Sankhyā : The Indian Journal of Statistics
2018, Volume 80-B, Part 1, pp. 151-177
© 2017, Indian Statistical Institute

A Two-Stage Design for Comparative Clinical Trials: the Heteroscedastic Solution

Elena M. Buzaianu
University of North Florida, Jacksonville, USA
Pinyuen Chen
Syracuse University, Syracuse, USA

Abstract

The problem of comparing several experimental treatments to a control arises frequently in clinical trials. Various multi-stage randomized phase II/III designs have been proposed for the purpose of selecting one or more promising experimental treatments and comparing them with a control, while controlling overall Type I and Type II error rates. In this paper, a hybrid selection and testing design for comparing the means of several experimental normal populations among themselves and with the mean of a control normal population is proposed. It is assumed that the variances of the experimental and the control normal populations are unknown and unequal. A Stein-type two-sample selection approach is used at both the selection and testing stages to solve the heteroscedastic problems caused by the unknown variances. The hybrid two-stage design allows for dropping the poorly performing treatments early on the basis of interim analysis results and for early termination if none of the experimental treatments seems promising. Numerical computations are given to show the advantage of the proposed procedure over a pure selection procedure. An example is provided to illustrate the use of the new procedure.

AMS (2000) *subject classification.* Primary 62F07; Secondary 62L05, 62P30.
Keywords and phrases. Clinical trials, Hybrid selection and testing design, Ranking and selection, Stein-type two-stage design

1 Introduction

The problem of comparing several experimental treatments (populations) to a control arises frequently in medical research. In this paper, we propose a seamless phase II/III design where several doses of a drug or several different treatments are simultaneously compared with a control. Several authors, under various formulations, have considered the problem of simultaneously comparing k experimental treatments among themselves and with

reference to a control. In this paper, we consider the goal of comparing the means $\mu_1, \mu_2, \ldots, \mu_k$ of k normally distributed experimental treatments $\pi_1, \pi_2, \ldots, \pi_k$ among themselves and with the unknown mean μ_0 of a control normal treatment π_0. The variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2$ and σ_0^2 of the k experimental treatment and the control treatment, respectively, are all assumed to be unknown and unequal. We would like to select one of the experimental treatments, provided that it is better than the control. If none of the k treatments is better than the control, then no experimental treatment is to be chosen. Here, "better than the control" means that the mean of the experimental treatment is sufficiently larger than the mean of the control treatment. We propose a hybrid selection and testing design in order to achieve this goal.

Dantzig (1940) showed that, in testing the hypothesis of normal means, there does not exist a one-sample test procedure whose power function is independent of the unknown variance. For testing normal means when the variances are unknown, Stein (1945) proposed two-sample procedures whose power functions are independent of the unknown variances. His idea was to estimate the unknown variances from the first sample and used the variance estimates in the second sample to construct the test procedures. Following Stein's idea, Bechhofer et al. (1954) studied two-sample selection procedures for selecting the best normal means when the variances are unknown and equal. Dudewicz and Dalal (1975) extended the Bechhofer et al. (1954)'s work from unknown equal variances case to unknown unequal variances case.

For selecting the best experimental treatment provided it is better than a specified (known) standard, Bechhofer and Turnbull (1978) proposed a ranking and selection problem formulation and two selection procedures, a one-sample procedure for common known variance case and a two-sample procedure for the common unknown variance case. Aiming at comparing medical trials from a coronary drug project study, Dunnett (1984) proposed a single-stage procedure for the goal of selecting the best normal population with regard to a control normal population, for the case of common known variance. Considering the same goal but for the case of unknown and unequal variances, Buzaianu and Chen (2015) proposed a Stein-type two-sample selection procedure for normal means with reference with a control.

By using the normal approximation to the binomial, Dunnett (1984) extended his single-stage procedure for comparing normal populations with regard to a control normal population to comparison of probabilities of success of Bernoulli experimental treatments with the unknown probability of a control Bernoulli treatment. Using statistical concepts that professionals

in medical and pharmaceutical sciences are more familiar with, Thall et al. (1988, 1989) proposed a hybrid two-stage selection and testing procedure for the goal of comparing several binomial treatments with a control treatment. At the first stage, ranking and selection techniques are used to screen out the most promising treatment, and at the second stage, hypothesis testing is used to determine if the chosen experimental treatment is better than the control treatment.

The purpose of this paper is to compare several normal experimental populations among themselves and with the control population by using the hybrid selection and testing design as considered by Thall et al. (1988, 1989). Instead of comparing Bernoulli success probabilities as in Thall et al. (1988, 1989), we will compare the means of normal populations when the experimental and control populations have unknown and unequal variances. A Stein-type two-sample method approach is used in each of the selection stage and testing stage of the proposed procedure.

One important difference between the Buzaianu and Chen's procedure (2015) and the proposed procedure is that the goal of Buzaianu and Chen's procedure (2015) is to select the population having the largest mean μ_k, given that it is larger than the mean of the control population ($\mu_k > \mu_0$), while the goal of the proposed procedure is to select one of the experimental treatments, provided that its mean is larger than the mean of the control. Another major difference between the two procedures is that Buzaianu and Chen's procedure (2015) is a pure ranking and selection procedure, while the proposed procedure combines ranking and selection with hypothesis testing. An advantage of the proposed design is that it allows for early termination if none of the experimental treatments seems promising, which may be particularly useful in medical trials where the prolonged administration of an ineffective treatment is undesirable. Another advantage of the proposed hybrid design is the potential reduction in sample size. Thall et al. (1988) showed that, in general, their hybrid design for comparing binomial treatments requires smaller sample sizes than Dunnet's selection procedure (1984), for the same probability requirements.

In Section 2, we provide the problem formulation and outline the proposed design. In Section 3, we state the main results with regard to the power, size, and the least favorable configuration. In Section 4, we derive the lower and upper bounds on the total expected sample size. In Section 5, we derive the formulas for calculating the design parameters and we provide tables of the design parameters that are needed to achieve certain size and power requirements. In Section 6, we provide numerical results of the upper

and lower bounds of the expected sample size for various combinations of the first sample size, the number of populations, and the power and the size of the procedure, and compare the proposed procedure with an analogous pure selection procedure (Buzaianu and Chen (2015)). In Section 7, we present an example illustrating the use of our procedure. We provide some discussion in Section 8. The Appendix includes the proofs of the results we state in Section 3.

2 Problem Formulation and Procedure

Suppose that there are k experimental treatments $\pi_1, \ldots \pi_k$ and a control treatment π_0. We assume that $\pi_i (0 \leq i \leq k)$ is normally distributed with unknown mean μ_i and unknown variance σ_i^2. The comparison of these treatments is with respect to their means. Their variances are assumed to be unknown and unequal and are nuisance parameters. The ranked values of the experimental population means are denoted by $\mu_{[1]} \leq \ldots \leq \mu_{[k]}$ and the population associated with $\mu_{[i]}$ is denoted by $\pi_{[i]}$. We will actually drop the square bracket notation and assume that $\mu_1 \leq \ldots \leq \mu_k$, and the population with μ_i is denoted simply by π_i.

The experimenter will specify constants δ_1 and δ_2 such that $0 \leq \delta_1 < \delta_2$. As interpreted by Thall et al. (1988), $\mu_0 + \delta_1$ represents a marginal improvement over the control μ_0 and $\mu_0 + \delta_2$ represents a significant improvement over the control μ_0. Any treatment π_ν with $\mu_\nu \geq \mu_0 + \delta_2$ is said to be acceptable. The goal is to select from the k experimental populations an acceptable one, if that exists. Otherwise, we will not select any experimental treatment. That is, we will select the control treatment.

The proposed procedure consists of two stages. The first stage is the selection stage where we select the best experimental treatment, provided it is better than the control. If the sample evidence shows that no experimental treatment is better than the control, the procedure terminates with the decision of accepting the null hypothesis $H_0 : \mu_1 = \mu_2 = \ldots = \mu_k = \mu_0$, meaning that all the experimental treatments are equivalent to the control treatment. Otherwise, we proceed to the second stage where we test the above null hypothesis. The size, calculated under the null hypothesis, and the power, calculated under the preference zone, will be defined right after we introduce our procedure in the next paragraph. Here, we adopt the notion for the null hypothesis of a homogeneous test since the size of our procedure is defined analogously to the size in hypothesis testing. There is no other analogy between our procedure and homogeneous test.

Following Stein's two-stage structure for hypothesis testing problems and Dudewicz and Dalal's two-stage structure for selection problems, for

prespecified integers $N_0 \geq 2, M_0 \geq 2$ and positive constants y_1, y_2, h_1, h_2, the proposed procedure is defined as follows:

Procedure P_{ST}.

STAGE I (Selection)

Step 1. Take an initial random sample $X_{i1}, X_{i2}, \ldots, X_{iN_0}$ of size N_0 from each population $\pi_i, i = 0, 1, \ldots, k$, and compute the following:

$$\overline{X}_{i(N_0)} = \frac{\sum_{j=1}^{N_0} X_{ij}}{N_0},$$

$$S^2_{i(N_0)} = \frac{\sum_{j=1}^{N_0} (X_{ij} - \overline{X}_{i(N_0)})^2}{N_0 - 1},$$

$n_i = \max\{N_0 + 1, [\frac{S^2_{i(N_0)} h_1^2}{y_1^2}] + 1\}$, where $[x]$ denotes the largest integer less than x.

Step 2. Take $n_i - N_0$ additional independent observations $X_{i,N_0+1}, X_{i,N_0+2}, \ldots, X_{i,n_i}$ from π_i and define $\widetilde{X}_i = \sum_{j=1}^{n_i} a_{ij} X_{ij}$, where the $a'_{ij}s$ are chosen so that $\sum_{j=1}^{n_i} a_{ij} = 1$, $a_{i1} = \ldots = a_{iN_0}$ and $S^2_{i(N_0)} \sum_{j=1}^{n_i} a_{ij}^2 = \frac{y_1^2}{h_1^2}$.

Step 3. If $\max_{i=\overline{1,k}}(\widetilde{X}_i) > \widetilde{X}_0 + y_1$, we denote by π_ν the population which produced this maximum and go to STAGE II. Else, we do not reject H_0 and select the control population.

STAGE II (Testing).

Step 1. Take M_0 additional observations from $W_{i1}, W_{i2}, \ldots W_{iM_0}$ from $\pi_i, i \in \{0, \nu\}$ and find the following:

$$\overline{W}_{i(M_0)} = \frac{\sum_{j=1}^{M_0} W_{ij}}{M_0},$$

$$S^2_{i(M_0)} = \frac{\sum_{j=1}^{M_0} (W_{ij} - \overline{W}_{i(M_0)})^2}{M_0 - 1},$$

$m_i = \max\{M_0 + 1, [\frac{S^2_{i(M_0)} h_2^2}{y_2^2}] + 1\}$, where $[x]$ denotes the largest integer less than x.

Step 2. Take $m_i - M_0$ additional independent observations X_{i,M_0+1}, $X_{i,M_0+2}, \ldots, X_{i,m_i}$ from $\pi_i, i \in \{0, \nu\}$ and define $\widetilde{W}_i = \sum_{j=1}^{m_i} b_{ij} W_{ij}$, where the $b'_{ij}s$ are chosen so that $\sum_{j=1}^{m_i} b_{ij} = 1$, $b_{i1} = \ldots = b_{iM_0}$ and $S^2_{i(M_0)}$ $\sum_{j=1}^{m_i} b_{ij}^2 = \frac{y_2^2}{h_2^2}$.

Step 3. If $\frac{\widetilde{X}_\nu + \widetilde{W}_\nu}{2} > \frac{\widetilde{X}_0 + \widetilde{W}_0}{2} + y_2$, we reject H_0 and select population π_ν. Else, do not reject H_0 and select the control population.

Remark 1. The proposed procedure consists of two stages, a selection stage followed by a testing stage. The procedure studied by Buzaianu and Chen (2015) has only one stage, the selection stage.

Remark 2. From Dudewicz and Dalal (1975), it can be shown that, for given constants $z > 0, N_0 \geq 0, S \geq 0$, and $n = \max\{N_0 + 1, [\frac{S^2}{z}] + 1\}$, constants $a_1, \ldots a_n$ can be chosen so that

$$\sum_{j=1}^{n} a_j = 1, a_1 = \ldots = a_{N_0}, S^2 \sum_{j=1}^{n} a_j^2 = z. \tag{2.1}$$

The explicit solutions are as follows:

$$a_1 = \ldots = a_{N_0} = \frac{c}{N_0} \text{ and } a_{N_0+1} = \ldots = a_n = \frac{1-c}{n-N_0},$$
$$\text{where } c = \frac{N_0}{n}[1 - \sqrt{1 - \frac{n}{N_0}(1 - \frac{(n-N_0)z}{S^2})}]$$

Therefore, it is possible to select such a_{ij}'s and b_{ij}'s so as to satisfy the equalities in (2.1). Explicit choices are $a_{i1} = \ldots = a_{i,N_0} = \frac{c_i^I}{N_0}$ and $a_{i,N_0+1} = \ldots = a_{i,n_i} = \frac{1-c_i^I}{n_i - N_0}$, where $c_i^I = \frac{N_0}{n_i}\left[1 - \sqrt{1 - \frac{n_i}{N_0}(1 - \frac{(n_i - N_0)y_1^2}{h_1^2 S^2_{i(N_0)}})}\right]$, $i = 0, 1, \ldots, k$ and $b_{i1} = \ldots = b_{i,m_0} = \frac{c_i^{II}}{M_0}$ and $b_{i,m_0+1} = \ldots = b_{i,m_i} = \frac{1-c_i^{II}}{m_i - M_0}$, where $c_i^{II} = \frac{M_0}{m_i}\left[1 - \sqrt{1 - \frac{m_i}{M_0}(1 - \frac{(m_i - M_0)y_2^2}{h_2^2 S^2_{i(M_0)}})}\right]$, $i = 0, \nu$.

Remark 3. The design parameters y_1 and y_2 are obtained by solving simultaneous integral equations, which will be stated and derived in Section 4. The algorithm of obtaining those design parameters is detailed in Section 5.

Because our procedure incorporates facets of both statistical selection and testing and hypothesis testing, suitable definitions of size, preference

zone (PZ), power function, and least favorable configuration (LFC) should be defined accordingly. These definitions follow closely those proposed by Thall et al. (1988) who considered the case of binomial treatments. An experimental treatment π_ν is selected if it is selected at the first stage and $H_0 : \mu_1 = \mu_2 = \ldots = \mu_k = \mu_0$ is rejected in favor of the alternative $\mu_\nu > \mu_0$ at the second stage. The level of our procedure is defined as the probability of selecting any experimental treatment π_ν under the null hypothesis H_0. In order to define the power for the hybrid procedure, we will assume that (1) there is at least one acceptable experimental treatment and (2) no μ_ν lies in the interval $(\mu_0 + \delta_1, \mu_0 + \delta_2)$. The necessity for the second assumption is that there is no procedure that can effectively differentiate between two populations whose means are arbitrarily close. We denote by $\vec{\mu}$ the vector parameter $(\mu_0, \mu_1, \mu_2, \ldots, \mu_k)$ and by μ_i the scalar associated with the mean of treatment π_i. The preference zone (PZ) will be the set of all parameter configurations $\vec{\mu}$'s satisfying assumptions (1) and (2) above. The power function of our procedure, denoted by $1-\beta(\vec{\mu})$, is defined for any configuration $\vec{\mu}$ in the PZ, as the probability of selecting any acceptable treatment π_ν. The LFC will be the parameter configuration in the PZ that minimizes the power function $1 - \beta(\vec{\mu})$.

3 Size, Power, and the Least Favorable Configuration

In this section, we derive the least favorable configuration, the power, and the type I error rate. As commented in Section 2, Dantzig (1940) has shown the non-existence of a single-stage testing procedure for normal means whose power is independent of the unknown variances, and Stein (1945) has shown the existence of a two-sample testing procedure for normal means whose power is independent of the unknown variances. We show that our two-sample two-stage design has power and type I error rate independent of the unknown variances. As a consequence, our LFC is also independent of the unknown variances.

The choice of $a'_{ij}s$ so that $\sum_{j=1}^{n_i} a_{ij} = 1$, $a_{i1} = \ldots = a_{iN_0}$, and $S^2_{i(N_0)} \sum_{j=1}^{n_i} a^2_{ij} = \frac{y_1^2}{h_1^2}$ guarantees (Dudewicz and Dalal, 1975) that, conditioning for $S_{i(N_0)} = s_{i(N_0)}$, $\frac{\widetilde{X}_i - \mu_i}{\sqrt{s^2_{i(N_0)} \sum_{j=1}^{n_i} a^2_{ij}}}$ has a normal distribution with mean zero and variance $\sigma_i^2/s^2_{i(N_0)}$. Moreover, it also guarantees that $\widetilde{X}_i$ and $S^2_{i(N_0)}$ are independent random variables. These imply that $T_i = \frac{\widetilde{X}_i - \mu_i}{\frac{y_1}{h_1}}$, $i = 1, \ldots, k$, are independent random variables following a Student's t distribution with $N_0 - 1$ degrees of freedom. Then, if $F_{\widetilde{X}_i}(x|\mu_i)$ denotes the distribution function of

$\widetilde{X}_i$, $F(\widetilde{X}_i|\mu_i)$ is the distribution function of $\frac{y_1}{h_1}T + \mu_i$ where T is a random variable following a Student's t distribution with $N_0 - 1$ degrees of freedom. Similarly, it can be shown that, if $G_{\widetilde{X}_i}(x|\mu_i)$ denotes the distribution function of $\widetilde{W}_i$, $G(\widetilde{W}_i|\mu_i)$ is the distribution function of $\frac{y_2}{h_2}T + \mu_i$ where T is a random variable following a Student's t distribution with $M_0 - 1$ degrees of freedom.

The following lemma shows that the distributions of the two statistics $\widetilde{X}$ and $\widetilde{W}$ used in our design satisfy an important monotonicity property, a condition for applying Mahamunulu (1967)'s Lemma, the major tool for deriving our LFC.

LEMMA 3.1. *If $F(\widetilde{X}_i|\mu_i)$ denotes the distribution function of $\widetilde{X}_i$ and $G(\widetilde{W}_i|\mu_i)$ denotes the distribution function of $\widetilde{W}_i$, then the collections $F(\widetilde{X}_i|\mu_i : i = 1, 2, \ldots, k)$ and $\{G(\widetilde{W}_0|\mu_0, G(\widetilde{W}_\nu|\mu_\nu, \nu = 1, 2, \ldots, k)\}$ form two separate stochastically increasing families. Moreover, $\widetilde{X}_0, \widetilde{X}_1, \ldots, \widetilde{X}_k, \widetilde{W}_0, \widetilde{W}_\nu$ are independent random variables.*

The following lemma is due to Mahamunulu (1967) and has been the main tool for deriving the LFC for many selection procedures. We stated it without proof as it can be seen from the original paper by Mahamunulu (1967).

LEMMA 3.2. *Let $F(x|\theta) = F_\theta(x)$ where $\theta \in \Theta$, be a stochastically increasing family of distribution functions on the real line. If Ψ is any nonincreasing (nondecreasing) function of x, then $E_\theta[\Psi(x)]$ is a nonincreasing (nondecreasing) function of θ.*

A typical configuration in the preference zone is

$$\mu_1 \leq \mu_2 \leq \ldots \leq \mu_{k-t} \leq \mu_0 + \delta_1^* < \mu_0 + \delta_2^* = \mu_{k-t+1} \leq \ldots \leq \mu_k, \qquad (3.1)$$

where $1 \leq m \leq k$, the lower bound on m following from the fact that at least one $\mu_i \geq \mu_0 + \delta_2^*$. Let $A = \{k - m + 1, \ldots, k\}$ denote the set of indices of acceptable populations, and $A^c = 1, \ldots, k - m$. Recall that 1 - $\beta(|\vec{\mu}) = P$ (select an acceptable population π_i, $i = 1, \ldots, k|\vec{\mu} \in \text{PZ}) = P(\text{CS}|\vec{\mu} \in \text{PZ})$. The following lemma provides some properties of the power function that will be used when deriving the least favorable configuration. The proof of the lemma is given in the Appendix. The proof first creates a monotone function by which the power function of our procedure is defined. Then, we apply Lemma 3.2 to the power function to show the monotonicity property of the power function.

LEMMA 3.3. *a. The power function 1 - $\beta(\vec{\mu})$ is nondecreasing in μ_i, $i \in A$, when the remaining $\mu_j' s, j = 0, 1, \ldots i-1, i+1, \ldots, k$ are held fixed. b. The power function 1 - $\beta(\vec{\mu})$ is nonincreasing in μ_i, $i \in A^c$, when the remaining $\mu_j' s, j = 0, 1, \ldots i-1, i+1, \ldots, k$ are held fixed.*

Theorem 3.4 derives the LFC for our proposed design by applying the monotonicity results obtained in Lemma 3.3. The proof of the theorem is given in the Appendix.

THEOREM 3.4. *For any given μ_0, δ_1^*, δ_2^*, whenever $\theta_k \geq \mu_0 + \delta_2^*$ and no μ_i lies in the interval $(\mu_0 + \delta_1^*, \mu_0 + \delta_2^*)$, the least favorable configuration for procedure P is $\mu_1 = \mu_2 = \ldots = \mu_{k-1} = \mu_0 + \delta_1^*, \mu_k = \mu_0 + \delta_2^*$.*

The following theorem gives the explicit formulas for the procedure parameters $(N_0, M_0,\ y_1, y_2, h_1, h_2)$ which are needed to achieve any size and power requirements. The derivation of the formulas was based on the LFC obtained in Theorem 3.4. The proof of the theorem is given in the Appendix.

THEOREM 3.5. *Let $\delta_1^*, \delta_2^* (0 \leq \delta_1^* < \delta_2^*), \alpha$, and $1-\beta$ be given. Then, a size of α and, for any parameter vector $\vec{\mu} \in PZ$, a power of $1-\beta$ can be achieved by choosing design parameters $(N_0, M_0, y_1, y_2, h_1, h_2)$ which simultaneously satisfy the equations:*

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{\frac{(y_1-\delta_1^*)h_1}{y_1}+t_0}^{\infty} F_{N_0}^{k-1}\left(t_k + \frac{(\delta_2^*-\delta_1^*)h_1}{y_1}\right) \times F_{M_0}\left(\frac{h_2 y_1}{y_2 h_1} t_k - \frac{y_1 h_2}{h_1 y_2} t_0 - v_0 - \frac{2h_2}{y_2}(y_2 - \delta_2^*)\right) \times f_{N_0}(t_k) f_{N_0}(t_0) f_{M_0}(v_0)\, dt_k\, dt_0\, dv_0 = 1-\beta \tag{3.2}$$

$$k\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{h_1+t_0}^{\infty} F_{N_0}^{k-1}(t_k) F_{M_0}\left(\frac{h_2 y_1}{y_2 h_1} t_k - \frac{y_1 h_2}{h_1 y_2} t_0 - v_0 - 2h_2\right) f_{N_0}(t_k) f_{N_0}(t_0) \times f_{M_0}(v_0)\, dt_k\, dt_0\, dv_0 = \alpha, \tag{3.3}$$

where F_{N_0} and f_{N_0} denote the distribution and the density functions, respectively, of a Student's t random variable with $N_0 - 1$ degrees of freedom, while F_{M_0} and f_{M_0} denote the distribution and the density functions, respectively, of a Student's t random variable with $M_0 - 1$ degrees of freedom.

The following theorem gives the explicit formulas for the probability of terminating the design at Stage I and the probability of continuation to Stage II. Those formulas are used to calculate the expected sample sizes which are needed when comparing our procedure P_{ST} to the existing procedure P_{S} later in Section 6 and also in computing the entries in Tables 2, 3, and 4.

THEOREM 3.6. *Let τ_0 denote the probability of terminating at STAGE I, computed under H_0, and τ_1 denote the probability of continuation to STAGE II, computed under the LFC. Then,*

$$\tau_0 = \int_{-\infty}^{\infty} F_{N_0}^k(t_0 + h_1) f_{N_0}(t_0)\, dt_0. \tag{3.4}$$

$$\tau_1 = 1 - \int_{-\infty}^{\infty} F_{N_0}^{k-1}\left(t_0 + h_1 - \frac{h_1\delta_1^*}{y_1}\right) F_{N_0}\left(t_0 + h_1 - \frac{h_1\delta_2^*}{y_1}\right) f_{N_0}(t_0)\, dt_0. \tag{3.5}$$

4 Expected Total Sample Size

The expected total sample size for our procedure is

$$E(n_T) = E(n_{T_I}) + P(\text{continue to STAGE II stage}) E(n_{T_{II}}). \tag{4.1}$$

Here, $E(n_{\mathrm{T_I}})$ and $E(n_{\mathrm{T_{II}}})$ are the expected sample sizes from STAGE I(selection) and STAGE II(testing), respectively. Also,

$$\begin{aligned} P(\text{continue to STAGE II}) &= 1 - P(\text{terminate early}) \\ &= 1 - \int_{-\infty}^{\infty} \prod_{i=1}^{k} F_{N_0}\left(t_0 + \frac{y_1 + \mu_0 - \mu_i}{\frac{y_1}{h_1}}\right) f_{N_0}(t_0)\, dt_0, \end{aligned} \tag{4.2}$$

where F_{N_0} and f_{N_0} are the distribution and density functions, respectively, of a Student's t random variable with $N_0 - 1$ degrees of freedom.

At the first (selection) stage, we have that

$$P(n_i = N_0 + 1) = P\left(\frac{S_i^2 h_1^2}{(\delta_0^* + c)^2} < N_0 + 1\right) = P\left(\chi_{N_0-1}^2 < \frac{(N_0^2 - 1)(y_1)^2}{\sigma_i^2 h_1^2}\right)$$

$$\begin{aligned} P(n_i = \nu) &= P\left(\nu - 1 < \frac{S_i^2 h_1^2}{(h_1)^2} < \nu\right) \\ &= P\left(\frac{(\nu - 1)(N_0 - 1)z^2}{\sigma_i^2} < \chi_{N_0-1}^2 \le \frac{\nu(N_0 - 1)z^2}{\sigma_i^2}\right) \end{aligned}$$

From Stein (1945), letting $z_1 = \frac{y_1}{h_1}$, we obtain that

$$E(n_i) = (N_0 + 1) P\left(\chi_{N_0-1}^2 < \frac{(N_0 - 1)z_1^2}{\sigma_i^2}\right)$$

$$+\frac{\sigma_i^2}{z_1^2}P\left(\chi^2_{N_0+1}>\frac{(N_0-1)z_1^2}{\sigma_i^2}\right)$$
$$+\theta_i P\left(\chi^2_{N_0-1}>\frac{(N_0-1)z^2}{\sigma_i^2}\right),$$

where $0\le\theta_i<1$. Then,

$$E(n_{T_1})=\sum_{i=0}^{k}(N_0+1)P\left(\chi^2_{N_0-1}<\frac{(N_0-1)z_1^2}{\sigma_i^2}\right)$$
$$+\sum_{i=0}^{k}\frac{\sigma_i^2}{z_1^2}P\left(\chi^2_{N_0+1}>\frac{(N_0-1)z_1^2}{\sigma_i^2}\right)$$
$$+\sum_{i=0}^{k}\theta_i P\left(\chi^2_{N_0-1}>\frac{(N_0-1)z_1^2}{\sigma_i^2}\right),$$

where $0\le\theta_i<1$. Therefore, a lower bound for $E(n_{\mathrm{T_I}})$ is

$$\mathrm{LB_I}=\sum_{i=0}^{k}(N_0+1)P\left(\chi^2_{N_0-1}<\frac{(N_0-1)z_1^2}{\sigma_i^2}\right)+\sum_{i=0}^{k}\frac{\sigma_i^2}{z_1^2}P\left(\chi^2_{N_0+1}>\frac{(N_0-1)z_1^2}{\sigma_i^2}\right)$$

while an upper bound for $E(n_{\mathrm{T_I}})$ is

$$\mathrm{UB_I}=\sum_{i=0}^{k}(N_0+1)P\left(\chi^2_{N_0-1}<\frac{(N_0-1)z_1^2}{\sigma_i^2}\right)+\sum_{i=0}^{k}\frac{\sigma_i^2}{z_1^2}P\left(\chi^2_{N_0+1}>\frac{(N_0-1)z_1^2}{\sigma_i^2}\right)$$
$$+\sum_{i=0}^{k}\frac{\sigma_i^2}{z_1^2}P\left(\chi^2_{N_0+1}>\frac{(N_0-1)z_1^2}{\sigma_i^2}\right).$$

Similarly, letting $z_2=\frac{y_2}{h_2}$, at STAGE II (testing), we have that

$$E(n_{\mathrm{T_{II}}})=\sum_{i=0,\nu}(M_0+1)P\left(\chi^2_{M_0-1}<\frac{(M_0-1)z_2^2}{\sigma_i^2}\right)$$
$$+\sum_{i=0}^{k}\frac{\sigma_i^2}{z_2^2}P\left(\chi^2_{M_0+1}>\frac{(N_0-1)z_2^2}{\sigma_i^2}\right)$$
$$+\sum_{i=0}^{k}\theta_i P\left(\chi^2_{M_0-1}>\frac{(M_0-1)z_2^2}{\sigma_i^2}\right),\ \text{where } 0\le\theta_i<1$$

Note that $E(n_{\mathrm{T_{II}}})$ depends on the experimental treatment π_ν that is selected for STAGE II. Therefore, a lower bound for $E(n_{\mathrm{T}_2})$ is

$$\mathrm{LB_{II}} = \min_{\nu=1,\ldots,k} \left\{ \sum_{i=0,\nu} (M_0+1) P\left(\chi^2_{M_0-1} < \frac{(M_0-1)z_2^2}{\sigma_i^2}\right) + \sum_{i=0,\nu} \frac{\sigma_i^2}{z_2^2} P\left(\chi^2_{M_0+1} > \frac{(m_0-1)z_2^2}{\sigma_i^2}\right) \right\}$$

while an upper bound for $E(n_{\mathrm{T_{II}}})$ is

$$\mathrm{UB_{II}} = \max_{\nu=1,\ldots,k} \left\{ (N_0+1) P\left(\chi^2_{N_0-1} < \frac{(N_0-1)z_2^2}{\sigma_i^2}\right) + \sum_{i=0,\nu} \frac{\sigma_i^2}{z_2^2} P\left(\chi^2_{N_0+1} > \frac{(N_0-1)z^2}{\sigma_i^2}\right) + \sum_{i=0,\nu} P\left(\chi^2_{N_0-1} > \frac{(N_0-1)z_2^2}{\sigma_i^2}\right) \right\}$$

Then, lower and upper bounds for $E(n_{\mathrm{T}})$ are

$$\mathrm{LB} = \mathrm{LB_I} + P(\text{continue to stage II}|\vec{\mu})\mathrm{LB_{II}} \tag{4.3}$$

$$\mathrm{UB} = \mathrm{UB_I} + P(\text{continue to stage II}|\vec{\mu})\mathrm{UB_{II}} \tag{4.4}$$

where P (continue to stage II) $= 1 - \int_{-\infty}^{\infty} \prod_{i=1}^{k} F_{N_0}(t_0 + \frac{y_1+\mu_0-\mu_i}{\frac{y_1}{h_1}}) f_{N_0}(t_0)\, dt_0$, and F and f are the distribution and density functions, respectively, of a Student's t random variable with $N_0 - 1$ degrees of freedom.

5 Design Parameters

Of special interest is the case when $\delta_1^* = 0$. Letting $d_1 = \frac{\delta_2^*}{y_1}$ and $d_2 = \frac{\delta_2^*}{y_2}$, the integral equations (3.2) and (3.3) become

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{(1-d_1)h_1+t_0}^{\infty} F_{N_0}^{k-1}(t_k+d_1h_1) F_{M_0}\left(\frac{h_2d_2}{h_1d_1}(t_k-t_0) - v_0 - 2h_2(1-d_2)\right) \times f_{N_0}(t_k) f_{N_0}(t_0) f_{M_0}(v_0)\, dt_k\, dt_0\, dv_0 = 1-\beta \tag{5.1}$$

$$k\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{h_1+t_0}^{\infty} F_{N_0}^{k-1}(t_k)F_{M_0}\left(\frac{h_2d_2}{h_1d_1}(t_k-t_0)-v_0-2h_2\right)f_{N_0}(t_k)f_{N_0}(t_0) \times f_{M_0}(v_0)\,dt_k\,dt_0\,dv_0=\alpha \tag{5.2}$$

Since the design allows for early termination at stage one (selection) if none of the experimental treatments seem promising, we would like to obtain designs that control $P(\text{early termination}|H_0)$ and $P(\text{continuation to the testing stage}|\text{LFC})$, that is the two probabilities take prespecified values τ_0^* and τ_1^*, respectively.

We will get h_1 using (3.4) by setting

$$\int_{-\infty}^{\infty} P(T_i<t_0+h_1, i=\overline{1,k})f_{N_0}(t_0)\,dt_0=\int_{-\infty}^{\infty} F_{N_0}^{k}(t_0+h_1)f_{N_0}(t_0)\,dt_0=\tau_0^*. \tag{5.3}$$

Table 1: Design parameters for procedure P_{ST} when $\alpha=.05$, $1-\beta=.90$, $\delta_1^*=0$

	$k=2$				$k=3$				$k=4$			
N_0	h_1	d_1	h_2	d_2	h_1	d_1	h_2	d_2	h_1	d_1	h_2	d_2
5	1.48416	2.79263	2.5	1.8	1.82393	2.44977	3.0	1.9	2.06491	2.27623	2.8	1.8
6	1.42949	2.76129	2.5	1.9	1.74884	2.43259	2.5	1.8	1.97165	2.2674	2.7	1.8
7	1.3946	2.74265	2.4	1.9	1.70139	2.42249	2.4	1.8	1.91317	2.26236	2.7	1.9
8	1.37042	2.73039	2.3	1.9	1.66873	2.4159	2.5	1.9	1.87314	2.25915	2.6	1.9
9	1.35269	2.72177	2.3	1.9	1.64489	2.41129	2.4	1.9	1.84403	2.25696	2.5	1.9
10	1.33913	2.71539	2.3	1.9	1.62673	2.40791	2.4	1.9	1.82193	2.25537	2.5	1.9
11	1.32843	2.71049	2.2	1.9	1.61245	2.40532	2.3	1.9	1.80459	2.25418	2.4	1.9
12	1.31977	2.70662	2.2	1.9	1.60092	2.40328	2.3	1.9	1.79061	2.25325	2.4	1.9
13	1.31262	2.70348	2.2	1.9	1.59142	2.40164	2.3	1.9	1.77912	2.2525	2.4	1.9
14	1.30662	2.70089	2.2	1.9	1.58345	2.40028	2.3	1.9	1.7695	2.25189	2.4	1.9
15	1.30151	2.69872	2.2	1.9	1.57668	2.39915	2.3	1.9	1.76132	2.25139	2.3	1.9
16	1.29711	2.69687	2.2	1.9	1.57086	2.39819	2.3	1.9	1.7543	2.25096	2.3	1.9
17	1.29328	2.69528	2.2	1.9	1.56579	2.39736	2.3	1.9	1.74819	2.2506	2.3	1.9
18	1.28991	2.69389	2.2	1.9	1.56134	2.39664	2.3	1.9	1.74284	2.25029	2.2	1.8
19	1.28693	2.69268	2.2	1.9	1.55741	2.39601	2.3	1.9	1.73811	2.25001	2.3	1.9
20	1.28427	2.69161	2.2	1.9	1.55391	2.39546	2.3	1.9	1.7339	2.24977	2.3	1.9
21	1.28189	2.69065	2.1	1.9	1.55076	2.39496	2.3	1.9	1.73012	2.24956	2.3	1.9
22	1.27974	2.68979	2.2	1.9	1.54793	2.39452	2.3	1.9	1.72672	2.24937	2.3	1.9
23	1.27779	2.68902	2.2	1.9	1.54537	2.39412	2.3	1.9	1.72364	2.2492	2.3	1.9
24	1.27601	2.68832	2.2	1.9	1.54303	2.39376	2.2	1.9	1.72083	2.24904	2.3	1.9
25	1.27438	2.68769	2.2	1.9	1.54089	2.39343	2.2	1.9	1.71827	2.2489	2.3	1.9
26	1.27289	2.68711	2.2	1.9	1.53893	2.39313	2.2	1.9	1.71592	2.24877	2.3	1.9
27	1.27152	2.68657	2.2	1.9	1.53713	2.39286	2.2	1.9	1.71376	2.24866	2.3	1.9
28	1.27025	2.68608	2.2	1.9	1.53546	2.39261	2.2	1.9	1.71176	2.24855	2.3	1.9
29	1.26907	2.68563	2.2	1.9	1.53391	2.39238	2.2	1.9	1.7099	2.24845	2.3	1.9
30	1.26797	2.68521	2.2	1.9	1.53247	2.39216	2.2	1.9	1.70818	2.24836	2.3	1.9

Once h_1 is obtained, we get d_1 using (3.5) by setting

$$1-\int_{-\infty}^{\infty} F_{N_0}^{k-1}(t_0+h_1(1-d_1))F_{N_0}(t_0+h_1(1-d_2))f_{N_0}(t_0)\,dt_0=\tau_1^*. \qquad (5.4)$$

Then, we search for d_2 and h_2 so that the probability requirements (5.1) and (5.2) are satisfied. Evaluations of integrals in (5.1) and (5.2) were done via simulations with 10^7 iterations. All computations and simulations were done by using the statistical software `R`. The code of the program is available per request from the first author.

Table 1 includes the (h_1, d_1, h_2, d_2) values that satisfy (5.1) and (5.2) for various values of N_0 and k, when $M_0 = 10$, $\alpha = .05$, $1 - \beta = .90$, $\tau_0^* = .70$, $\tau_1^* = .93$ and the searching range for d_2 and h_2 being 1.5(.1)2.7.

6 Optimal Design Parameters and Sample Size Comparison

We would like to obtain design parameters that minimize the expected total sample size $E(n_{\mathrm{T}})$ subject to the probability constrains (3.2) and (3.3). The expected total sample size depends on the mean configuration through $P($ continue at stage II). Since we would like to obtain a design that performs well both under H_0 and LFC, we will redefine $E(n_{\mathrm{T}})$ as the simple average $\frac{E(N|H_0)+E(N|\mathrm{LFC})}{2}$. Then, lower and upper bounds for $E(n_{\mathrm{T}})$ ((4.3) and (4.4)) become

$$\begin{aligned}\mathrm{LB} &= \mathrm{LB_I}+\frac{1}{2}[P(\text{continue to stage II}|H_0)\\ &\quad +P(\text{continue to stage II(testing)}|\mathrm{LFC})]\mathrm{LB_{II}}\end{aligned}$$

Table 2: Comparison of sample sizes for P_{ST} and P_{S} when $k = 2$, $\alpha = .05$, $1 - \beta = .90$

	Procedure P_{ST}								Procedure P_{S}	
N_0	α_{sim}	$1-\beta_{\mathrm{sim}}$	LB_{H_o}	UB_{H_o}	$\mathrm{LB}_{\mathrm{LFC}}$	$\mathrm{UB}_{\mathrm{LFC}}$	LB	UB	LB	UB
5	0.0469	0.8925	72.41	82.22	92.38	118.53	82.39	100.37	129.64	132.21
6	0.0500	0.9027	68.57	79.14	89.95	118.76	79.26	98.95	114.02	116.47
7	0.0498	0.9016	65.95	75.71	86.25	112.80	76.10	94.26	105.55	107.87
8	0.0505	0.8999	64.57	73.58	83.87	108.24	74.22	90.91	100.52	102.71
9	0.0496	0.9003	64.38	73.31	83.67	107.97	74.02	90.64	97.41	99.49
10	0.0489	0.9006	64.64	73.49	83.94	108.16	74.29	90.83	95.49	97.49
11	0.0507	0.8984	64.80	72.91	83.15	105.26	73.97	89.09	94.34	96.28
12	0.0503	0.8987	65.67	73.69	84.03	106.05	74.85	89.87	93.71	95.61
13	0.0500	0.8988	66.77	74.70	85.13	107.05	75.95	90.88	93.44	95.31
14	0.0496	0.8991	68.07	75.91	86.43	108.26	77.25	92.08	93.45	95.28
15	0.0494	0.8992	69.54	77.29	87.90	109.64	78.72	93.47	93.67	95.46
20	0.0487	0.8996	78.74	86.16	97.10	118.51	87.92	102.34	97.05	98.59
25	0.0484	0.8998	89.95	97.10	108.31	129.45	99.13	113.27	103.35	104.58
30	0.0482	0.8998	102.81	109.70	121.17	142.05	111.99	125.87	111.78	112.79

$$\begin{aligned} \text{UB} &= \text{UB}_{\text{I}} + \frac{1}{2}[P(\text{continue to stage II}|H_0) \\ &\quad + P(\text{continue to stage II(testing)}|\text{LFC})]\text{UB}_{\text{II}} \end{aligned}$$

Since we do not have an analytic expression for the total expected sample size, we will look into minimizing its upper bound (UB) as defined above.

Tables 2, 3, and 4 provide the lower and upper bounds for the total expected sample size $E(n_{\text{T}})$, the total expected sample size under the null hypothesis $E(N|H_0)$, and total expected sample size under the least favorable configuration $E(N|\text{LFC})$ for cases considered in Section 5 and when $\sigma_i = i + 1, i = 0, 1, \ldots, k$. The expected sample sizes were obtained via simulations with 10^7 iterations.

We notice that, for fixed k, these upper bounds first decrease as N_0 increases, then they decrease. Thus, we can find the optimal N_0 in terms of the total sample size and so we can find an optimal design, that is the design that produces the smallest upper bound on the total expected sample size. For example, for the case when $k = 2$, the optimal design happens when $N_0 = 11$, in which case UB = 89.09.

Table 3: Comparison of sample sizes for P_{ST} and P_{S} when $k = 3$, $\alpha = .05$, $1 - \beta = .90$

	Procedure P_{ST}								Procedure P_{S}	
N_0	α_{sim}	$1-\beta_{\text{sim}}$	LB_{H_o}	UB_{H_o}	LB_{LFC}	UB_{LFC}	LB	UB	LB	UB
6	0.0468	0.8925	148.54	169.94	168.52	228.57	158.53	199.25	272.38	275.91
7	0.0463	0.8915	140.87	160.68	159.91	215.29	150.39	187.98	249.34	252.74
8	0.0496	0.9033	137.68	161.00	159.06	225.49	148.37	193.24	235.11	238.38
9	0.0500	0.9024	134.52	156.15	154.83	216.15	144.68	186.15	225.68	228.84
10	0.0493	0.9031	132.93	154.49	153.23	214.49	143.08	184.49	219.16	222.23
11	0.0508	0.9012	131.56	151.49	150.86	207.20	141.21	179.35	214.52	217.52
12	0.0503	0.9017	131.20	151.06	150.50	206.76	140.85	178.91	211.16	214.11
13	0.0499	0.9020	131.25	151.02	150.54	206.73	140.90	178.88	208.69	211.61
14	0.0495	0.9022	131.63	151.32	150.92	207.02	141.27	179.17	206.89	209.77
15	0.0492	0.9024	132.28	151.89	151.58	207.60	141.93	179.74	205.59	208.44
16	0.0490	0.9027	133.18	152.70	152.47	208.41	142.83	180.56	204.67	207.50
17	0.0488	0.9029	134.28	153.72	153.58	209.43	143.93	181.58	204.08	206.87
18	0.0488	0.9031	135.56	154.93	154.85	210.64	145.20	182.78	203.75	206.50
19	0.0484	0.9030	136.98	156.29	156.28	212.00	146.63	184.15	203.65	206.36
20	0.0485	0.9031	138.54	157.79	157.84	213.50	148.19	185.65	203.76	206.41
21	0.0482	0.9034	140.21	159.41	159.51	215.11	149.86	187.26	204.05	206.65
22	0.0481	0.9034	141.98	161.12	161.27	216.83	151.62	188.98	204.50	207.04
23	0.0481	0.9035	143.83	162.93	163.12	218.63	153.48	190.78	205.11	207.60
24	0.0503	0.9011	145.31	162.85	163.67	214.46	154.49	188.66	205.87	208.29
25	0.0503	0.9011	147.32	164.81	165.67	216.41	156.50	190.61	206.76	209.12

Table 4: Comparison of sample sizes for P_{ST} and P_{S} when $k = 4$, $\alpha = .05$, $1 - \beta = .90$

	Procedure P_{ST}								Procedure P_{S}	
N_0	α_{sim}	$1-\beta_{\mathrm{sim}}$	LB_{H_o}	UB_{H_o}	$\mathrm{LB}_{\mathrm{LFC}}$	$\mathrm{UB}_{\mathrm{LFC}}$	LB	UB	LB	UB
6	0.0456	0.8956	288.23	329.74	310.27	430.40	299.25	380.07	539.99	544.56
7	0.0498	0.9042	272.86	318.53	296.61	429.89	284.73	374.21	490.94	495.40
8	0.0494	0.9045	262.05	304.59	284.58	408.36	273.32	356.47	460.47	464.80
9	0.0500	0.9040	254.66	294.21	276.05	390.68	265.35	342.44	439.98	444.20
10	0.0490	0.9049	249.92	289.41	271.31	385.88	260.61	337.65	425.46	429.58
11	0.0503	0.9035	246.07	282.68	266.38	372.16	256.22	327.42	414.78	418.82
12	0.0496	0.9042	243.73	280.28	264.04	369.76	253.88	325.02	406.70	410.68
13	0.0492	0.9046	242.13	278.61	262.44	368.09	252.29	323.35	400.45	404.39
14	0.0489	0.9050	241.12	277.52	261.42	367.00	251.27	322.26	395.54	399.46
15	0.0508	0.9030	240.09	273.71	259.39	356.48	249.74	315.09	391.65	395.53
16	0.0503	0.9032	239.94	273.47	259.23	356.25	249.58	314.86	388.53	392.39
17	0.0503	0.9036	240.11	273.57	259.41	356.34	249.76	314.95	386.03	389.87
18	0.0459	0.8910	239.65	267.53	257.01	336.84	248.33	302.19	384.05	387.84
19	0.0500	0.9039	241.26	274.57	260.55	357.34	250.90	315.96	382.48	386.24
20	0.0498	0.9040	242.16	275.40	261.45	358.18	251.80	316.79	381.28	385.00
22	0.0495	0.9042	244.46	277.60	263.75	360.37	254.10	318.98	379.78	383.40
24	0.0492	0.9045	247.28	280.33	266.58	363.11	256.93	321.72	379.27	382.78
26	0.0489	0.9046	250.52	283.49	269.82	366.26	260.17	324.87	379.58	382.97
28	0.0489	0.9047	254.11	286.98	273.4	369.76	263.76	328.37	380.59	383.86
30	0.0487	0.9050	258.02	290.80	277.32	373.58	267.67	332.19	382.17	385.35

We also compare our hybrid selection and testing procedure P_{ST} to the pure selection procedure defined by Buzaianu and Chen (2015), denoted by P_{S}, which we describe below. The goal of procedure P_{S} is to select the population having the largest mean μ_k, given that it is larger than the mean of the control population ($\mu_k > \mu_0$). Before starting the experiment, the researcher should specify five constants $\delta_0^*, \delta_1^*, \delta_2^*, P_1^*, P_2^*$, so that $0 < \delta_1^*, \delta_2^* < \infty$, $-\delta_1^* < \delta_0^* < \infty$, $2^{-k} < P_0^* < 1$, and $(1 - 2^{-k})k^{-1} < P_1^* < 1$. We denote by π_0 the event of selecting the control population π_0 and by Π_k the event of selecting the population with mean μ_k. Then, the proposed procedure is guaranteed to satisfy the following probability requirements:

$$P(\pi_0|\vec{\mu} : \mu_k \leq \mu_0 - \delta_0^*) \geq P_0^*. \tag{6.1}$$

$$P(\pi_k|\vec{\mu} : \mu_k \geq \mu_0 + \delta_1^*, \mu_k \geq \mu_{k-1} + \delta_2^*) \geq P_1^*. \tag{6.2}$$

For prespecified integers N_0 and positive constants h and c, the procedure is as follows:

Procedure P_{S}

Stage 1. Take an initial random sample $X_{i1}, X_{i2}, \ldots, X_{iN_0}$ of size $N_0 \geq 2$ from each population $\pi_i, i = 0, 1, \ldots, k$, and compute the following:

$$\overline{X}_{i(N_0)} = \frac{\sum_{j=1}^{N_0} X_{ij}}{N_0}$$

$$S^2_{i(N_0)} = \frac{\sum_{j=1}^{N_0} (X_{ij} - \overline{X}_{i(N_0)})^2}{N_0 - 1}.$$

$n_i = \max\left\{N_0+1, \left[\frac{h^2 S^2_{i(N_0)}}{(\delta_0^* + c)^2}\right] + 1\right\}$, where $[x]$ denotes the smallest integer $\geq x$.

Stage 2. Take $n_i - N_0$ additional independent observations X_{i,N_0+1}, $X_{i,N_0+2}, \ldots, X_{i,n_i}$ from π_i and define $\widetilde{X}_i = \sum_{j=1}^{n_i} a_{ij} X_{ij}$ where the $a'_{ij}s$ are chosen so that

$$\sum_{j=1}^{n_i} a_{ij} = 1, a_{i1} = \ldots = a_{iN_0}, \text{ and } S^2_{i(N_0)} \sum_{j=1}^{n_i} a^2_{ij} = \frac{(\delta_0^* + c)^2}{h^2}.$$

If $\max_{i=\overline{1,k}}(\widetilde{X}_i) > \widetilde{X}_0 + c$, then select the population π_ν which produced the maximum and assert that it has mean μ_k. Else, select the control population.

The two procedures have slightly different problem formulations. The goal of procedure P_{ST} is to select an acceptable population, while the goal of procedure P_{S} is to select the best population. Moreover, Buzaianu and Chen (2015)'s preference zone consists of all vectors $\vec{\mu}$ with $\mu_k \geq \mu_0 + \delta_1^*$ and $\mu_k \geq \mu_{k-1} = \delta_2^*$, while our preference zone consists of all vectors $\vec{\mu}$ so that there is at least one $\mu_i \geq \mu_0 + \delta_2^*$ and no μ_i lies in $\mu_0 + \delta_1^*$, $\mu_0 + \delta_2^*$. However, for the special case when Buzaianu and Chen (2015)'s $\delta_0^* = 0$, their(δ_1^*) = our(δ_2^*), their$(\delta_1^* - \delta_2^*)$ = our(δ_1^*), our(α) = their$(1 - P_1^*)$, our$(1 - \beta)$ = their(P_1^*), the two problem formulations are comparable.

The entries in Tables 2, 3, and 4 indicate that overall, the hybrid selection and testing procedure P_{ST} is superior to the pure selection procedure P_{S} with respect to the expected total sample size. For example, when $k = 4$, the optimal design for the hybrid procedure happens at $N_0 = 18$, where the lowest upper bound on the expected total sample size is attained (UB = 302.19), while the optimal design for the pure selection procedure happens at $N_0 = 24$ where the minimum on the the upper bound on the expected total sample

size is attained at UB = 328.78. However, as N_0 increases, under the LFC, the upper bound on the expected sample size for procedure P_{ST} is higher than that of procedure P_{S}, while under H_0, the upper bound on the expected sample size for procedure P_{ST} is lower than that of procedure P_{S}.

7 An Example

To illustrate our procedure, we will consider an example based on an experiment reported by Ramsey and Schafer (1997) from the original study of Weindruch et al. (1986). In this study, female mice were randomly assigned to various treatment groups to investigate whether restricting dietary intake increases life expectancy. The diet treatments were as follows:

Treatment 0. "N/N85": mice fed normally before and after weaning. After weaning, ration was controlled at 85 kcal/week. This is the control group because caloric intake is held reasonably constant.

Treatment 1. "N/R50": normal diet before weaning and reduced calorie diet (50 kcal/week) after weaning.

Treatment 2. "R/R50": reduced calorie diet of 50 kcal/week both before and after weaning.

Treatment 3. "N/R50 lopro": normal diet before weaning, restricted diet (50 kcal/week) after weaning and dietary protein content decreased with advancing age.

Treatment 4. "N/R40": normal diet before weaning and reduced diet (40 kcal/week) after weaning.

Let μ_i denote the average lifetime expectancy under treatment i. The goal is to select one of the experimental treatments if its mean is sufficiently larger than the mean of the control treatment. Let us specify $\delta_0^* = 0$, $\delta_1^* = \delta_2^* = \delta^* = 2.00$, $\alpha = .05$, and $1 - \beta = .90$. We decide to start with $N_0 = 8$ observations from each treatment at the first step of the selection stage and $M_0 = 10$ at the first step of the testing stage, if needed. Consulting Table 1 with $k = 4$, $N_0 = 8$, $\alpha = .05$, and $1 - \beta = .90$, we obtain $h_1 = 1.873714$, $d_1 = 2.25915$, $h_2 = 2.6$, $d_2 = 1.9$, from which we obtain $y_1 = 0.885287$ and $y_2 = 1.052632$.

STAGE I(SELECTION) After taking eight observations from each treatment at the first step of the selection stage (Table 5), we obtain the following:

Treatment 0: $\overline{X}_0 = 39.0750$ $S^2_{0(N_0)} = 2.389$, $n_0 = 11$.
Treatment 1: $\overline{X}_1 = 48.2875$ $S^2_{1(N_0)} = 0.7641$, $n_1 = 9$.
Treatment 2: $\overline{X}_2 = 51.0500$ $S^2_{2(N_0)} = 0.0784$, $n_2 = 9$.

Table 5: The data—stage I (selection)

Step	Treatment 0	Treatment 1	Treatment 2	Treatment 3	Treatment 4
1	42.3	49.7	51.9	50.7	54.6
	40.1	49.3	51.7	50.6	54
	39.5	48.6	51.4	50.5	53.8
	38.6	48.3	51.3	50.3	53.3
	38.4	48	50.9	50.1	52.9
	38.3	47.7	50.5	50.1	52.7
	37.8	47.5	50.5	50	52.5
	37.6	47.2	50.2	50	52.4
2	37.4	47.1	50	49.8	52
	37.3				
	36.8				

Treatment 3: $\overline{X}_3 = 50.2875$ $S^2_{3(N_0)} = 0.078$, $n_3 = 9$.

Treatment 4: $\overline{X}_4 = 53.2750$ $S^2_{4(N_0)} = 0.6279$, $n_4 = 9$.

Therefore, at the second step of the selection stage, we will collect three more observations from the control treatment and one from each of the experimental treatments. The data at the second step stage are listed in Table 5. Then, after collecting these observations, we obtain the following:

Treatment 0: $c_0^I = 0.8031, \widetilde{X}_0 = 38.6993$.

Treatment 1: $c_1^I = 0.4875, \widetilde{X}_1 = 47.6789$.

Treatment 2: $c_2^I = 0.2409, \widetilde{X}_2 = 50.2530$.

Treatment 3: $c_3^I = -0.6712, \widetilde{X}_3 = 49.4728$.

Treatment 4: $c_4^I = 0.4225, \widetilde{X}_4 = 52.5388$.

Based on this information, Treatment 4 and the Control will continue to the Testing Stage.

STAGE II(TESTING) At the first step of this stage, we collect $M_0 = 10$ more observations (the data are in Table 6) form each of Treatment 4 and the Control and obtain the following:

Treatment 0: $\overline{W}_0 = 35.7182$ $S^2_{0(N_0)} = 0.369$, $m_0 = 11$.

Treatment 4: $\overline{W}_4 = 50.1364$ $S^2_{4(N_0)} = 1.3138$, $m_4 = 11$.

Based on these, at the second step of the testing stage, we will collect one more observation from the control treatment and one from the selected experimental treatment.

The data at the second step stage are listed in Table 6. Then, after collecting these observations, we obtain the following:

Treatment 0: $c_0^{II} = 0.3424, \widetilde{W}_0 = 35.0211$

Table 6: The data—stage II (testing)

Step	Treatment 0	Treatment 4
1	36.5	51.8
	36.5	51.3
	36.5	51.3
	36.4	51
	35.9	50.8
	35.5	50.3
	35.5	50.1
	35.3	49.8
	35.3	48.7
	34.9	48.3
2	34.6	48.1

Treatment 4: $c_4^{II} = 0.7337, \widetilde{W}_4 = 49.7434$

Then, $\frac{\widetilde{X}_4+\widetilde{W}_4}{2} = \frac{52.5388+49.7434}{2}$ and $\frac{\widetilde{X}_0+\widetilde{W}_0}{2}+y_2 = \frac{38.6993+35.0211}{2}+1.052632$.

Since $\frac{\widetilde{X}_4+\widetilde{W}_4}{2} > \frac{\widetilde{X}_0+\widetilde{W}_0}{2} + y_2$, experimental Treatment 4 is selected as an acceptable treatment.

8 Discussion

The two fundamental underlying distributions in medical trials, as well as in most statistical applications, are binomial and normal. Even for trials with binomial outcomes, clinical practitioners use normal approximations to binomial in calculating the confidence limits in interval estimation, in calculating probabilities of errors and powers in hypothesis testing, and in calculating the probability of correct selection in ranking and selection problems. For example, Thall et al. (1988), who considered a two-stage selection and testing design for comparing experimental clinical trials to a control treatment, all with binomial outcomes, used a variance-stabilizing transformation to express the probability of type I error and the power in terms of the normal density and distribution functions.

The procedure we propose is a hybrid selection and testing procedure for comparing experimental treatments to a control treatment, all with normal outcomes with unknown means, which are the parameters of interest, and unknown variances, which are nuisance parameters. We show in this paper that the Stein-type two-sample methodology can be adopted in a more complex two-stage selection and testing design. An important feature of our procedure is that probabilities of errors, power, and the probability

of correct selection are independent of the unknown variances under our design. The design also allows for early termination if none of the experimental treatments seems promising. Our two-sample two-stage design is adaptive in the sense that the second sample sizes in both the selection stage and testing stage are calculated from formulas based on the estimates of the unknown variances that are obtained from the first sample at each stage.

Although our work lies mainly in the frequentist domain, it would be desirable to formulate a Bayesian approach for the hybrid selection/testing two-stage design as Bayesian adaptive methodology has become more used and more useful in clinical trials in recent years. Berry et al. (2010) give a thorough review and examples of Bayesian approach in adaptive clinical trials. It is interesting to observe that the comments posted in Cross Validated for the topic "Bayesian and frequentist reasoning in plain English" (https://stats.stackexchange.com/questions/22/Bayesian-and-frequentist-reasoning-in-plain-english) predominately consider the two approaches as two completely different sampling plans, exclusive to each other. We also see several hundred informative comparative discussions on Bayesian approach versus frequentist approach in google search. The speech "Bayesians and Frequentists" given in 2008 at 37$^{\text{th}}$ Annual Meeting of the American College of Clinical Pharmacology by George Casella gave a clear and fair comparison between the two approaches (http://www.stat.ufl.edu/archived/casella/Talks/BayesRefresher.pdf). Casella's speech used an example in amino-glycoside-associated nephrotoxicity (AAN) to illustrate the basic differences between Bayesian approach and frequentist approach. Similarly, in this paper, the main differences between our approach and a possible Bayesian approach would be (1) there a fixed unknown parameter of interest μ versus a varied (or distributional) μ under the Bayesian approach (see pages 7–9 and 12 of the speech) and (2) the data are repeatable in our study and the data are fixed for a Bayesian approach (see pages 8 and 9 of the speech). The pro of our approach is that our design could be implemented to new clinical trials without any prior knowledge of the parameters that we are inferencing. Since the sample size of our second sample is a random number dependent of the first sample result, the fixed data requirement for Bayesian approach would not be possible. The con of our approach is that our design might take longer and more expensive to complete than the Bayesian approach that is known to require shorter time to complete. Since the two approaches are basically incomparable due to their assumptions on the parameters and the type of the data to be analyzed, a Bayesian approach to the two-sample two-stage design for clinical trials will be a very challenging research problem.

Appendix

Proof of Lemma 3.3.

$$P(\text{select an acceptable population } \vec{\mu})$$
$$= P\left(\max_{i\in A}\widetilde{X}_i > \max_{j\in A^c}\widetilde{X}_j, \max_{i\in A}\widetilde{X}_i \geq \widetilde{X}_0 + y_1, \frac{\widetilde{X}_\nu + \widetilde{W}_\nu}{2} > \frac{\widetilde{X}_0 + \widetilde{W}_0}{2} + y_2\right),$$

where ν is the index so that $X_\nu = \max\{\widetilde{X}_1, \ldots \widetilde{X}_k\}$.

In order to use Mahamunulu's result, we define a function $\Psi = \Psi(\widetilde{x}_0, \widetilde{x}_1, \ldots, \widetilde{x}_k, \widetilde{w}_0, \widetilde{w}_\nu)$ as follows:

$$\Psi(\widetilde{x}_0, \widetilde{x}_1, \ldots, \widetilde{x}_k, \widetilde{w}_0, \widetilde{w}_\nu) = 1, \text{ if } \max_{i\in A}\widetilde{X}_i > \max_{j\in A^c}\widetilde{X}_j, \max_{i\in A}\widetilde{X}_i \geq \widetilde{X}_0 + y_1$$
$$\text{and } \frac{\widetilde{X}_\nu + \widetilde{W}_\nu}{2} > \frac{\widetilde{X}_0 + \widetilde{W}_0}{2} + y_2$$
$$= 0, \text{else}$$

a. Let $i \in A$ be fixed. First, we show that ψ is a nondecreasing function of $\widetilde{x}_i$ when $\widetilde{w}_\nu, \widetilde{w}_0$ and all other $\tilde{x_j}'s$ are held fixed. Let $\tilde{x_i} < \tilde{x_i^*}$ and fix $\widetilde{w}_\nu, \widetilde{w}_0$ and all other $\tilde{x_j}'s$. For ease of notation, we will denote $\Psi(\widetilde{x}_0, \widetilde{x}_1, \ldots, \widetilde{x}_k, \widetilde{w}_0, \widetilde{w}_\nu)$ by $\Psi = \Psi(., \widetilde{x}_i)$.

If $\Psi(., \widetilde{x}_i) = 1$, then

$$\begin{aligned}\max(\widetilde{x}_{k-m+1}, \ldots, \widetilde{x}_i^*, \ldots, \widetilde{x}_k) &\geq \max(\widetilde{x}_{k-m+1}, \ldots, \widetilde{x}_i, \ldots, \widetilde{x}_k)\\ &> \max(\widetilde{x}_1, \ldots, \widetilde{x}_{k-m}),\end{aligned}$$

$\max(\widetilde{x}_{k-m+1}, \ldots, \widetilde{x}_i^*, \ldots, \widetilde{x}_k) \geq \max(\widetilde{x}_{k-m+1}, \ldots, \widetilde{x}_i, \ldots, \widetilde{x}_k) > \widetilde{x}_0 + y_1$ and

$$\frac{max\{\widetilde{x}_1, \ldots, \widetilde{x}_i^*, \ldots \widetilde{x}_k\} + \widetilde{w}_\nu}{2} \geq \frac{max\{\widetilde{x}_1, \ldots, \widetilde{x}_i, \ldots \widetilde{x}_k\} + \widetilde{w}_\nu}{2} > y_2 + \frac{x_0 + w_0}{2}.$$

Therefore, $\Psi(., \widetilde{x}_i^*) = 1$ and so $\Psi(., \widetilde{x}_i) \leq \Psi(., \widetilde{x}_i^*)$. If $\Psi(., \widetilde{x}_i) = 0$, then obviously $\Psi(., \widetilde{x}_i) \leq \Psi(\widetilde{x}_i^*)$. Consequently, Ψ is a nondecreasing function of $\widetilde{x}_i$, when $\widetilde{w}_\nu, \widetilde{w}_0$ and all other $\tilde{x_j}'s$ are held fixed. By Mahamunulu's Lemma, $E[\Psi|\widetilde{x}_0, \widetilde{x}_1, \ldots, \widetilde{x}_{i-1}, \widetilde{x}_{i+1}, \ldots, \widetilde{x}_k, \widetilde{w}_\nu, \widetilde{w}_0]$ is a nondecreasing function of μ_i and since

$$P(\text{select an acceptable population})$$
$$= E[\Psi(\widetilde{X}_0, \widetilde{X}_1, \ldots, \widetilde{X}_k, \widetilde{W}_\nu, \widetilde{W}_0)]$$
$$= E[E[\psi(\widetilde{X}_0, \widetilde{X}_1, \ldots, \widetilde{X}_k, \widetilde{W}_\nu, \widetilde{W}_0)|(\widetilde{x}_0, \widetilde{x}_1, \ldots, \widetilde{x}_{i-1}, \widetilde{x}_{i+1}, \ldots, \widetilde{x}_k, \widetilde{w}_\nu, \widetilde{w}_0)]]$$

we obtain that P(select an acceptable population) is a nondecreasing function of μ_i.

b. Let $i \in A^c$ be fixed. First, we show that ψ is a nonincreasing function of $\widetilde{x}_i$ when $\widetilde{w}_\nu, \widetilde{w}_0$ and all other $\tilde{x_j}$ are held fixed. Let $\tilde{x_i} < \tilde{x_i^*}$. If $\Psi(.,\widetilde{x}_i^*) = 1$, then $\max(\widetilde{x}_{k-m+1}, \ldots, \widetilde{x}_k) > \widetilde{x}_0 + y_1$ and $\max(\widetilde{x}_{k-m+1}, \ldots, \widetilde{x}_k) > \max(\widetilde{x}_1, \ldots, \widetilde{x}_i^*, \ldots, \widetilde{x}_{k-k}) \geq \max(\widetilde{x}_1, \ldots, \widetilde{x}_i, \ldots, \widetilde{x}_{k-m})$. From here, we conclude that $\nu \in A$, and so $\tilde{x_\nu}$ does not depend on the value taken by $\widetilde{x}_i$, and so $\frac{\widetilde{w}_\nu + \widetilde{w}_\nu}{2} > \frac{max\{\widetilde{x}_1, \ldots, \widetilde{x}_i^*, \ldots \widetilde{x}_k\} + \widetilde{w}_\nu}{2} > y_2 + \frac{\tilde{x_0} + \widetilde{w}_0}{2}$. Therefore, $\Psi(.,\widetilde{x}_i^*) = 1$ and consequently, $\Psi(.,\widetilde{x}_i) \geq \Psi(.,\widetilde{x}_i^*)$. If $\Psi(.,\widetilde{x}_i^*) = 0$, then obviously $\Psi(.,\widetilde{x}_i) \geq \Psi(.,\widetilde{x}_i^*)$. We obtain that Ψ is a nondecreasing function of $\widetilde{x}_i$, when all other $\widetilde{x}_j' s$ are held fixed. By Mahamunulu's Lemma $E[\Psi|\widetilde{x}_0, \widetilde{x}_1, \ldots, \widetilde{x}_{i-1}, \widetilde{x}_{i+1}, \ldots, \widetilde{x}_k, \widetilde{w}_\nu, \widetilde{w}_0]$ is a nonincreasing function of μ_i. Since

$$
\begin{aligned}
&P(\text{select an acceptable population}) \\
&= E[\Psi(\widetilde{X}_0, \widetilde{X}_1, \ldots, \widetilde{X}_k, \widetilde{W}_\nu, \widetilde{W}_0)] \\
&= E[E[\psi(\widetilde{X}_0, \widetilde{X}_1, \ldots, \widetilde{X}_k, \widetilde{W}_\nu, \widetilde{W}_0)|(\widetilde{x}_0, \widetilde{x}_1, \ldots, \widetilde{x}_{i-1}, \widetilde{x}_{i+1}, \ldots, \widetilde{x}_k, \widetilde{w}_\nu, \widetilde{w}_0)]]
\end{aligned}
$$

we conclude that P(select an acceptable population) is a nonincreasing function of μ_i.

Proof of Theorem 3.4. A typical configuration in the preference zone is

$$\mu_1 \leq \mu_2 \leq \ldots \leq \mu_{k-t} \leq \mu_0 + \delta_1^* < \mu_0 + \delta_2^* = \mu_{k-t+1} \leq \ldots \leq \mu_k. \tag{8.1}$$

where $1 \leq m \leq k$, the lower bound on m following from the fact that at least one $\mu_i \geq \mu_0 + \delta_2^*$. From Lemma 3.3 P(CS) is minimized at the configuration

$$\mu_1 \leq \mu_2 = \ldots = \mu_{k-t} = \mu_0 + \delta_1^* < \mu_0 + \delta_2^* = \mu_{k-t+1} = \ldots \leq \mu_k.$$

It remains to show that over all configurations (8.1), the minimum power is achieved when m=1. Consider all mean vectors of the form (8.1). For each $m = 1, 2, \ldots, k$, let $A_m = \{k - m + 1, \ldots, k\}$ and $A_m^c = \{1, 2, \ldots, k - m\}$. If $m_1 < m_2$, then, $A_{m_1} \subset A_{m_2}$ and $A_{m_1}^c \supset A_{m_2}^c$. Then for any fixed values $(\widetilde{x}_0, \widetilde{x}_1, \ldots, \widetilde{x}_k)$, we have

$$max_{i \in A_{m_1}} \tilde{x_i} \leq max_{j \in A_{m_2}} \tilde{x_j} \text{ and}$$

$$max_{i \in A_{m_1}^c} \tilde{x_i} \geq max_{j \in A_{m_2}^c} \tilde{x_j}$$

Therefore $\{max_{i \in A_{m_1}} \tilde{x_i} \geq max_{j \in A_{m_1}^c} \tilde{x_j}\} \subset \{max_{i \in A_{m_2}} \tilde{x_i} \geq max_{j \in A_{m_2}^c} \tilde{x_j}\}$.

Also, $\{max_{i \in A_{m_1}} \tilde{x_i} \geq y_1 + \widetilde{x}_0\} \subset \{max_{i \in A_{m_2}} \tilde{x_i} \geq y_1 + \widetilde{x}_0\}$.

Note that for any $(\widetilde{x}_1, \ldots, \widetilde{x}_k)$ satisfying $max_{i\in A_m}\tilde{x}_i \leq max_{j\in A_m^c}\tilde{x}_j$, the index ν such that $\widetilde{x}_\nu = max_{i=1,\ldots,k}(\tilde{x}_i)$ is in A, and so $\mu_\nu = \mu_0+\delta_2^*$. Therefore, the distributions of $\widetilde{X}_\nu$ and $\widetilde{W}_\nu$ do not depend on m, implying that the distribution of $\frac{\widetilde{X}_\nu+\widetilde{W}_\nu}{2}$ is independent of the value of m. Thus, when $m_1 < m_2$,

$$\begin{aligned} &P(max_{i\in A_{m_1}}\widetilde{X}_i > max_{j\in A_{m_1}^c}\widetilde{X}_j, max_{i\in A_{m_1}}\widetilde{X}_i > \widetilde{X}_0 \\ &\quad +y_1, \frac{\widetilde{X}_\nu+\widetilde{W}_\nu}{2} > \frac{\widetilde{X}_0+\widetilde{W}_0}{2}+y_2) \\ &\leq P(max_{i\in A_{m_2}}\widetilde{X}_i > max_{j\in A_{m_2}^c}\widetilde{X}_j, max_{i\in A_{m_2}}\widetilde{X}_i > \widetilde{X}_0 \\ &\quad +y_1, \frac{\widetilde{X}_\nu+\widetilde{W}_\nu}{2} > \frac{\widetilde{X}_0+\widetilde{W}_0}{2}+y_2) \end{aligned}$$

Therefore, $P(CS)$ is an increasing function of m and so the minimum happens when $m = 1$.

Proof of Theorem 3.5.

$$\begin{aligned} 1-\beta(\vec{\mu}) &\geq 1-\beta(LFC) \\ &= P(\text{select acceptable population}|LFC) \\ &= P(\text{select}\pi_k|LFC) \\ &= P(\widetilde{X_k} > \widetilde{X_k}, i=\overline{1,k-1}, \widetilde{X_k} > \widetilde{X_0}+y_1, \frac{\widetilde{X_k}+\widetilde{W_k}}{2} \\ &> \frac{\widetilde{X_0}+\widetilde{W_0}}{2}+y_2|LFC) \\ &= P(\frac{\widetilde{X_k}-\mu_k}{\frac{y_1}{h_1}} > \frac{\widetilde{X_i}-\mu_i}{\frac{y_1}{h_1}}+\frac{\mu_i-\mu_k}{\frac{y_1}{h_1}}, i=\overline{1,k-1}, \\ &\quad \frac{\widetilde{X_k}-\mu_k}{\frac{y_1}{h_1}} \geq \frac{\widetilde{X_0}-\mu_0}{\frac{y_1}{h_1}}+\frac{\mu_0-\mu_k}{\frac{y_1}{h_1}}+\frac{y_1}{\frac{y_1}{h_1}}, \\ &\quad \frac{y_1}{2h_1}(\frac{\widetilde{X_k}-\mu_k}{\frac{y_1}{h_1}})+\frac{y_2}{2h_2}(\frac{\widetilde{W_k}-\mu_k}{\frac{y_2}{h_2}}) \geq \frac{y_1}{2h_1}(\frac{\widetilde{X_0}-\mu_0}{\frac{y_1}{h_1}}) \\ &\quad +\frac{y_2}{2h_2}(\frac{\widetilde{W_0}-\mu_0}{\frac{y_2}{h_2}})+y_2-\mu_k+\mu_0|LFC) \end{aligned}$$

Let $T_i = \frac{\widetilde{X_i}-\mu_i}{\frac{y_1}{h_1}}, i=\overline{1,k-1}$ and $V_i = \frac{W_k-\mu_k}{\frac{y_2}{h_2}}, i=0,k$. Then, $T_1, T_2, \ldots, T_k$, V_0, V_k are independent random variables and have Student's t distributions

with $N_0 - 1$ degrees of freedom and V has a Student's t distribution with $M_0 - 1$ degrees of freedom. Therefore,

$$\begin{aligned}
1-\beta(\vec{\mu}) &\geq P(T_k > T_i + \frac{\mu_i-\mu_k}{\frac{y_1}{h_1}}, i=\overline{1,k-1}, T_k > T_0 + \frac{y_1+\mu_0-\mu_k}{\frac{y_1}{h_1}} \\
&\qquad (\frac{y_1}{2h_1})T_k + (\frac{y_2}{2h_2})V_k \geq (\frac{y_1}{2h_1})T_0 + (\frac{y_2}{2h_2})V_0 + y_2 - \mu_k + \mu_0 | LFC) \\
&= P(T_k > T_i - \frac{\delta_2^* - \delta_1^*}{\frac{y_1}{h_1}}, i=\overline{1,k-1}, T_k > T_0 + \frac{y_1-\delta_2^*}{\frac{y_1}{h_1}} \\
&\qquad (\frac{y_1}{2h_1})T_k + (\frac{y_2}{2h_2})V_k \geq (\frac{y_1}{2h_1})T_0 + (\frac{y_2}{2h_2})V_0 + y_2 - \delta_2^*) \\
&= P(T_k > T_i - \frac{\delta_2^* - \delta_1^*}{\frac{y_1}{h_1}}, i=\overline{1,k-1}, T_k > T_0 + \frac{y_1-\delta_2^*}{\frac{y_1}{h_1}} \\
&\qquad V_k \geq \frac{2h_2}{y_2}(-(\frac{y_1}{2h_1})T_k + (\frac{y_1}{2h_1})T_0 + (\frac{y_2}{2h_2})V_0 + y_2 - \delta_2^*)) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{\frac{(y_1-\delta_1^*)h_1}{y_1}+t_0}^{\infty} F_{n_0}^{k-1}(t_k + \frac{(\delta_2^*-\delta_1^*)h_1}{y_1}) F_{M_0}(\frac{h_2y_1}{y_2h_1}t_k - \frac{y_1h_2}{h_1y_2}t_0 \\
&\qquad - v_0 - \frac{2h_2}{y_2}(y_2-\delta_2^*)) \times f_{N_0}(t_k) f_{N_0}(t_0) f_{M_0}(v_0)\, dt_k\, dt_0\, dt_{v_0}
\end{aligned} \tag{8.2}$$

Since $\alpha = P(\pi_i \text{ is selected for some } i = \overline{1,k} | H_0)$, the size can be obtained by evaluating (8.2) at $\delta_1^* = \delta_2^* = 0$ and multiplying it by k. This gives

$$\begin{aligned}
\alpha = k \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{h_1+t_0}^{\infty} & F_{N_0}^{k-1}(t_k) F_{M_0}(\frac{h_2y_1}{y_2h_1}t_k - \frac{y_1h_2}{h_1y_2}t_0 - v_0 - 2h_2) f_{N_0}(t_k) f_{N_0} \\
& \times (t_0) f_{M_0}(v_0)\, dt_k\, dt_0\, dv_0
\end{aligned}$$

PROOF OF THEOREM 3.6. P(terminating early $|\vec{\mu}$)

$$\begin{aligned}
&= P(max_{i=\overline{1,k}}(\widetilde{X_i}) \leq \widetilde{X_0} + y_1 | \vec{\mu}) \\
&= P((\widetilde{X_i}) \leq \widetilde{X_0} + y_1, i = \overline{1,k} | \vec{\mu}) \\
&= P(\frac{\widetilde{X_i} - \mu_i}{\frac{y_1}{h_1}} < \frac{\widetilde{X_0} - \mu_0}{\frac{y_1}{h_1}} + \frac{\mu_0 - \mu_i}{\frac{y_1}{h_1}} + \frac{y_1}{\frac{y_1}{h_1}}, i = \overline{1,k} | \vec{\mu}) \\
&= P(T_i \leq T_0 + \frac{y_1 + \mu_0 - \mu_i}{\frac{y_1}{h_1}}, i = \overline{1,k} | \vec{\mu}) \\
&= \int_{-\infty}^{\infty} P(T_i \leq t_0 + \frac{y_1 + \mu_0 - \mu_i}{\frac{y_1}{h_1}}, i = \overline{1,k} | \vec{\mu}) f_{N_0}(t_0)\, dt_0
\end{aligned}$$

$$= \int_{-\infty}^{\infty} \prod_{i=1}^{k} P(T_i \le t_0 + \frac{y_1 + \mu_0 - \mu_i}{\frac{y_1}{h_1}} | \vec{\mu}) f_{N_0}(t_0)\, dt_0$$
$$= \int_{-\infty}^{\infty} \prod_{i=1}^{k} F_{N_0}(t_0 + \frac{y_1 + \mu_0 - \mu_i}{\frac{y_1}{h_1}} | \vec{\mu}) f_{N_0}(t_0)\, dt_0$$

where T_i's are Student's t random variables with $N_0 - 1$ degrees of freedom as defined earlier. We obtain that

$$\tau_0 = P(\text{early termination} | H_0)$$
$$= \int_{-\infty}^{\infty} P(T_i \le t_0 + h_1, i = \overline{1,k}) f(t_0)\, dt_0 = \int_{-\infty}^{\infty} F_{N_0}^{k}(t_0 + h_1) f_{N_0}(t_0)\, dt_0.$$

$$P(\text{continuation to STAGE II} | \vec{\mu})$$
$$1 - P(max_{i=\overline{1,k}}(\widetilde{X_i}) \le \widetilde{X_0} + y_1 | \vec{\mu})$$
$$= 1 - P((\widetilde{X_i}) \le \widetilde{X_0} + y_1, i = \overline{1,k} | \vec{\mu})$$
$$= 1 - P(\frac{\widetilde{X_i} - \mu_i}{\frac{y_1}{h_1}} > \frac{\widetilde{X_0} - \mu_0}{\frac{y_1}{h_1}} + \frac{\mu_0 - \mu_i}{\frac{y_1}{h_1}} + \frac{y_1}{\frac{y_1}{h_1}}, i = \overline{1,k} | \vec{\mu})$$
$$= 1 - P(T_i \le T_0 + \frac{y_1 + \mu_0 - \mu_i}{\frac{y_1}{h_1}}, i = \overline{1,k} | \vec{\mu})$$
$$= 1 - \int_{-\infty}^{\infty} P(T_i \le t_0 + \frac{y_1 + \mu_0 - \mu_i}{\frac{y_1}{h_1}}, i = \overline{1,k} | \vec{\mu}) f_{N_0}(t_0)\, dt_0$$
$$= 1 - \int_{-\infty}^{\infty} \prod_{i=1}^{k} P(T_i \le t_0 + \frac{y_1 + \mu_0 - \mu_i}{\frac{y_1}{h_1}} | \vec{\mu}) f_{N_0}(t_0)\, dt_0$$
$$= 1 - \int_{-\infty}^{\infty} \prod_{i=1}^{k} F_{N_0}(t_0 + \frac{y_1 + \mu_0 - \mu_i}{\frac{y_1}{h_1}} | \vec{\mu}) f_{N_0}(t_0)\, dt_0$$

where T_i's are Student's t random variables with $N_0 - 1$ degrees of freedom as defined earlier. We obtain that

$$\tau_1 = P(\text{continuation to STAGE II} | \text{LFC })$$
$$= 1 - \int_{-\infty}^{\infty} F_{N_0}^{k-1}(t_0 + h_1 - \frac{h_1 \delta_1^*}{y_1}) F_{N_0}(t_0 + h_1 - \frac{h_1 \delta_2^*}{y_1}) f_{N_0}(t_0)\, dt_0.$$

References

BECHHOFER, R.E. and TURNBULL, B.W. (1978). Two $(k+1)$-decision selection procedures for comparing k normal means with a specified standard. *J. Am. Stat. Assoc.* **73**, 385–392.

BECHHOFER, R.E., DUNNETT, C.W. and SOBEL, M. (1954). A two sample multiple decision procedure for ranking means of normal populations with a common unknown variance. *Biometrika* **41**, 170–176.

BERRY, S.M., CARLIN, B.P., LEE, J.J. and MULLER, P. (2010). *Bayesian adaptive methods for clinical trials*. CRC Press, Boca Raton.

BUZAIANU, E.M. and CHEN, P. (2015). A Stein-type two-sample procedure for comparing normal means. *Seq. Anal.* **34**, 441–460.

DANTZIG, G. (1940). On the non-existenice of tests of "student's" hypothesis having power functions independent of σ. *Ann. Mathemat. Statist.* **11**, 186–192.

DUDEWICZ, E.J. and DALAL, S.R. (1975). Allocation of observations in ranking and selection with unequal variances. *Sankhya* **37B**, 28–78.

DUNNETT, C.W. (1984). Selection of the best treatment in comparison to a control with an application to a medical trial. In *Design of experiments: ranking and selection*. (T. J. Santner and A. C. Tamhane, eds.). Dekker, New York.

MAHAMUNULU, D.M. (1967). Some fixed-sample ranking and selection problems. *Ann. Math. Stat.* **38**, 1079–1091.

RAMSEY, F.L. and SCHAFER, D.W. (1997). *The statistical sleuth: a course in methods of data analysis*, 3rd edn. Cengage Learning, Boston.

STEIN, C. (1945). A two-sample test for a linear hypothesis whose power is independent of the variance. *Ann. Mathemat. Statist.* **16**, 243–258.

THALL, P.F., SIMON, R. and ELLENBERG, S.S. (1988). Two-stage designs for comparative clinical trials. *Biometrika* **75**, 303–310.

THALL, P.F., SIMON, R. and ELLENBERG, S.S. (1989). A two-stage design for choosing among several experimental treatments and a control in clinical trials. *Biometrics* **45**, 537–547.

WEINDRUCH, R., WALFORD, R.L., FLIGIEL, S. and GUTHRIE, D. (1986). The retardation of aging in mice by dietary restriction: longevity, cancer, immunity and lifetime energy intake. *J. Nutr.* **116**, 641–54.

Elena M. Buzaianu
Department of Mathematics and Statistics, University of North Florida, 1 UNF Drive, Jacksonville, FL 32224, USA
E-mail: e.buzaianu@unf.edu

Pinyuen Chen
Syracuse University, Syracuse, NY, USA

Paper received: 4 November 2016.