( \omega_C \) equals 1 - \( \beta \), i.e., that \( P_{\omega_C, \text{UMP}_C} (\text{UMP}_C \text{ rejects at least } H_{0C}) = 1 - \beta \). This follows from parts (ii) and (iii) of Lemma 5, using that at \( \omega_C \) we have \( p_1 \Delta_1 + p_2 \Delta_2 = p_1 \Delta_{\min} + p_2 \Delta_{\min} = \Delta_{\min} \).

Assume, for the sake of contradiction, that \( A \) has greater overall power than UMP_C at \( \omega_1 \), and greater or equal power as UMP_C to reject \( H_{0C} \) at \( \omega_C \). By definition, any \( A \in A \) strongly controls the familywise Type I error rate at level \( \alpha \). Let \( \phi(X) \) denote the indicator variable taking value 1 if \( A \) rejects \( H_{0C} \), and 0 otherwise. Then \( \phi(X) \) is a valid test at level \( \alpha \) of \( \mu^{(0)} \) versus the alternative \( \mu^{(1)} \), with power at least \( 1 - \beta \) at \( \mu^{(1)} \). Therefore, \( \phi(X) \) is a test of \( \mu^{(0)} \) at level \( \alpha \) with power at \( \mu^{(1)} \) at least that of the most powerful test. By the above arguments, \( \phi(X) = 1 \) if and only if \( Z^{\text{TOTAL}} > \Phi^{-1}(1 - \alpha) \) (except on an event with probability 0, which we ignore below). Therefore \( A \) must reject \( H_{0C} \) if and only if \( Z^{\text{TOTAL}} > \Phi^{-1}(1 - \alpha) \).

Define the event \( E \) to be that \( A \) rejects at least one null hypotheses and \( Z^{\text{TOTAL}} \leq \Phi^{-1}(1 - \alpha) \). We show the probability of \( E \) equals 0, at \( \mu^{(0)} \). The proof is by contradiction. At \( \mu^{(0)} \), the z-statistic \( Z^{\text{TOTAL}} \) has a standard normal distribution (which follows from Lemma 5), and so the probability of \( Z^{\text{TOTAL}} > \Phi^{-1}(1 - \alpha) \) is exactly \( \alpha \). If the probability of \( E \) were greater than zero, then by the above arguments, the familywise Type I error rate of \( A \) would exceed \( \alpha \), which would contradict the assumption that \( A \) strongly controls the familywise Type I error rate at level \( \alpha \). Therefore, \( E \) has probability zero at \( \mu^{(0)} \). This implies \( E \) has probability zero at any \( \mu \in \mathbb{R}^4 \), by Lemma 2.

We next show \( D(X^{(1)}) = C \) with probability 1. Assume for the sake of contradiction that \( D(X^{(1)}) = 1 \) with nonzero probability. Consider the parameter values \( \mu^{(0)'} = (\mu^{(0)'_1, \mu^{(0)'_2, \mu^{(0)'_3, \mu^{(0)'_4}}}) = (0, p_2, p_1, 0) \), which is in \( H_{0C} \) since

\[
p_1 \Delta_1 + p_2 \Delta_2 = p_1 (\mu^{(0)'}_{11} - \mu^{(0)'}_{10}) + p_2 (\mu^{(0)'}_{21} - \mu^{(0)'}_{20}) = p_1 (p_2 - 0) + p_2 (0 - p_1) = 0.
\]

We will show the familywise Type I error rate of \( A \) at \( \mu^{(0)'} \) exceeds \( \alpha \), leading to a contradiction.

The following claims follow from Lemma 3: the first stage z-statistic \( Z^{(1)} \) has a standard normal distribution; conditioned on \( D(X^{(1)}) \), the stage two z-statistic \( Z^{(2)} \) is independent of \( X^{(1)} \); conditioned on \( X^{(1)} \), the stage two z-statistic \( Z^{(2)} \) has a normal distribution with conditional variance 1 and conditional mean \( E(Z^{(2)} | X^{(1)}) \) equal to 0 if \( D(X^{(1)}) = C \) and equal to

\[
(\mu^{(0)'_1} - \mu^{(0)'_2}) (f_{21} n_1)^{1/2} / (2 \sigma) = p_2 (f_{21} n_1)^{1/2} / (2 \sigma) > 0 \text{ if } D(X^{(1)}) = 1.
\]

Define the random variable \( U \) to have a standard normal distribution and be independent of \( X^{(1)} \). Below, we let \( P_{D, \mu^{(0)'}} \) and \( E_{D, \mu^{(0)'}} \) denote probability and expectation, respectively, under \( \mu = \mu^{(0)'} \), and design \( D \).

By the above arguments, the probability that \( A \) rejects \( H_{0C} \) at \( \mu^{(0)'} \) equals the probability that \( Z^{\text{TOTAL}} > \Phi^{-1}(1 - \alpha) \) at \( \mu^{(0)'} \), which equals

\[
P_{D, \mu^{(0)'}} \{ Z^{\text{TOTAL}} > \Phi^{-1}(1 - \alpha) \}
= P_{D, \mu^{(0)'}} \left\{ f_1^{1/2} Z^{(1)} + f_2^{1/2} Z^{(2)} > \Phi^{-1}(1 - \alpha) \right\}
= P_{D, \mu^{(0)'}} \left[ f_1^{1/2} Z^{(1)} + f_2^{1/2} \left\{ Z^{(2)} - E(Z^{(2)} | X^{(1)}) + E(Z^{(2)} | X^{(1)}) \right\} > \Phi^{-1}(1 - \alpha) \right]
= E_{D, \mu^{(0)'}} \text{} P_{D, \mu^{(0)'}} \left[ f_1^{1/2} Z^{(1)} + f_2^{1/2} \left\{ Z^{(2)} - E(Z^{(2)} | X^{(1)}) + E(Z^{(2)} | X^{(1)}) \right\} > \Phi^{-1}(1 - \alpha) | X^{(1)} \right] \tag{99}
\]
\begin{align*}
&= E_{D,\mu^{(0)}} P_{D,\mu^{(0)}} \left[f_1^{1/2} Z^{(1)}ight] \\
&\quad + f_2^{1/2} \left\{ U + 1[D(X^{(1)}) = 1] \frac{p_2(f_2 n)^{1/2}}{2\sigma} \right\} > \Phi^{-1}(1 - \alpha) \mid X^{(1)} \right] (100) \\
&= P_{D,\mu^{(0)}} \left[f_1^{1/2} Z^{(1)} + f_2^{1/2} \left\{ U + 1[D(X^{(1)}) = 1] \frac{p_2(f_2 n)^{1/2}}{2\sigma} \right\} > \Phi^{-1}(1 - \alpha) \mid X^{(1)} \right] (101) \\
&= P_{D,\mu^{(0)}} \left[f_1^{1/2} Z^{(1)} + f_2^{1/2} U > \Phi^{-1}(1 - \alpha) \mid D(X^{(1)}) = C \right] P_{\mu^{(0)}}(D(X^{(1)}) = C) \\
&\quad + P_{D,\mu^{(0)}} \left[f_1^{1/2} Z^{(1)} + f_2^{1/2} U > \Phi^{-1}(1 - \alpha) \mid D(X^{(1)}) = 1 \right] \times \\
&\quad \quad \quad \quad \quad \quad \quad \quad \quad P_{\mu^{(0)}}(D(X^{(1)}) = 1) (102) \\
&> P_{D,\mu^{(0)}} \left[f_1^{1/2} Z^{(1)} + f_2^{1/2} U > \Phi^{-1}(1 - \alpha) \mid D(X^{(1)}) = C \right] P_{\mu^{(0)}}(D(X^{(1)}) = C) \\
&\quad + P_{D,\mu^{(0)}} \left[f_1^{1/2} Z^{(1)} + f_2^{1/2} U > \Phi^{-1}(1 - \alpha) \mid D(X^{(1)}) = 1 \right] P_{\mu^{(0)}}(D(X^{(1)}) = 1) \\
&= P_{D,\mu^{(0)}} \left[f_1^{1/2} Z^{(1)} + f_2^{1/2} U > \Phi^{-1}(1 - \alpha) \right] = \alpha,
\end{align*}

where (99) and (101) follow from the tower property of conditional expectation; (100) follows from the definition of $U$ above and that conditioned on $X^{(1)}$, the stage two $z$-statistic $Z^{(2)}$ has a normal distribution with conditional variance 1 and conditional mean $1[D(X^{(1)}) = 1] p_2(f_2 n)^{1/2}/(2\sigma)$; (102) follows from $f_1^{1/2} Z^{(1)} + f_2^{1/2} U$ having a standard normal distribution, and the assumption above that $D(X^{(1)}) = 1$ with nonzero probability. Therefore, the familywise Type I error rate of $A$ exceeds $\alpha$. This contradicts the assumption that $A$ strongly controls the familywise Type I error rate at level $\alpha$. This contradiction implies $D(X^{(1)}) = 1$ with probability zero, i.e., $D(X^{(1)}) = C$ with probability 1.

Define the standard design $S_A \in S$ (which by definition enrolls the combined population in stage 2) to be that using the multiple testing procedure $M$ (recall $M$ is the multiple testing procedure used by $A$) when $D(X^{(1)}) = C$ and that rejects no null hypothesis otherwise. Then $S_A$ has identical power and Type I error as $A$ at any $\mu \in \mathbb{R}^4$. By the above assumptions on $A$, the standard design $S_A$ has greater overall power than UMP$_C$ at $\omega_1$, and greater or equal power as UMP$_C$ to reject $H_{0C}$ at $\omega_C$. But by Theorem 1, at $\omega_1$ the overall power of $S_A$ can be no larger than that of UMP$_C$. This contradiction shows that no adaptive design $A \in \mathcal{A}$ can simultaneously have greater overall power than UMP$_C$ at $\omega_1$, and greater or equal power as UMP$_C$ to reject $H_{0C}$ at $\omega_C$. This completes the proof of Theorem 3. \hfill \square

G. ALGORITHM FOR SOLVING OPTIMIZATION PROBLEM IN SECTION 8

Assume $\sigma_{st}^2 = \sigma_1^2$ for all $s \in \{1, 2\}$, $t \in \{0, 1\}$. We describe how, for each $q$ and $p_1$, we solved the optimization problem (8) subject to the constraint (9). We use the following property:

**Lemma 12.** For any $f_1 \in [0.05, 0.95]$, the probability in (8) is a non-decreasing function of $\theta$, and the probability in (9) is an non-increasing function of $\theta$. 

http://biostats.bepress.com/jhubiostat/paper260
Proof: Since both null hypotheses $H_{0C}$, $H_{01}$ are false at $\omega_1$, the probability in (8) equals

$$P_{\theta, \omega_1} \{ Z_{\text{TOTAL}} > z_{1-\alpha} \},$$

where we let $z_{1-\alpha} = \Phi^{-1}(1 - \alpha)$. We show the above display is a non-decreasing function of $\theta$.

Define the following function of $\theta$ and $X^{(1)}$:

$$h(\theta, X^{(1)}) = 1[D_\theta(X^{(1)}) = C] \rho_1 \Delta_{\min} \left\{ \frac{n_2 p_1}{2(\sigma_{10}^2 + \sigma_{11}^2)} \right\}^{1/2} + 1[D_\theta(X^{(1)}) = 1] \Delta_{\min} \left\{ \frac{n_2}{2(\sigma_{10}^2 + \sigma_{11}^2)} \right\}^{1/2}.$$

By parts (iv) and (v) of Lemma 3, the above display is the conditional mean of $Z^{(2)}$ given $X^{(1)}$ at alternative $\omega_1$. We will show for any possible value of $X^{(1)}$ that $h(\theta, X^{(1)})$ is non-decreasing in $\theta$. First, it follows from the definition of $D_\theta$ from Section 5 that for any $\theta_1 < \theta_2$ and any possible value of $X^{(1)}$,

$$1[D_{\theta_1}(X^{(1)}) = 1] \leq 1[D_{\theta_2}(X^{(1)}) = 1].$$

(103)

Also, we have

$$\Delta_{\min} \left\{ \frac{n_2}{2(\sigma_{10}^2 + \sigma_{11}^2)} \right\}^{1/2} - \rho_1 \Delta_{\min} \left\{ \frac{n_2 p_1}{2(\sigma_{10}^2 + \sigma_{11}^2)} \right\}^{1/2} > 0,$$

which follows by $\rho_1 p_1^{1/2} < 1$. Then we have

$$h(\theta, X^{(1)}) = \left( 1 - 1[D_\theta(X^{(1)}) = 1] \right) \rho_1 \Delta_{\min} \left\{ \frac{n_2 p_1}{2(\sigma_{10}^2 + \sigma_{11}^2)} \right\}^{1/2} + 1[D_\theta(X^{(1)}) = 1] \Delta_{\min} \left\{ \frac{n_2}{2(\sigma_{10}^2 + \sigma_{11}^2)} \right\}^{1/2}$$

(105)

$$= \rho_1 \Delta_{\min} \left\{ \frac{n_2 p_1}{2(\sigma_{10}^2 + \sigma_{11}^2)} \right\}^{1/2} + 1[D_\theta(X^{(1)}) = 1] \left[ \Delta_{\min} \left\{ \frac{n_2}{2(\sigma_{10}^2 + \sigma_{11}^2)} \right\}^{1/2} - \rho_1 \Delta_{\min} \left\{ \frac{n_2 p_1}{2(\sigma_{10}^2 + \sigma_{11}^2)} \right\}^{1/2} \right],$$

(106)

where (105) follows from $1[D_\theta(X^{(1)}) = C] + 1[D_\theta(X^{(1)}) = 1] = 1$ for all possible values of $\theta$ and $X^{(1)}$. By (103) and (104), we have for any possible value of $X^{(1)}$ that (106) is a non-decreasing function of $\theta$. This shows for any value of $X^{(1)}$, that $h(\theta, X^{(1)})$ is a non-decreasing function of $\theta$.

Define the random variable $U$ to have a standard normal distribution and be independent of $X$. We have

$$P_{\theta, \omega_1} \{ Z_{\text{TOTAL}} > z_{1-\alpha} \}$$

(107)

$$= P_{\theta, \omega_1} \left\{ \int_1^{1/2} Z^{(1)} + \int_2^{1/2} Z^{(2)} > z_{1-\alpha} \right\}$$

$$= E_{D_\theta, \omega_1} \left\{ \int_1^{1/2} Z^{(1)} + \int_2^{1/2} Z^{(2)} > z_{1-\alpha} \mid X^{(1)} \right\}$$

(108)

$$= E_{D_\theta, \omega_1} E_{D_\theta, \omega_1} \left\{ \int_1^{1/2} Z^{(1)} + \int_2^{1/2} \left\{ U + h(\theta, X^{(1)}) \right\} > z_{1-\alpha} \mid X^{(1)} \right\},$$

(109)

where (108) follows from the tower property of conditional expectation; (109) follows from the definition of $U$ above and from parts (iv) and (v) of Lemma 3 which imply that conditioned on
$X^{(1)}$, the stage two $z$-statistic $Z^{(2)}$ has a normal distribution with variance 1 and mean $h(\theta, X^{(1)})$ at alternative $\omega_1$. Since we showed above that for any possible value of $X^{(1)}$, we have $h(\theta, X^{(1)})$ is a non-decreasing function of $\theta$, this implies the conditional probability
\[
P_{D_{\theta, \omega_1}} \left[ f_1^{1/2} Z^{(1)} + f_2^{1/2} \left\{ U + h(\theta, X^{(1)}) \right\} > z_{1-\alpha} \mid X^{(1)} \right],
\]
in (109) is a non-decreasing function of $\theta$ (with probability 1). Therefore, the expression (109) is a non-decreasing function of $\theta$, which implies (107) is a non-decreasing function of $\theta$. The above argument shows the probability in (8) is a non-decreasing function of $\theta$. A similar argument shows the probability in (9) is a non-increasing function of $\theta$. This proves Lemma 12. □

We compute the value of $\theta$ maximizing the probability in (8) under constraint (9) by binary search, where for each candidate value of $\theta$, the left side of (9) is computed using the multivariate normal distribution function, using the R package mvtnorm. This was done for each $f_1 \in F$, and the corresponding value of (8) was recorded. The solution $(\theta(q, p_1), f_1(q, p_1))$ was set to be the pair for which the value of (8) was largest. R code implementing the above optimization is given in the Supplementary Materials.

H. PROOF THAT EACH DESIGN IN $A^*_4$ STRONGLY CONTROLS THE FAMILYWISE TYPE I ERROR RATE AT LEVEL $\alpha = 0.05$

The class of designs $A^*_4$ was defined in Section 8. We verified that each design $A^q_{\nu_1} \in A^*_4$ strongly controls the familywise Type I error rate at level $\alpha = 0.05$. This was done using the method in (Rosenblum and van der Laan, 2011, Section F of Supplementary Materials). Below we describe how this was implemented.

For any design $A_{\theta, f_1} \in A_4$, define $A^\alpha_{\theta, f_1}$ to be the design identical to $A_{\theta, f_1}$ except that it rejects both $H_{0C}$ and $H_{01}$ if all of the following hold:
\[
D_{\theta, f_1}(X^{(1)}) = C, \quad Z^{TOTAL} > \Phi^{-1}(1-\alpha), \quad \text{and} \quad Z^{TOTAL}_1 > \Phi^{-1}(1-\alpha) + 0.055. \tag{110}
\]
Define the class of augmented designs $A^*_4 = \{ A^\alpha_{\theta, f_1} : \theta \in \mathbb{R}, f_1 \in [0.05, 0.95] \}$. It follows that $A^*_4 \subset A^*_4$.

Consider any $\theta \in \mathbb{R}$ and $f_1 \in [0.05, 0.95]$. Let $h(\nu_1, \nu_2, \theta, f_1)$ denote the probability that the multivariate normal random vector $(X_1, X_2, X_3)$ with mean vector $(\nu_1, (\nu_1 - \nu_2)/2^{1/2}, \nu_2/f_1^{1/2})$ and covariance matrix
\[
\begin{pmatrix}
1 & 2^{-1/2} & 0 \\
2^{-1/2} & 1 & -2^{-1/2}f_1^{1/2} \\
0 & -2^{-1/2}f_1^{1/2} & 1
\end{pmatrix}, \tag{111}
\]
is in the region $(-\infty, \theta) \times (0, \infty) \times (\Phi^{-1}(0.95) + 0.055, \infty)$. Let $h''(\nu_1, \theta, f_1)$ denote the probability that a bivariate normal random vector $(X'_1, X'_2)$ with mean vector $(\theta - \nu_1, \nu_1(1-p_1)/2^{1/2} - \Phi^{-1}(0.95))$ and covariance matrix with 1s on the main diagonal and $- (f_1p_1)^{1/2}$ off the diagonal, is in the region $[0, \infty) \times [0, \infty)$. Define
\[
h''(\nu_1, \theta) = \Phi(\nu_1 - \theta)\Phi(-\Phi^{-1}(0.95) - 0.055).
\]

It follows from the proof in (Rosenblum and van der Laan, 2011, Section F of Supplementary Materials) that for any $\theta \in \mathbb{R}$ and $f_1 \in [0.05, 0.95]$, to verify that the design $A^\alpha_{\theta, f_1} \in A^*_4$ strongly controls the familywise Type I error rate at level 0.05, under the assumption from Section 8 that
Adaptive Enrichment Designs Not Dominated by Any Standard Design

\[ \sigma^2_{st} = \sigma^2 \] for all \( s \in \{1, 2\}, t \in \{0, 1\} \), it suffices to upper bound the following quantity by 0.05:

\[
\sup_{\nu_1 \in [0.5], \nu_2 \in [-3f_1^{1/2}, 0]} \left\{ \hat{h}(\nu_1, \nu_2, \theta, f_1) + \hat{h}'(\nu_1, \theta, f_1) + \hat{h}''(\nu_1, \theta) \right\}.
\] (112)

Furthermore, it was shown in (Rosenblum and van der Laan, 2011, Section F of Supplementary Materials) that such an upper bound could be established by performing a grid search over pairs \((\nu_1, \nu_2)\) in the set

\[ \tilde{H} = \{0, 10, 12, 14\} \times \{0, 8, 10, 12\}, \]

for \( \tau_1 = 1/5000 \) and \( \tau_2 = 1/5000 \), computing (112) at each grid point using the R code included in the Supplementary Material. We executed this computation for each design in \( A^{p_1+} \in A^1 \), and the maximum value of (112) was 0.0496, rounded up at the fourth decimal place. This result, combined with the upper bound on the approximation error of the grid search from (Rosenblum and van der Laan, 2011, Section F of Supplementary Materials), implies that each design \( A^{p_1+} \in A^1 \) strongly controls the familywise Type I error rate at level 0.05.

I. Plots of Power Tradeoffs at \( p_1 = 1/2 \)

Define the set of designs \( A^1 = \{ A^{p_1+} : p_1 = 1/2, q \in \{0.01, 0.02, \ldots, 0.80\}\} \cup \{\text{UMP}\}. In Figure 2a, for each design in this set, we plot the point \((x, y)\) where \( x \) is its power to reject \( H_{0C} \) at \( \omega_C \), and \( y \) is its overall power at \( \omega_1 \).

The adaptive designs \( A^{0.7, 0.5+}, A^{0.75, 0.5+}, A^{0.78, 0.5+} \), which are highlighted in Figure 2a, have 17%, 12%, 7% greater overall power at \( \omega_1 \) than \( \text{UMP}_C \), respectively. (Here and below, each percent represents the absolute difference, e.g., 17% represents the difference 0.17 = 0.47 – 0.30.) The tradeoff is that \( A^{0.7, 0.5+}, A^{0.75, 0.5+}, A^{0.78, 0.5+} \) have 10%, 5%, 2% less power to reject \( H_{0C} \) at \( \omega_C \), respectively. All designs in Figure 2a have overall power 0.8 at \( \omega_C \).

Each of Theorems 1–3 has a graphical interpretation in terms of Figure 2a. For any design \( A \in \mathcal{A} \), let \((x_A, y_A)\) denote the coordinates of the corresponding point, were it plotted in Figure 2a; i.e., \( x_A \) is the power of \( A \) to reject \( H_{0C} \) at \( \omega_C \), and \( y_A \) is the overall power of \( A \) at \( \omega_1 \). Let \((x_{\text{UMP}}, y_{\text{UMP}}) = (0.80, 0.34)\) denote the point corresponding to \( \text{UMP}_C \), represented by a white circle in Figure 2a.

By Theorem 1, for any standard design \( S \in \mathcal{S} \) with overall power at least 0.8 at \( \omega_C \), the corresponding point \((x_S, y_S)\) must satisfy \( x_S \leq x_{\text{UMP}} \) and \( y_S \leq y_{\text{UMP}} \). By part (iii) of Theorem 2, for any design \( A \in \mathcal{A}_a \), the point \((x_A, y_A)\) must satisfy \( y_A > y_{\text{UMP}} \). By Theorem 3, for any design \( A \in \mathcal{A} \), if \( y_A > y_{\text{UMP}} \) then \( x_A < x_{\text{UMP}} \). Therefore, since \( \mathcal{A}_a \subset \mathcal{A} \), we have for any design \( A \in \mathcal{A}_a \), the point \((x_A, y_A)\) is above and to the left of \((x_{\text{UMP}}, y_{\text{UMP}})\).

Figure 2b plots power to reject \( H_{01} \) at \( \omega_1 \) versus power to reject \( H_{0C} \) at \( \omega_C \). The curve is similar to Figure 2a, except shifted down slightly.

The curves in Figures 2a and 2b are essentially flat for values on the horizontal axis less than 0.4. This shows there is little gained in terms of power at \( \omega_1 \) by sacrificing more than 40% power to reject \( H_{0C} \) at \( \omega_C \) compared to \( \text{UMP}_C \).

J. Local Alternatives

We sketch the notion of local alternatives mentioned in Section 10, and present a conjecture related to how our result may be extended to non-normal distributions, in an asymptotic sense. By (5), the total sample size \( n(\alpha, \beta) \) is proportional to \( \Delta^{-2}_{\min} \). Equivalently, \( \Delta_{\min} \) is proportional
Fix $\alpha = 0.05, \beta = 0.2, \sigma^2_{st} = \sigma^2$ for each $s \in \{1, 2\}, t \in \{0, 1\}$, and define the following sequence, for each $n > 0$,

$$
\Delta_{\min, n} = n^{-1/2} \left[ 2 \left( \Phi^{-1}(1 - \alpha) + \Phi^{-1}(1 - \beta) \right)^2 \left( p_1(\sigma^2_{10} + \sigma^2_{11}) + p_2(\sigma^2_{20} + \sigma^2_{21}) \right) \right]^{1/2}.
$$

Define the local alternatives $\omega_{C,n}$ and $\omega_{1,n}$ analogous to $\omega_C$ and $\omega_1$, respectively, except using $\Delta_{\min, n}$ instead of $\Delta_{\min}$ in (3). We modify the assumption from Section 3 to be that each $Y_i$, rather than being a draw from a normal distribution with mean $\mu_{S_iT_i}$ and variance $\sigma^2_{S_iT_i}$, is a random draw from an arbitrary, fixed (i.e., not changing with sample size) distribution with mean $\mu_{S_iT_i}$ and variance $\sigma^2_{S_iT_i}$. We conjecture that for any $A_{\theta,f} \in A_*$, in the limit as $n$ goes to infinity, its overall power at $\omega_{C,n}$ converges to $1 - \beta$, its overall power at $\omega_{1,n}$ converges to a value exceeding (6), and the Type I error at $\mu = (0, 0, 0, 0)$ converges to 0.05.

[Received January 2014]