**Estimation of the selected treatment mean in two stage drop-the-losers design**

Masihuddin (✉) and Neeraj Misra (✉)

Department of Mathematics & Statistics, Indian Institute of Technology Kanpur, Kanpur, Uttar Pradesh, India

**ABSTRACT**

A common problem faced in clinical studies is that of estimating the effect of the most effective (e.g. the one having the largest mean) treatment among $k (\geq 2)$ available treatments. The most effective treatment is adjudged based on numerical values of some statistic corresponding to the $k$ treatments. A proper design for such problems is the so-called "Drop-the-Losers Design (DLD)". We consider two treatments whose effects are described by independent Gaussian distributions having different unknown means and a common known variance. To select the more effective treatment, the two treatments are independently administered to $n_1$ subjects each and the treatment corresponding to the larger sample mean is selected. To study the effect of the adjudged more effective treatment (i.e. estimating its mean), we consider the two-stage DLD in which $n_2$ subjects are further administered the adjudged more effective treatment in the second stage of the design. We obtain some admissibility and minimaxity results for estimating the mean effect of the adjudged more effective treatment. The maximum likelihood estimator is shown to be minimax and admissible. We show that the uniformly minimum variance conditionally unbiased estimator (UMVUE) of the selected treatment mean is inadmissible and obtain an improved estimator. In this process, we also derive a sufficient condition for inadmissibility of an arbitrary location and permutation equivariant estimator and provide dominating estimators in cases, where this sufficient condition is satisfied. The mean squared error and the bias performances of various competing estimators are compared via a simulation study. A real data example is also provided for illustration purpose.

**ARTICLE HISTORY**

Received 8 June 2022
Accepted 10 February 2023

**KEYWORDS**

Two-stage drop-the-losers design; adaptive trial; mean squared error; UMVUE; admissible; minimax; inadmissible estimator

1. **Introduction**

In clinical and drug development studies, while evaluating the effects of multiple treatments (or drugs), a healthcare practitioner is often interested in choosing the most effective treatment out of $k (\geq 2)$ active treatments. The effectiveness of a treatment is evaluated on the basis of the characteristic (or a parametric function) associated with it. A treatment’s mean effect is usually a primary characteristic of interest. For selecting the most effective treatment, a natural selection rule is the one that selects the treatment corresponding to the largest (or smallest sample) mean as the most effective treatment. Such a selection rule is known to be optimal under quite general scenarios (see, for example (Bahadur and Goodman 1952), (Eaton 1967) and (Misra and Dhariyal 1994)). After choosing one of the treatments/drugs as the most effective one, a practical problem is that of estimating the mean effect of the chosen treatment. For this purpose, an adaptive design, popularly known as drop-the-losers design (DLD), has been employed by many researchers in biomedical sciences (see, (Thall et al. 1988), (Sill 2000) and (Chow and Chang 2006). An adaptive design typically consists of several phases. Data
analyses are performed at each stage, and adaptations are made based on updated information to achieve maximum likelihood of success. For a detailed informational review on developments in inference procedures for a two-stage adaptive trial setting, one may refer to (Bauer et al. 2016) and (Robertson et al. 2023).

A drop-the-losers design (DLD) is a multistage adaptive design, under which interim analyses are performed at each stage, and losers (e.g., inferior treatment groups) are dropped based on predetermined criteria. For an introduction of two-stage drop-the-losers methodology see (Thall et al. 1988) and (Thall et al. 1989). For further developments, one may refer to (Bauer and Kieser 1999), (Sampson and Sill 2005), (Wu et al. 2010) and (Lu et al. 2013).

(Cohen and Sackrowitz 1989) proposed an unbiased estimate for the mean of the adjudged best performing treatment using a two-stage adaptive design framework. They have derived an uniformly minimum variance conditionally unbiased estimator (UMVCUE) for the best performing candidate in the Gaussian setting with a common variance, the same sample size for each of the competing treatments at stage I and single observation at stage II. The cases of known and unknown common variances are studied separately. For the case of known variance, (Bowden and Glimm 2008) extended the results of (Cohen and Sackrowitz 1989) by considering the possibility of multiple observations (with number of observations different from that in Stage I) at stage II. They proposed a two-stage unbiased estimate for the best performing treatment at the interim. Their method is valid when the quantity of interest is the best, second best, or the 3rd best treatment among k treatments. In the Gaussian setting, for the case of common unknown variance, (Robertson and Glimm 2019) provided an UMVCUE which works for multiple selected candidates, as well as unequal stage one and stage two sample sizes. (Sill and Sampson 2007) extended the methodology of (Cohen and Sackrowitz 1989) to the bivariate normal case where the covariance matrix is assumed to be known.

In the literature, most of the work in two-stage adaptive design setting is either in hypothesis testing framework or on investigating estimators for the selected treatment mean when bias is the main criterion. In a two-stage context, however, proposing efficient estimators when mean squared error is the key criterion, has received relatively little attention in the literature. (Lu et al. 2013) considered the problem of estimating the mean of the selected normal population in two-stage adaptive design and have shown that for k ≥ 3 the naive estimator, which is the weighted average of the first and second stage sample means (with weights being proportional to the corresponding sample sizes) is not minimax under squared error and LINEX loss functions. In this article, building upon the work of (Lu et al. 2013), we consider the problem of estimating the mean of the selected treatment in a two-stage drop-the-losers setting and obtain various decision theoretic results under the criterion of mean squared error.

The paper is organized as follows: In Section 2, we introduce some notations and preliminaries that will be used throughout the paper. In Section 3, we prove that the naive estimator, which is the weighted average of the first and second stage sample means of adjudged effective treatment, is minimax and admissible for estimating the selected treatment mean. In Section 4, we derive a sufficient condition for inadmissibility of an arbitrary location and permutation equivariant estimator. Some additional estimators, having nice performances in terms of mean squared error, are also derived in Section 4 of the paper. In Section 5, we report a simulation study to numerically assess performances of the various competing estimators under the criterion of mean squared error and bias. In Section 6, we present a data analysis to illustrate implementation of the proposed methodology.

## 2. Preliminaries

We consider a Drop-the-Losers design which is divided into two stages, with a single data-driven selection made in the middle. In the first stage of the design, referred to as Stage I, the investigators would independently administer two treatments (say, τ₁ and τ₂) to n₁ subjects each. Suppose that the effects of two treatments τ₁ and τ₂ are characterized by two Gaussian populations $N(μ_i, σ^2_i), i = 1, 2$, having different unknown means $(μ_i ∈ (−∞, ∞), i = 1, 2)$ and a common known variance $σ^2(>0)$. The
treatment corresponding to \( \max \{ \mu_1, \mu_2 \} \) is considered to be the more effective treatment or better treatment. Let \( \bar{X}_i, i = 1, 2 \), be the sample means (treatment effect estimates) corresponding to the two populations. Consider the natural selection rule that selects the treatment yielding the larger sample mean as the better treatment. Let \( S \in \{1, 2\} \) denote the index of the selected treatment \( \tau_S \) (i.e. \( S = 1 \), if \( \bar{X}_1 > \bar{X}_2 \), \( S = 2 \), if \( \bar{X}_2 \geq \bar{X}_1 \)). Treatment \( \tau_S \) is then taken forward in the second stage, referred to as Stage II of the two-stage trial, for further confirmatory analysis. We draw another sample of size \( n_2 \) independently for the treatment \( \tau_S \) selected in Stage I. Let \( Y \) be the stage II sample mean from the selected treatment \( \tau_S \). The goal is to estimate \( \mu_S \), the treatment effect of the selected treatment. Clearly, \( \bar{X}_i \sim N\left( \mu_i, \frac{\sigma_i^2}{n_i} \right), i = 1, 2 \), are independently distributed and, conditioned on \( S \), \( Y \sim N\left( \mu_S, \frac{\sigma_S^2}{n_2} \right) \).

We will be using the following notations throughout the paper:

- \( \mathbb{R} \): the real line \((-\infty, \infty)\);
- \( \mathbb{R}^k \): the \( k \) dimensional Euclidean space, \( k \in \{2, 3, \ldots\} \);
- \( N(\lambda, \sigma^2) \): normal distribution with mean \( \lambda \in \mathbb{R} \) and standard deviation \( \sigma \in (0, \infty) \);
- \( N_2(\mu, \Sigma) \): the bivariate normal distribution with mean vector \( \mu \in \mathbb{R}^2 \) and positive definite covariance matrix \( \Sigma \);
- \( \phi(\cdot) \): probability density function (p.d.f.) of \( N(0, 1) \);
- \( \Phi(\cdot) \): cumulative distribution function (c.d.f.) of \( N(0, 1) \);
- For real numbers \( a \) and \( b \)

\[
I(a \geq b) = \begin{cases} 
1, & \text{if } a \geq b \\
0, & \text{if } a < b.
\end{cases}
\]

In addition to the notations introduced above, we make use of the following notations throughout the paper:

\[
X = (\bar{X}_1, \bar{X}_2, Y); \quad \mu = (\mu_1, \mu_2); \quad \Theta = \mathbb{R}^2; \quad S = \max(\bar{X}_1, \bar{X}_2) \text{ (maximum of } \bar{X}_1 \text{ and } \bar{X}_2); \quad 3_{-S} = \min(\bar{X}_1, \bar{X}_2) \text{ (minimum of } \bar{X}_1 \text{ and } \bar{X}_2); \quad D_1 = 3_{-S} - S; \quad D_2 = Y - S; \quad \theta_1 = \min(\mu_1, \mu_2); \quad \theta_2 = \max(\mu_1, \mu_2); \quad \theta = \theta_2 - \theta_1.
\]

Also, for any \( \mu \in \Theta \), \( \mathbb{P}_\mu(\cdot) \) will denote the probability measure induced by \( X = (\bar{X}_1, \bar{X}_2, Y) \), when \( \mu \in \Theta \) is the true parameter value, and \( \mathbb{E}_\mu(\cdot) \) will denote the expectation operator under the probability measure \( \mathbb{P}_\mu(\cdot), \mu \in \Theta \). Note that \( \theta \geq 0 \) and \( \mathbb{P}_\mu(D_1 \leq 0) = 1, \forall \mu \in \Theta \).

We consider estimation of

\[
\mu_S = \begin{cases} 
\mu_1, & \text{if } \bar{X}_1 \geq \bar{X}_2 \\
\mu_2, & \text{if } \bar{X}_1 < \bar{X}_2
\end{cases}
\]

under the squared error loss function

\[
L_X(\mu, a) = (a - \mu)^2, \mu \in \Theta, a \in \mathcal{A} = \mathbb{R}.
\]

Under the squared error loss function (2.2), the risk function (also referred to as the mean squared error) of an estimator \( \delta(\cdot) \) is defined by

\[
R(\mu, \delta) = \mathbb{E}_\mu((\delta(X) - \mu_S)^2), \mu \in \Theta.
\]

Suppose, we have a prior distribution (density) \( \Pi \) on \( \Theta \). Then, the Bayes risk of an estimator \( \delta \) with respect to prior \( \Pi \) is defined as

\[
r(\Pi, \delta) = \mathbb{E}_\Pi(R(\mu, \delta))
= \int_\Theta \int_\Xi L_X(\mu, \delta(x))f(x|\mu)\Pi(\mu)dx \, d\mu.
\]
An estimator \( \delta_{\Pi} \) that minimizes the Bayes risk \( r(\Pi, \delta) \) among all estimators \( \delta \) of \( \mu_S \) is called a Bayes estimator with respect to the prior \( \Pi \).

An estimator \( \delta(X) \) is said to be conditionally unbiased if for every \( \mu \in \Theta \),

\[
E_\mu((\delta(X)|S = s)) = \mu, \forall \mu \in \Theta.
\]

A natural estimator for estimating \( \mu_S \), the underlying true response of the selected treatment, is the weighted average of sample means at two stages, that is,

\[
\delta_M(X) = \frac{n_1 \overline{X}_S + n_2 \overline{Y}}{n_1 + n_2}.
\]  

(2.3)

Clearly, \( \delta_M(X) \) is the maximum likelihood estimator (MLE) of \( \mu_S \). It is worth mentioning here that the statistic

\[
T = \left(\frac{n_1 \overline{X}_S + n_2 \overline{Y}}{n_1 + n_2}, \overline{X}_{3-S}, S\right) = (T_1, T_2, S) \text{(say)}
\]

is minimal sufficient, but not complete. However, given \( S = s \), the statistic \((T_1, T_2)\) is a complete-sufficient statistic (cf. Bowden and Glimm 2008). Conditioned on \( S = s \), (Bowden and Glimm 2008) derived the uniformly minimum variance conditionally unbiased estimator (UMVCUE) of \( \mu_S \) as

\[
\delta_{BG}(X) = \frac{n_1 \overline{X}_S + n_2 \overline{Y}}{n_1 + n_2} - \sigma \sqrt{\frac{n_1}{n_2(n_1 + n_2)}} \frac{\Phi\left(\frac{\sqrt{n_1}}{\sqrt{n_2(n_1 + n_2)}} \left(n_2D_2-(n_1+n_2)D_1\right)\right)}{\phi\left(\frac{\sqrt{n_1}}{\sqrt{n_2(n_1 + n_2)}} \left(n_2D_2-(n_1+n_2)D_1\right)\right)} \Phi\left(\frac{\sqrt{n_1}}{\sqrt{n_2(n_1 + n_2)}} \left(n_2D_2-(n_1+n_2)D_1\right)\right)^{-1/2}
\]  

(2.4)

where, \( D_1 = \overline{X}_{3-S} - \overline{X}_S \) and \( D_2 = \overline{Y} - \overline{X}_S \).

In this paper, we aim to study properties of the natural estimator \( \delta_M \) and the UMVCUE \( \delta_{BG} \) under the criterion of mean squared error. Specifically we wish to explore if these estimators can be improved. In this direction, in Section 3 of the paper, we prove that the natural estimator \( \delta_M \) is minimax and admissible. In Section 4, we prove that the UMVCUE \( \delta_{BG} \) is inadmissible for estimating \( \mu_S \) and we obtain an estimator improving upon it. In the process we derive a sufficient condition for inadmissibility of an arbitrary location and permutation equivariant estimator and obtain dominating estimators in situations where this sufficient condition is satisfied. We then apply this general result to various single-stage and two-stage estimators and find improved estimators. In Section 5, we make numerical comparisons of various competing estimators. A real-life data analysis is presented in Section 6.

The following two lemmas will be useful in proving main results of the paper.

**Lemma 2.1.** For any \( a \in \mathbb{R} \) and \( b \in \mathbb{R} \),

(i) \( \int_{-\infty}^{\infty} \Phi(ax + b)\phi(x)dx = \Phi\left(\frac{b}{\sqrt{1+a^2}}\right) \);

(ii) \( \int_{-\infty}^{\infty} x\Phi(ax + b)\phi(x)dx = \frac{-ab}{(1+a^2)^{3/2}} \phi\left(\frac{b}{\sqrt{1+a^2}}\right) \);

(iii) \( \int_{-\infty}^{\infty} x^2\Phi(ax + b)\phi(x)dx = \Phi\left(\frac{b}{\sqrt{1+a^2}}\right) - \frac{a^2b}{(1+a^2)^{3/2}} \phi\left(\frac{b}{\sqrt{1+a^2}}\right). \)

**Proof.** For proofs of identities (i) and (ii), refer to Lemma 3.2 of Masihuddin and Misra [15].

(iii) Using the relation, \( x\phi(x) = -\phi'(x) \) and an application of the identities given in (i) and (ii) yields
\[
\int_{-\infty}^{\infty} x^2 \Phi(ax + b) \phi(x) \, dx = \int_{-\infty}^{\infty} \{ \Phi(ax + b) + ax \phi(ax + b) \} \phi(x) \, dx
\]
\[
= \Phi \left( \frac{b}{\sqrt{1 + a^2}} \right) - \frac{a^2 b}{(1 + a^2)^{3/2}} \phi \left( \frac{b}{\sqrt{1 + a^2}} \right).
\]

**Lemma 2.2.** Let \( U = \frac{n_1 \bar{X}_1 + n_2 \bar{Y}}{n_1 + n_2} - \mu_S \), \( \sigma_s^2 = \frac{a^2}{n_1 + n_2} \) and \( \rho = \frac{n_2}{n_1 + n_2} \). Then,

(i) the p.d.f. of \( U \) is given by,
\[
f_U(u) = \frac{1}{\sigma_s} \left[ \Phi \left( \frac{\sqrt{\nu_1}(u + \theta)}{\sigma \sqrt{1 + \rho}} \right) + \Phi \left( \frac{\sqrt{\nu_1}(u - \theta)}{\sigma \sqrt{1 + \rho}} \right) \right] \frac{1}{\sigma_s}, \quad -\infty < u < \infty;
\]

(ii) \( \mathbb{E}_\mu(U^2) = \sigma_s^2 = \frac{n_1}{n_1 + n_2} \).

**Proof.** See the Appendix.

In case of single-stage sampling alone, it is well known that the naive estimator \( \delta_0(X) = \bar{X}_S \) has good risk properties. It is minimax and admissible under the squared error loss function (see (Sackrowitz and Samuel-Cahn 1986) and (Hwang 1993)). Therefore, it is natural to investigate the minimaxity and admissibility of the two-stage natural estimator \( \delta_M(X) \) defined in (2.3).

In the following section, we prove that the naive estimator \( \delta_M(X) = \frac{n_1 \bar{X}_1 + n_2 \bar{Y}}{n_1 + n_2} \) is minimax and admissible for estimating the selected treatment mean \( \mu_S \) under the squared error loss function (2.2).

### 3. Minimaxity and admissibility of the MLE \( \delta_M(X) \)

In the Bayesian set-up, suppose that the unknown parameter \( \mu = (\mu_1, \mu_2) \in \Theta \) is considered to be a realization of a random vector \( H = (H_1, H_2) \), having a specified probability distribution called prior distribution. Consider a sequence of prior densities \( \{\pi_m\}_{m \geq 1} \) for \( H = (H_1, H_2) \) as follows:
\[
\pi_m(h_1, h_2) = \frac{1}{2\pi m^2} e^{-\frac{1}{2m^2} \sum_{i=1}^2 h_i^2}, \quad -\infty < h_i < \infty, \quad i = 1, 2, \quad m = 1, 2, \ldots \tag{3.1}
\]

Recall that, \( \bar{X}_i (i = 1, 2) \) is the sample mean of the first stage sample from the \( i \)th population and \( \bar{Y} \) is the second stage sample mean of the sample drawn from the population selected at the first stage.

It is easy to see that, the posterior distribution of \( (H_1, H_2) \) given \( (\bar{X}_1, \bar{X}_2, \bar{Y}) = (x_1, x_2, y) \) with respect to the prior distribution \( \pi_m(m = 1, 2, \ldots) \) is such that, the random variables \( H_1 \) and \( H_2 \) are independently distributed as \( N(\lambda_{1,m}, \tau_{1,m}^2) \) and \( N(\lambda_{2,m}, \tau_{2,m}^2) \), respectively, where
\[
\left( \lambda_{1,m}, \lambda_{2,m}, \tau_{1,m}^2, \tau_{2,m}^2 \right) = \begin{cases} 
\left( \frac{n_1 x_1 + n_2 x_2}{n_1 + n_2 m}, \frac{n_2 x_1}{n_1 + n_2 m}, \frac{n_1 x_2}{n_1 + n_2 m}, \frac{n_2 x_2}{n_1 + n_2 m} \right), & \text{if } x_1 \geq x_2 \\
\left( \frac{n_1 x_1 + n_2 x_2}{n_1 + n_2 m}, \frac{n_2 x_1}{n_1 + n_2 m}, \frac{n_1 x_2}{n_1 + n_2 m}, \frac{n_2 x_2}{n_1 + n_2 m} \right), & \text{if } x_1 < x_2
\end{cases}
\]

Therefore, under the squared error loss function (2.2), the Bayes estimator of the selected treatment mean \( \mu_S \) is
\[
\delta_{\pi_m}(X) = \begin{cases} 
\mathbb{E}(H_1 | X), & \text{if } \bar{X}_1 \geq \bar{X}_2 \\
\mathbb{E}(H_2 | X), & \text{if } \bar{X}_1 < \bar{X}_2
\end{cases}
\]
\[ \delta \overset{\sim}{m} X(\cdot) = \begin{cases} \left( n_1\overline{X}_1 + s_2\overline{Y} \right), & \text{if } \overline{X}_1 \geq \overline{X}_2 \\ \left( n_1 + n_2 \right) + s_2, & \text{if } \overline{X}_1 < \overline{X}_2 \end{cases} \]

\[ = \left( \frac{n_1\overline{X}_S + n_2\overline{Y}}{n_1 + n_2} + \frac{s_s}{m^2} \right) \]

The posterior risk of the Bayes rule \( \delta_{\Pi_m}(X) \) is obtained as

\[ r_{\delta_{\Pi_m}}(X) = \frac{1}{\frac{n_1 + n_2}{\sigma^2} + \frac{1}{m^2}}, \]

which does not depend on \( X = (\overline{X}_1, \overline{X}_2, \overline{Y}) \). Hence, the Bayes risk of the estimator \( \delta_{\Pi_m} \) is

\[ r_{\delta_{\Pi_m}}^*(\Pi_m) = \frac{1}{n_1 + n_2 + \frac{1}{m^2}}, m = 1, 2, \ldots \] (3.2)

Using Lemma 2.2(ii), we have the risk of the estimator \( \delta_M(X) = \frac{n_1\overline{X}_S + n_2\overline{Y}}{n_1 + n_2} \), as

\[ R(\mu, \delta_M) = \mathbb{E}_\mu \left( \frac{n_1\overline{X}_S + n_2\overline{Y}}{n_1 + n_2} - \mu_S \right)^2 \]

\[ = \sigma^2 \] (3.4)

Therefore, the Bayes risk of the natural estimator \( \delta_M(X) \), under the prior \( \Pi_m \), is

\[ r_{\delta_M}^*(\Pi_m) = \frac{\sigma^2}{n_1 + n_2}, m = 1, 2, \ldots \] (3.5)

Now, we provide the following theorem which proves the minimaxity of the natural estimator \( \delta_M \).

**Theorem 3.1.** The MLE \( \delta_M(X) = \frac{n_1\overline{X}_S + n_2\overline{Y}}{n_1 + n_2} \) is minimax for estimating the selected treatment mean \( \mu_S \) under the squared error loss function (2.2).

**Proof.** Consider \( \delta \) to be any other estimator. Since, \( \delta_{\Pi_m} \), given by (3.2), is the Bayes estimator of \( \mu_S \) with respect to the sequence of priors \( \Pi_m \), we have,

\[ \sup_{\mu \in \Theta} R(\mu, \delta) \geq \int_{\Theta} R(\mu, \delta)\Pi_m(\mu_1, \mu_2) d\mu \]

\[ \geq \int_{\Theta} R(\mu, \delta_{\Pi_m})\Pi_m(\mu_1, \mu_2) d\mu \]

\[ = r_{\delta_{\Pi_m}}^*(\Pi_m) = \frac{1}{\frac{n_1 + n_2}{\sigma^2} + \frac{1}{m^2}}, m = 1, 2, \ldots \] (using (3.3))

\[ \Rightarrow \sup_{\mu \in \Theta} R(\mu, \delta) \geq \lim_{m \to \infty} r_{\delta_{\Pi_m}}^*(\Pi_m) = \frac{\sigma^2}{n_1 + n_2} = \sup_{\mu \in \Theta} R(\mu, \delta_M), \text{ (using (3.4))} \]
which implies that $\delta_M(X) = \frac{n_1\overline{X_1} + n_2\overline{X_2}}{n_1 + n_2}$ is minimax for estimating $\mu_S$.

Now we apply the principle of invariance. The estimation problem at hand is invariant under the location group of transformations $G = \{g_b : b \in \mathbb{R}\}$, where $g_b(x_1, x_2, y) = (x_1 + b, x_2 + b, y + b), (x_1, x_2, y) \in \mathbb{R}^3, b \in \mathbb{R}$, and also under the group of permutations $G_p = \{g_1, g_2\}$, where $g_1(x_1, x_2, y) = (x_1, x_2, y), g_2(x_1, x_2, y) = (x_2, x_1, y), (x_1, x_2, y) \in \mathbb{R}^3$. For the goal of estimating $\mu_S$, it is easy to verify that any location and permutation equivariant estimator will be of the form

$$\delta_p(X) = \frac{n_1\overline{X_1} + n_2\overline{Y}}{n_1 + n_2} + \psi(D_1, D_2),$$

for some function $\psi : (-\infty, 0] \times \mathbb{R} \to \mathbb{R}$, where $D_1 = \overline{X}_{-S} - \overline{X}_S$ and $D_2 = \overline{Y} - \overline{X}_S$. Let $D_1$ denote the class of all location and permutation equivariant estimators of the type (3.6). Note that the MLE $\delta_M(X)$ and UMVCUE $\delta_{BG}$, defined by (2.3) and (2.4) respectively, belong to the class $D_1$.

Next, we will show that the MLE $\delta_M$ is admissible in the class $D_1$ of location and permutation equivariant estimators. Notice that, the risk function of any estimator $\delta \in D_1$ depends on $\mu \in \Theta$ through $\theta = \theta_2 - \theta_1$. Therefore, we denote $R(\mu, \delta)$ by $R(\theta, \delta), \theta \geq 0$. Consequently, the Bayes risk of any estimator $\delta \in D_1$ depends on the prior distribution of $H = (H_1, H_2)$ through the distribution of $(H_2 - H_1)$, where $H_1 = \min\{H_1, H_2\}, H_2 = \max\{H_1, H_2\}$.

The following theorem establishes the admissibility of the estimator $\delta_M$, within the class $D_1$ of location and permutation equivariant estimators.

**Theorem 3.2.** For estimating $\mu_S$, under the squared error loss function (2.2), the MLE $\delta_M(X) = \frac{n_1\overline{X_1} + n_2\overline{X_2}}{n_1 + n_2}$ is admissible within the class $D_1$.

**Proof.** On contrary, let us assume that $\delta_M$ is inadmissible in $D_1$, i.e., there exists an estimator $\delta'$ such that

$$R(\theta) \leq R(\theta), \forall \theta \geq 0$$

and

$$R(\theta) < R(\theta), \text{ for some } \theta_0 \geq 0.$$  

Let $R(\theta_0) - R(\theta_0) = \epsilon$, so that $\epsilon > 0$. Define

$$B = \{\theta \geq 0 : R_\phi(\theta_0) - R_\phi(\theta_0) > \frac{\epsilon}{2}\}.$$  

Then, $B \cap \phi(\theta_0 \in B) \cap B$ is open (since $R(\theta_0) - R(\theta_0)$ is a continuous function of $\theta \in [0, \infty)$).

Let $r_\phi(\Pi_m)$ denote the Bayes risk of $\theta'$ under the prior $\Pi_m, m = 1, 2, \ldots$. Since $\delta_{\Pi_m}$, defined by (3.2), is the Bayes estimator with respect to prior $\Pi_m$, we have

$$r_\delta\Pi_m - r_{\delta_{\Pi_m}}(\Pi_m) \geq r_\delta\Pi_m - r_{\delta'}(\Pi_m)$$

$$= \int_\Theta [R_\delta(\theta) - R_\delta'(\theta)] \Pi_m(\mu_1, \mu_2) d\mu$$

$$\geq \int_B [R_\delta(\theta) - R_\delta'(\theta)] \Pi_m(\mu_1, \mu_2) d\mu$$
\[
\geq \frac{\varepsilon}{2} \mathbb{P}_{\Pi_m}( (H_{[2]} - H_{[1]}) \in B), \quad m = 1, 2, \ldots.
\]  
(3.7)

Now using (3.3), (3.5) and (3.7), we get

\[
\lim \inf_{m \to \infty} \frac{1}{m^2 \left( \frac{n_1 + n_2}{\sigma^2} \right) \left( \frac{n_1 + n_2}{\sigma^2} + \frac{1}{m^2} \right) \mathbb{P}_{\Pi_m}( (H_{[2]} - H_{[1]}) \in B)} \geq \frac{\varepsilon}{2}.
\]  
(3.8)

However, for any \(0 < a < b < \infty\), we have

\[
\mathbb{P}_{\Pi_m}(a < H_{[2]} - H_{[1]} \leq b) = 2 \left[ \Phi \left( \frac{b}{\sqrt{2m}} \right) - \Phi \left( \frac{a}{\sqrt{2m}} \right) \right]
\]

\[
= \frac{\sqrt{2}(b - a)}{m} \Phi(d^*), \quad a < d^* < b \sqrt{2m}. \]

Therefore,

\[
\lim \inf_{m \to \infty} \frac{1}{m^2 \left( \frac{n_1 + n_2}{\sigma^2} \right) \left( \frac{n_1 + n_2}{\sigma^2} + \frac{1}{m^2} \right) \mathbb{P}_{\Pi_m}( (H_{[2]} - H_{[1]}) \in B)} = 0,
\]  
(3.9)

which contradicts (3.8). Hence, the theorem follows.

**Remark 3.2.1.** Note that \(\delta_0(X) = \overline{X}_S\) is an admissible and minimax estimator based on the single-stage data (see Sackrowitz and Samuel-Cahn (1986). Using Lemma 3.1(ii) of Misra and Misra (2021) and Lemma 2.2(ii) we have

\[ R\left( \mu, \delta_M \right) = \frac{\sigma^2}{n_1 + n_2} < \frac{\sigma^2}{n_1} = R\left( \mu, \delta_0 \right), \forall \mu \in \Theta, \]

that is, under the two-stage drop-the-losers design, the estimator \(\delta_0(X) = \overline{X}_S\) is dominated by the two-stage admissible and minimax estimator \(\delta_M\).

4. **Inadmissibility of the UMVCUE**

It is a well-known result that for Gaussian populations there does not exists an unbiased estimator of the selected treatment mean \(\mu_S\) using single-stage data only (see Putter and Rubinstein 1968), (Vellaisamy 2009) and (Misra and Misra 2021)). A similar result was provided by (Tappin 1992) for binomial populations. For the goal of estimating \(\mu_S\), conditioned on \(S\), (Bowden and Glimm 2008) proposed the following two-stage UMVCUE

\[
\delta_{BG}(X) = \frac{n_1 \overline{X}_S + n_2 \overline{Y}}{n_1 + n_2} + \psi_{BG}(D_1, D_2)
\]  
(4.1)

with, \(\psi_{BG}(D_1, D_2) = -\sigma \sqrt{\frac{n_1}{n_1(n_1+n_2)}} \phi \left( \frac{\overline{X}_S - \overline{X}_{3-S} - \overline{X}_S}{\sqrt{\frac{n_1}{n_1(n_1+n_2)}}} \right) \), \(D_1 = \overline{X}_{3-S} - \overline{X}_S\) and \(D_2 = \overline{Y} - \overline{X}_S\). It is worth mentioning here that, conditioned on \(S\), \(\overline{Y}\) is an unbiased estimator of \(\mu_S\), \(E_{\mu}(\overline{Y}(T_1, T_2)) = \delta_{BG}(X)\). Clearly \(\delta_{BG}(X) \in D_1\), where \(D_1\) is the class of location and permutation equivariant estimators defined in the last section (see (3.6)). In order to find an improvement over \(\delta_{BG}\), we now use the idea of (Brewster and Zidek 1974) to obtain a sufficient condition for inadmissibility of an arbitrary location and permutation equivariant estimators \(\delta_\psi(X) = \frac{n_1 \overline{X}_S + n_2 \overline{Y}}{n_1 + n_2} + \psi(D_1, D_2)\), for some function function \(\psi(\cdot) : (-\infty, 0] \times \mathbb{R} \to \mathbb{R}\), of \(\mu_S\).
As a direct consequence of this result, we will see later in the section that the estimator $\delta_{BG}(\cdot)$, given in (4.1), is inadmissible and a dominating estimator is provided.

Recall that $X_{3-S} = \min(X_1, X_2)$, $X_S = \max(X_1, X_2)$; $D_1 = X_{3-S} - X_S$, $D_2 = Y - X_S$. Let $Z = X_S - \mu_S$. The following lemma will be useful in obtaining the main result of this section.

**Lemma 4.1.** Let $d_1 \in (-\infty, 0]$ and $d_2 \in \mathbb{R}$ be fixed. Then,

(a) the conditional p.d.f. of $Z$, given $(D_1, D_2) = (d_1, d_2)$, is given by

$$f_{1,2}(z|d_1, d_2) = \frac{\phi\left(\frac{\sqrt{n_1}}{\sigma}(z + d_1 - \theta)\right) + \phi\left(\frac{\sqrt{n_1}}{\sigma}(z + d_1 + \theta)\right) \phi\left(\frac{\sqrt{n_1}}{\sigma}(z + d_2)\right) \phi\left(\frac{\sqrt{n_1}}{\sigma}t\right)}{\int_{-\infty}^{\infty} \phi\left(\frac{\sqrt{n_1}}{\sigma}(t + d_1 - \theta)\right) + \phi\left(\frac{\sqrt{n_1}}{\sigma}(t + d_1 + \theta)\right) \phi\left(\frac{\sqrt{n_1}}{\sigma}(t + d_2)\right) \phi\left(\frac{\sqrt{n_1}}{\sigma}t\right) dt},$$

$z \in \mathbb{R}; \mu \in \Theta, \theta \geq 0$.

(b) the conditional expectation of $Z$, given $(D_1, D_2) = (d_1, d_2)$, is given by

$$\mathbb{E}_\mu[Z|(D_1, D_2) = (d_1, d_2)] = \frac{n_1 \theta}{n_2 + 2n_1} \left\{ \frac{1 - e^{-\frac{n_1}{n_2}}} {1 + e^{-\frac{n_1}{n_2}}} + \frac{n_1 \mu_S + n_2 \mu_S}{n_1 + n_2} \right\} - \frac{n_1 d_1 + n_2 d_2}{2n_1 + n_2},$$

$\mu \in \Theta, \theta \geq 0$.

Proof. See the Appendix.

It is worth noting that the risk function of the estimator $\delta_\psi(X)$, given by (3.6), depends on $\mu = (\mu_1, \mu_2)$ only through $\theta$. Therefore, for notational convenience, we denote $R(\mu, \delta_\psi)$ by $R_\theta(\delta_\psi)$. The risk function of an estimator of the form (3.6) is

$$R_\theta(\delta_\psi) = \mathbb{E}_\mu[R_1(\theta, \psi(D_1, D_2))], \theta \geq 0,$$

where, for any fixed $d_1 \in (-\infty, 0]$ and $d_2 \in \mathbb{R},$

$$R_1(\theta, \psi(d_1, d_2)) = \mathbb{E}_\mu\left[ \left( \frac{n_1 X_S + n_2 Y}{n_1 + n_2} - \psi(D_1, D_2) - \mu_S \right)^2 \right], (D_1, D_2) = (d_1, d_2), \theta \geq 0. \quad (4.3)$$

Our objective is to minimize the risk function $R_\theta(\delta_\psi)$ given by (4.2) with respect to $\psi$. This objective may be achieved by minimizing the conditional risk $R_1(\theta, \psi(d_1, d_2))$, given by (4.3). For any fixed $\theta \in [0, \infty)$, the choice of $\psi$ that minimizes (4.3) is obtained as

$$\psi_\theta(d_1, d_2) = -\mathbb{E}_\mu\left[ \left( \frac{n_1 X_S + n_2 Y}{n_1 + n_2} - \mu_S \right)^2 \right], (D_1, D_2) = (d_1, d_2)$$

$$= -\mathbb{E}_\mu[Z|(D_1, D_2) = (d_1, d_2)] - \frac{n_2 d_2}{n_1 + n_2}. \quad (4.4)$$

Further, an application of Lemma 4.1(b) implies,
\[
\psi_\theta(d_1, d_2) = \frac{n_1 \theta}{n_2 + 2n_1} \left( 1 - e^{-\frac{2n_1((n_1 + n_2)d_1 - n_2d_2)}{2n_1(n_1 + n_2)}} \right) + \frac{n_1((n_1 + n_2)d_1 - n_2d_2)}{(n_1 + n_2)(2n_1 + n_2)}, \theta \geq 0.
\]

It can be seen that \(\psi_\theta(d_1, d_2)\) is increasing in \(\theta \in [0, \infty)\), whenever \(d_1 \leq \frac{n_2d_2}{n_1 + n_2}\), and decreasing in \(\theta \in [0, \infty)\), whenever \(d_1 > \frac{n_2d_2}{n_1 + n_2}\). Thus, we have

\[
\inf_{\theta \geq 0} \psi_\theta(d_1, d_2) = \left( \frac{n_1((n_1 + n_2)d_1 - n_2d_2)}{(n_1 + n_2)(2n_1 + n_2)}, \text{ if } d_1 \leq \frac{n_2d_2}{n_1 + n_2} = \psi_* (d_1, d_2), \text{ (say)} \right) \tag{4.5}
\]

and

\[
\sup_{\theta \geq 0} \psi_\theta(d_1, d_2) = \left( \frac{n_1((n_1 + n_2)d_1 - n_2d_2)}{(n_1 + n_2)(2n_1 + n_2)}, \text{ if } d_1 \leq \frac{n_2d_2}{n_1 + n_2} = \psi^* (d_1, d_2), \text{ (say)} \right) \tag{4.6}
\]

Note that, for any fixed \(d_1 \in (-\infty, 0], d_2 \in \mathbb{R}\) and \(\mu \in \Theta\), the conditional risk, given by (4.3), is strictly decreasing on \((-\infty, \psi_\theta(d_1, d_2))\) and strictly increasing on \((\psi_\theta(d_1, d_2), \infty)\), where \(\psi_\theta(d_1, d_2), \theta \geq 0\), is defined by (4.4). Using this fact along with (4.5) and (4.6), we also note that:

(i) for any fixed \(d_1 \in (-\infty, 0], d_2 \in \mathbb{R}\) and \(\mu \in \Theta\), \(R_1(\theta, \psi_\theta(d_1, d_2))\) is strictly decreasing in \(\psi(d_1, d_2) \in (-\infty, \psi_* (d_1, d_2))\);

(ii) for any fixed \(d_1 \in (-\infty, 0], d_2 \in \mathbb{R}\) and \(\mu \in \Theta\), \(R_1(\theta, \psi_\theta(d_1, d_2))\) is strictly increasing in \(\psi(d_1, d_2) \in (\psi^* (d_1, d_2), \infty)\).

Thus, we have the following theorem.

**Theorem 4.1.** Let \(\delta_\psi(X) = \frac{n_1X + n_2\bar{Y}}{n_1 + n_2} + \psi(D_1, D_2)\), for a real valued function \(\psi\) defined on \((-\infty, 0] \times \mathbb{R}\), be a location and permutation equivariant estimator of \(\mu_S\). Suppose that

\[
\mathbb{P}_{\mu_0} \left[ \psi(D_1, D_2) < \frac{n_1((n_1 + n_2)D_1 - n_2D_2)}{(n_1 + n_2)(2n_1 + n_2)} \leq 0 \right]
\]

\[
+ \mathbb{P}_{\mu_0} \left[ 0 \leq \frac{n_1((n_1 + n_2)D_1 - n_2D_2)}{(n_1 + n_2)(2n_1 + n_2)} \leq \psi(D_1, D_2) \right] > 0, \text{ for some } \mu \in \Theta,
\]

Then, the estimator \(\delta_\psi(\cdot)\) is inadmissible for estimating \(\mu_S\) under the squared error loss function (2.2) and is dominated by

\[
\delta_\psi^I(X) = \left\{ \begin{array}{ll}
\frac{n_1(X + \bar{X}) + n_2\bar{Y}}{2n_1 + n_2}, & \text{if } \psi(D_1, D_2) < \frac{n_1}{2n_1 + n_2} \left( \bar{X}_3 - \frac{n_2X + n_2\bar{Y}}{n_1 + n_2} \right) \leq 0,
\end{array} \right.
\]

\[
\frac{n_1(X + \bar{X}) + n_2\bar{Y}}{2n_1 + n_2}, & \text{for } 0 \leq \frac{n_1}{2n_1 + n_2} \left( \bar{X}_3 - \frac{n_2X + n_2\bar{Y}}{n_1 + n_2} \right) < \psi(D_1, D_2),
\right.
\]

\[
\delta_\psi(X), & \text{otherwise.}
\right. \tag{4.7}
\]

As an immediate consequence of the above theorem we have the following corollary dealing with the admissibility of the UMVCUE \(\delta_{BG}(X)\) and \(\delta_0(X) = \bar{X}_S\).

**Corollary 4.1.1.** (a) The UMVCUE \(\delta_{BG}(X)\), given in (4.1), is inadmissible for estimation \(\mu_S\) under the squared error loss function (2.2) and is dominated by

\[
\delta_{BG}^I(X) = \left\{ \begin{array}{ll}
\frac{n_1(X + \bar{X}) + n_2\bar{Y}}{2n_1 + n_2}, & \text{if } \bar{X}_3 \cdot s \leq \frac{n_2X + n_2\bar{Y}}{n_1 + n_2} \leq \bar{X}_3 \cdot s + \frac{(2n_1 + n_2)(\bar{X}_Q - \phi_Q)}{(n_1 + n_2)(2n_1 + n_2)},
\end{array} \right.
\]

\[
\delta_{BG}(X), & \text{otherwise.}
\right. \tag{4.8}
\]
4.1. The obtained statistic \( T = \frac{n_1 \bar{X}_1 + n_2 \bar{Y} - (n_1 + n_2) \bar{X}_S}{\sigma} \).

(b) The single-stage admissible and minimax estimator \( \delta_0(X) = \bar{X}_S = \frac{n_1 \bar{X}_1 + n_2 \bar{Y}}{n_1 + n_2} - \frac{n_0 D_3}{n_1 + n_2} \) is also inadmissible for estimating \( \mu_S \) and is dominated by

\[
\delta_0^I(X) = \begin{cases} 
\frac{n_1 (\bar{X}_1 + \bar{X}_S) + n_2 \bar{Y}}{2n_1 + n_2}, & \text{if } -\frac{n_1}{2n_1 + n_2} \left( \bar{X}_3 - \bar{S} - \frac{n_1 \bar{X}_1 + n_2 \bar{Y}}{n_1 + n_2} \right) \leq 0 \\
\delta_0(X), & \text{otherwise.}
\end{cases}
\]

Remark 4.1.1. Under the single-stage sample set-up, Dahiya (1974) considered various estimators for estimating the selected treatment mean \( \mu_S \). It follows from Theorem 4.1 that, under the two-stage DLD, all the estimators proposed by Dahiya (1974) are inadmissible for estimating \( \mu_S \) w.r.t. the criterion of mean squared error.

4.1. Some additional admissibility results

Note that the random estimand \( \mu_S \) depends on observations only through the sufficient statistic \( T = \left( \frac{n_1 \bar{X}_1 + n_2 \bar{Y}}{n_1 + n_2}, \bar{X}_3, S \right) = (T_1, T_2, S) \). Then, for the goal of estimating \( \mu_S \), an improved estimator \( \delta_\psi^I \) obtained using Theorem 4.1, can be further improved in terms of MSE if it is not purely a function of the sufficient statistic \( T = \left( \frac{n_1 \bar{X}_1 + n_2 \bar{Y}}{n_1 + n_2}, \bar{X}_3, S \right) = (T_1, T_2, S) \). In fact any location and permutation equivariant estimator \( \delta_\psi \) of \( \mu_S \) is not a function of the sufficient statistic can be improved using the Rao-Blackwell theorem on \( \delta_\psi \) to obtain a dominating estimator as \( \delta_\psi^*(T) = \mathbb{E}_\mu(\delta_\psi(X) | T) \). For estimating \( \mu_S \) under the squared error loss (2.2), suppose we have an estimator of the type

\[
\delta_\psi(X) = \frac{n_1 \bar{X}_S + n_2 \bar{Y}}{n_1 + n_2} + \psi(D_1, D_2)
\]

\[
= \frac{n_1 \bar{X}_S + n_2 \bar{Y}}{n_1 + n_2} + \psi(\bar{X}_3 - \bar{S}, \bar{Y} - \bar{X}_S).
\]

Then, the estimator

\[
\delta_\psi^*(X) = \mathbb{E}_\mu\left[ \frac{n_1 \bar{X}_S + n_2 \bar{Y}}{n_1 + n_2} + \psi(\bar{X}_3 - \bar{S}, \bar{Y} - \bar{X}_S) | T \right]
\]

\[
= T_1 + \mathbb{E}_\mu\left[ \psi(\bar{X}_3 - \bar{S}, \bar{Y} - \bar{X}_S) | T \right],
\]

dominates \( \delta_\psi \), i.e. \( R(\mu_S, \delta_\psi^*) < R(\mu_S, \delta_\psi) \) \( \forall \mu \in \Theta \) (unless \( \delta_\psi = \delta_\psi^* \) a.e.)

**Example:** For estimating the selected treatment mean \( \mu_S \) under the two-stage drop-the-losers design, the single-stage admissible and minimax estimator \( \delta_0(X) = \bar{X}_S \) can also be improved using the Rao-Blackwell theorem. In terms of MSE, \( \delta_0 \) is dominated by an estimator \( \delta_0^{RB}(T_1, T_2, S) = \mathbb{E}(\delta_0(X) | T) \) given as

\[
\delta_0^{RB}(T) = T_1 + \frac{\sqrt{n_2} \sigma}{\sqrt{n_1(n_1 + n_2)}} \phi \left( \frac{\sqrt{n_1(n_1 + n_2)}(T_1 - T_2)}{\sqrt{n_2} \sigma} \right),
\]

\[
\psi \left( \frac{\sqrt{n_1(n_1 + n_2)}(T_1 - T_2)}{\sqrt{n_2} \sigma} \right).
\]
Derivation of $\delta_0^{RB}$: Let $Y_1, Y_2$ be as defined in the proof of Lemma 2.2, given in the Appendix. Note that, for $(t_1, t_2) \in \mathbb{R}^2$, $\delta_0^{RB}(t_1, t_2, 1) = \mathbb{E}_\mu \left( X_1 \middle| \frac{n_1 X_1 + n_2 Y_1}{n_1 + n_2} = t_1, X_2 = t_2, X_1 > t_2 \right)$ and $\delta_0^{RB}(t_1, t_2, 2) = \mathbb{E}_\mu \left( X_2 \middle| \frac{n_1 X_1 + n_2 Y_2}{n_1 + n_2} = t_1, X_1 = t_2, X_2 > t_2 \right)$. We have,

$$
\begin{bmatrix}
\frac{X_1}{n_1 X_1 + n_2 Y_1} \\
\frac{X_2}{n_1 X_1 + n_2 Y_2}
\end{bmatrix}
\sim N_2 \left( \begin{bmatrix}
\mu_1 \\
\mu_2
\end{bmatrix}, \begin{bmatrix}
\frac{\sigma_1^2}{n_1 + n_2} & \frac{\sigma_1^2}{n_1 + n_2} \\
\frac{\sigma_2^2}{n_1 + n_2} & \frac{\sigma_2^2}{n_1 + n_2}
\end{bmatrix} \right).
$$

Given that $\left( \frac{n_1 X_1 + n_2 Y_1}{n_1 + n_2}, \frac{n_1 X_1 + n_2 Y_2}{n_1 + n_2} \right) = (t_1, t_2)$, $X_1$ follows univariate Gaussian distribution with mean $\mu^* = t_1$ and variance $\sigma_1^2 = \frac{n_2 \sigma^2}{n_1 (n_1 + n_2)}$. Therefore,

$$
\delta_0^{RB}(t_1, t_2, 1) = \mathbb{E}_\mu \left( X_1 \middle| \frac{n_1 X_1 + n_2 Y_1}{n_1 + n_2} = t_1, X_2 = t_2, X_1 > t_2 \right) = t_1 + \sigma_1 \phi \left( \frac{t_1 - \mu}{\sigma_1} \right).
$$

By symmetry,

$$
\delta_0^{RB}(t_1, t_2, 2) = \mathbb{E}_\mu \left( X_2 \middle| \frac{n_1 X_1 + n_2 Y_2}{n_1 + n_2} = t_1, X_1 = t_2, X_2 > t_2 \right) = t_1 + \sigma_1 \phi \left( \frac{t_1 - \mu}{\sigma_1} \right).
$$

Therefore,

$$
\delta_0^{RB}(T) = T_1 + \frac{\sqrt{n_2} \sigma}{\sqrt{n_1 (n_1 + n_2)}} \phi \left( \frac{\sqrt{n_1 (n_1 + n_2)} (T_1 - T_2)}{\sqrt{n_2} \sigma} \right).
$$

Note that the estimator $\delta_0$, defined by (4.9), also dominates $\delta_0(X) = \bar{X}_S$. Since the estimator $\delta_0(X)$ is not a function of minimal sufficient statistic $T$, it can be further improved by an estimator $\delta_1$, defined as

$$
\delta_1(T) = \mathbb{E} \left( \delta_0(X) \middle| T \right)
$$

$$
= \begin{cases}
\frac{(n_1 + n_2) T_1 + n_2 T_2}{2n_1 + n_2} \mathbb{P} \left( -n_2 (\bar{Y} - \bar{X}_S) < n_1 (T_1 - T_2) \middle| T \right), & \text{if } T_1 > T_2, \\
+ \mathbb{E}_\mu \left[ \bar{X}_S \right] \mathbb{P} \left( -n_2 (\bar{Y} - \bar{X}_S) \geq n_1 (T_1 - T_2) \middle| T \right), & \text{if } T_1 > T_2, \\
\frac{(n_1 + n_2) T_1 + n_2 T_2}{2n_1 + n_2} \mathbb{P} \left( -n_2 (\bar{Y} - \bar{X}_S) \geq n_1 (T_1 - T_2) \middle| T \right), & \text{if } T_1 \leq T_2.
\end{cases}
$$

Using the conditional distribution of $\bar{X}_S$ given $(T_1, T_2) = (t_1, t_2)$ obtained in the derivation of the estimator $\delta_0^{RB}$ (4.10) we obtain,
\[ \delta_1(T) = \begin{cases} 
\frac{(n_1+n_2)T_1+n_1T_2}{2n_1+n_2} & \text{if } T_1 > T_2, \\
\frac{\phi \left( \frac{T_1-T_2}{\tau_1} \right)}{\phi \left( \frac{T_1}{\tau_1} \right)} - \phi \left( \frac{n_1(T_1-T_2)}{(n_1+n_2)\tau_1} \right) + \phi \left( \frac{n_1T_2}{(n_1+n_2)\tau_1} \right) \frac{\sigma_1\phi \left( \frac{n_1(T_1-T_2)}{(n_1+n_2)\tau_1} \right)}{\phi \left( \frac{T_1}{\tau_1} \right)} + T_1 \phi \left( \frac{n_1(T_1-T_2)}{(n_1+n_2)\tau_1} \right) \frac{\sigma_1\phi \left( \frac{n_1(T_1-T_2)}{(n_1+n_2)\tau_1} \right)}{\phi \left( \frac{T_1}{\tau_1} \right)} 
\end{cases} \]  

(4.11)

where \( \sigma_1 = \sigma \sqrt{\frac{n_1}{n_1(n_1+n_2)}} \). In addition to numerical comparison of performances of various estimators, in the following section, we compare the performance of the estimator \( \delta_1(T) \) with \( \delta_0 \) and \( \delta_0^{RB} \) numerically, through simulation, under the criterion of MSE and bias. In conformity with our theoretical findings, numerical study suggests that the estimator \( \delta_1 \), defined by (4.11), has better MSE performance than \( \delta_0^{\ell} \), given by (4.9). In terms of MSE, we will also see in the next Section that \( \delta_1 \) performs better than \( \delta_0^{RB} \), given by (4.10).

**4.2. Estimation of the mean of worse selected treatment**

In practice, it might be of interest to estimate the worse selected treatment among the two available treatments (say \( \tau_1 \) and \( \tau_2 \)) under consideration. We define the treatment corresponding to \( \min \{ \mu_1, \mu_2 \} \) as the worse treatment. For selecting the worse treatment, we consider the natural selection rule that selects the treatment with the smaller sample mean. Let \( J \in \{ 1, 2 \} \) denote the index of the selected treatment \( \tau_J \) in Stage I (i.e. \( J = 1 \), if \( \bar{X}_1 < \bar{X}_2 \), \( J = 2 \), if \( \bar{X}_2 \leq \bar{X}_1 \)). The selected treatment \( \tau_J \) is then carried forward in Stage II of the DLD, in which \( n_2 \) patients are further randomized to receive the treatment \( \tau_J \). Let \( \bar{Y} \) be the sample mean of the sample in Stage II. Then, \( \bar{X}_i \sim N \left( \mu_i, \sigma_i^2/n_i \right), i = 1, 2 \), are independently distributed and, conditioned on \( J \), \( \bar{Y} \sim N \left( \mu_J, \sigma_J^2/n_J \right) \). Here, our goal is to estimate the selected treatment mean

\[ \mu_J = \begin{cases} 
\mu_1, & \text{if } \bar{X}_1 < \bar{X}_2 \\
\mu_2, & \text{if } \bar{X}_2 \leq \bar{X}_1 
\end{cases} \]  

(4.12)

under the squared error loss function.

Applying the transformation \( \bar{X}_1 \rightarrow -\bar{X}_1, \bar{X}_2 \rightarrow -\bar{X}_2 \) and \( \bar{Y} \rightarrow -\bar{Y} \), we get \( \mu_1 \rightarrow -\mu_1 \). Thus, the results obtained in Sections 3 and 4 can be directly translated to the problem of estimating the selected worse treatment mean \( \mu_J \). A naive estimator of \( \mu_J \) is \( \delta_M(X) = \frac{n_1\bar{X}_1 + n_2\bar{Y}}{n_1 + n_2} \), which is the MLE of \( \mu_J \). The UMVCUE of \( \mu_J \) is obtained as

\[ \delta_U(X) = \frac{n_1\bar{X}_1 + n_2\bar{Y}}{n_1 + n_2} + \sigma \sqrt{\frac{n_1}{n_2(n_1+n_2)}} \Phi \left( \frac{n_1}{\sqrt{n_2(n_1+n_2)}} \left( \frac{E_1 - n_2E_2}{\sigma} \right) \right) \]  

(4.13)

with, \( E_1 = \bar{X}_J - \bar{X}_3 - J \) and \( E_2 = \bar{X}_J - \bar{Y} \). From the results obtained in Sections 3 and 4, the following findings are easily obtained: (i) The estimator \( \delta_M(X) = \frac{n_1\bar{X}_1 + n_2\bar{Y}}{n_1 + n_2} \) is minimax for estimating mean \( \mu_J \), under the squared error loss function.

(ii) Consider the following class of equivariant estimators of \( \mu_J \):

\[ C_L = \left\{ \delta_f(X) : \frac{n_1\bar{X}_J + n_2\bar{Y}}{n_1 + n_2} - \phi(E_1, E_2) \right\} \]
where, \( E_1 = \bar{X}_j - \bar{X}_{3-j} \) and \( E_2 = \bar{X}_j - \bar{Y} \) and \( \varphi \) is a real valued function defined on \((-\infty, 0] \times \mathbb{R}\).

Then, one can easily obtain a result analogous to Theorem 4.1., providing a sufficient condition for inadmissibility of an arbitrary estimator in class \( C_L \).

(iii) Analogous to Theorem 3.2, one can also prove that \( \delta_{M}^{*}(X) = \frac{n_1\bar{X}_j + n_2\bar{Y}}{n_1 + n_2} \) is admissible within the the class \( C_L \) for estimating \( \mu \) under the squared error loss function.

(iv) For estimating the selected worse treatment mean \( \mu_j \), one can easily obtain estimators analogous to \( \delta_{0}^{RB} \) and \( \delta_{1} \) using the Rao-Blackwellization, as

\[
\delta_{0}^{RB}(W) = W - \frac{\sqrt{n_2} \sigma}{\sqrt{n_1(n_1 + n_2)}} \phi \left( \frac{\sqrt{n_1(n_1 + n_2)}}{\sqrt{n_2} \sigma} (W_2 - W_1) \right)
\]

and

\[
\delta_{1}^*(W) = \left\{ \begin{array}{ll}
\frac{(n_1 + n_2)W_1 + n_1W_2}{2n_1 + n_2} & \text{if } W_1 < W_2 \\
\sigma \sqrt{\frac{n_2}{n_1(n_1 + n_2)}} - W_1 & \text{if } W_1 = W_2 \\
\frac{(n_1 + n_2)W_1 + n_1W_2}{2n_1 + n_2} & \text{if } W_1 \geq W_2
\end{array} \right.
\]

where \( \sigma = \sigma \sqrt{\frac{n_2}{n_1(n_1 + n_2)}} \), \( W_1 = \frac{n_1\bar{X}_j + n_2\bar{Y}}{n_1 + n_2} \) and \( W_2 = \bar{X}_{3-j} \).

### 5. Simulation study

In this section, we consider a simulation study to compare the risk (MSE) and bias performances of various estimators of the selected treatment mean \( (\mu_j) \) under the two-stage drop-the-losers design. We consider the following estimators for our numerical study: \( \delta_M, \delta_{BG}, \delta_{BG}^{l}, \delta_{0}, \delta_{0}^{RB} \) and \( \delta_{1} \) (see (2.3), (2.4), (4.8), (4.9), (4.10) and (4.11)). For our simulation study, we take \( \sigma = 1 \). Since the MSE and bias of these estimators of \( \mu_j \), depend on \( \mu = (\mu_1, \mu_2) \in \Theta \) only through \( \theta = \theta_2 - \theta_1 \), we simulate the MSE and the bias of various estimators against different configurations of \( \theta \geq 0 \) and sample sizes \( n_1 \) and \( n_2 \). The simulated values of the MSE and the bias based on 100,000 simulations are plotted in Figures 1–12 and Figures S1–8 (please refer the supplementary material for Figures S1–8). The percentage risk improvement of the estimator \( \delta_{0}^{l} \) over \( \delta_{0} \) have been tabulated in Table 1. It is observed that \( \delta_{0}^{l} \) provides considerable risk improvements over the single-stage admissible and minimax estimator \( \delta_{0} \).

In Figures 1–4 and Figures S1–2, we have plotted the simulated MSE values of the estimators \( \delta_{M}, \delta_{BG}, \delta_{BG}^{l}, \delta_{0}^{RB}, \delta_{1} \) and \( \bar{Y} \) for different configurations of \( n_1, n_2 \) against \( \theta \). In Figures 5–6 and Figures S3–4 we have plotted the simulated MSE values of the estimators \( \delta_{BG} \) and its improved version \( \delta_{BG}^{l} \). In Figures 7–8 and Figures S5–6, we have plotted the simulated MSE of these estimators for fixed \( \theta \) and the total sample size \( n = n_1 + n_2 \). In Figures 9–12 and Figures S7–8, we have plotted the simulated bias values of the estimators \( \delta_{M}, \delta_{BG}, \delta_{BG}^{l}, \delta_{0}^{RB}, \delta_{1} \) and \( \bar{Y} \) for different configurations of \( n_1, n_2 \) against \( \theta \).

The following concluding remarks are evident from Figures 1–12 and Figures S1–8:

**Remark 5.0.1.**

(i) The MSE of the estimator \( \delta_{BG}^{l} \) is nowhere larger than that of the UMVCUE \( \delta_{BG} \) (see Figures 5–6, Figures S3–4), which confirms our theoretical finding.

(ii) In terms of MSE, the maximum likelihood estimator \( \delta_{M} \), which is minimax and admissible, performs reasonably well in comparison to other estimators of \( \mu_j \). For \( n_1=n_2, \delta_{BG} \) and \( \delta_{0}^{RB} \) have
very similar MSE performance (see Figure 2, Figure S2).
For \( n_1 \geq n_2 \), the estimator \( \delta_1 \) beats all other estimators, except \( \delta_M \), in terms of the MSE. It is also observed that, \( \delta_1 \) uniformly dominates \( \delta_0^{RB} \) in terms of the MSE. Also, when \( n_1 > n_2 \), \( \overline{Y} \) has very poor MSE performance in comparison to other estimators (see Figure 1-2, Figure S1). For relatively small \( n_1 \) and relatively large \( n_2 \), \( \overline{Y} \) performs better as compared to the estimators \( \delta_0^{RB} \) and \( \delta_1 \). Also, in this case, the MSEs of \( \delta_{BG} \) and \( \delta_{BG}^I \) are comparable to that of \( \delta_M \) (see Figures 3-4).

(iii) For fixed \( \theta \) and the total size \( n = n_1 + n_2 \), \( \delta_M \) has constant MSE \( \left( = \frac{\sigma^2}{n_1 + n_2} \right) \) under different allocations of \( n_1 \) and \( n_2 \). For a fixed smaller value of \( \theta \), the MSE of \( \delta_{BG} \) and \( \delta_{BG}^I \) are increasing as \( n_1 \) increases and \( n_2 \) decreases, whereas \( \delta_1 \) and \( \delta_0^{RB} \) have a reverse tendency in this case. For larger value of \( \theta \), \( \delta_1 \) has comparable performance to \( \delta_M \) in terms of MSE, whenever \( n_1 \leq n_2 \) (see Figures 7-8 and Figures S5-6).

(iv) In terms of bias, the two-stage UMVCUE (\( \delta_{BG} \)) and the estimator \( \overline{Y} \) perform better (have zero bias) among all the estimators of \( \mu_S \). After \( \delta_{BG} \) and \( \overline{Y} \), \( \delta_{BG}^I \) performs better than all other estimators and its bias is closer to the bias of UMVCUE \( \delta_{BG} \). Also note that, except \( \delta_{BG} \) and \( \overline{Y} \), all the estimators are positively biased and have maximum bias when \( \theta \) is zero. (see Figures 9-12, Figures S7-8).

(v) Consideration of the bias of estimators is of great significance in adaptive clinical trial designs. As pointed out by (Robertson et al. 2023), using biased estimators may lead to wrong decisions, which could waste limited resources, result in patients not getting effective treatments and put patients at unnecessary risk. However, there is always a trade-off between the variance and the bias and generally decrease in values of one results in an increase in values of the other. The MSE acts as a balancer between the bias and the variance. In practice, it is desirable to find estimators having smaller MSEs and not large values of biases.

It is evident from our simulation study that the bias of the estimator \( \delta_{BG}^I \) is close to that of the UMVCUE, \( \delta_{BG} \). Also, for large values of \( \theta \) and \( n_2 > n_1 \), the MSE of \( \delta_{BG}^I \) is comparable to that of \( \delta_M \), the best performing estimator in terms of the MSE. We also note that, for moderate and large values of \( \theta \), the bias of the estimator \( \delta_M \) is comparable to that of the UMVCUE \( \delta_{BG} \). It may be worth pointing out here that the bias of an estimator near \( \theta = 0 \) should not be of much concern, as in this case the null hypothesis of homogeneity of treatment effects is not likely to be rejected.

It is also interesting to note that, for moderate and large values of \( \theta \), other estimators (except \( \delta_M \) and \( \delta_{BG}^I \)) also have biases comparable to that of UMVCUE, \( \delta_{BG} \), but they have larger MSEs in many cases. (vi) The MSEs and biases of all the competing estimators of \( \mu_S \) approach to zero for large values of the sample sizes \( n_1 \) and \( n_2 \).

When mean squared error is the key criterion for choosing estimators of the selected treatment mean \( \mu_S \), we recommend the estimator \( \delta_M \) for use in practice. In some scenarios, \( \delta_1 \) is a suitable alternative to \( \delta_M \). When bias is the key criterion for choosing estimators, we recommend using \( \delta_{BG} \) and \( \delta_{BG}^I \).

Therefore, in the light of the above discussion based on simulation study, we recommend estimators \( \delta_{BG}^I \) and \( \delta_M \) for estimation of the selected treatment mean \( \mu_S \).

Note: Some of the pertaining to MSE and bias comparisons of various estimators are not included here and have been provided as supplementary material.
Figure 1. Risk plots of different competing estimators for $n_1 = 6, n_2 = 4$.

Figure 2. Risk plots of different competing estimators for $n_1 = 5, n_2 = 5$.

Figure 3. Risk plots of different competing estimators for $n_1 = 3, n_2 = 12$. 

JOURNAL OF BIOPHARMACEUTICAL STATISTICS 179
Figure 4. Risk plots of different competing estimators for $n_1 = 6, n_2 = 18$.

Figure 5. Risk plots of the estimator $\delta_{BG}$ vs. $\delta_{BG}'$ for $n_1 = 8, n_2 = 3$.

Figure 6. Risk plots of the estimator $\delta_{BG}$ vs. $\delta_{BG}'$ for $n_1 = 5, n_2 = 5$. 
Figure 7. Risk plots of different competing estimators for fixed $\theta = 0$ and $n = n_1 + n_2 = 20$.

Figure 8. Risk plots of different competing estimators for fixed $\theta = 0.3$ and $n = n_1 + n_2 = 20$.

Figure 9. Bias plots of different competing estimators for $n_1 = 5$, $n_2 = 3$. 
Figure 10. Bias plots of different competing estimators for $n_1 = 9, n_2 = 4$.

Figure 11. Bias plots of different competing estimators for $n_1 = 5, n_2 = 5$.

Figure 12. Bias plots of different competing estimators for $n_1 = 5, n_2 = 10$. 
Table 1. Percentage risk improvement of estimator $\delta_3$ over $\delta_2$.

| $\theta$ | $(n_1, n_2) = (5, 5)$ | $(n_1, n_2) = (10, 5)$ | $(n_1, n_2) = (10, 10)$ | $(n_1, n_2) = (10, 15)$ |
|----------|------------------------|------------------------|------------------------|------------------------|
| 0.00     | 13.83                  | 7.10                   | 15.01                  | 22.17                  |
| 0.05     | 15.17                  | 6.78                   | 15.41                  | 21.80                  |
| 0.1      | 15.04                  | 6.92                   | 14.80                  | 21.75                  |
| 0.2      | 14.95                  | 7.42                   | 16.40                  | 23.89                  |
| 0.3      | 17.20                  | 8.18                   | 17.09                  | 24.50                  |
| 0.5      | 17.62                  | 8.30                   | 17.92                  | 26.71                  |
| 1.0      | 16.44                  | 4.47                   | 10.55                  | 18.63                  |
| 1.5      | 10.16                  | 0.69                   | 3.33                   | 6.37                   |
| 2.0      | 3.89                   | 0.063                  | 0.51                   | 2.05                   |
| 2.5      | 1.67                   | 0.00                   | 0.049                  | 0.44                   |
| 3.0      | 0.42                   | 0.00                   | 0.00                   | 0.05                   |

6. Real data example

In this section, we illustrate the implementation of the findings of our paper to a data set. For the purpose of illustration, we take a part of data set given in (Abdalghani et al. 2021). (one may also refer the book by (Daniel and Cross 2018)). We have presented it in Table 2 below. The data reflects the serum concentrations (treatment effects) of binding protein-3 after use of growth hormone (GH) and insulin-like growth factor I (IGF-I) on biochemical markers of bone metabolism in patients of idiopathic osteoporosis. Let $\Pi_1$ and $\Pi_2$ denote the populations receiving the treatments GH and IGF-I, respectively. As verified in (Abdalghani et al. 2021), the data are assumed to be normally distributed with different means and a common variance.

It is worth noting that the quality of the two treatments are assessed in terms of their average effects. Therefore, the population corresponding the larger value of mean effect ($\max\{\mu_1, \mu_2\}$) is considered to be more effective. In stage I of the drop-the-losers design, we first draw a sample of size $n_1$ from the two populations, and choose the population corresponding to larger sample mean ($\bar{X}_S = \max\{\bar{X}_1, \bar{X}_2\}$) and drop the population corresponding to smaller sample mean effects ($\bar{X}_{3-S} = \min\{\bar{X}_1, \bar{X}_2\}$). In stage II, we draw another sample of size $n_2$ from the selected population in stage I and name it as $\bar{Y}$. From the data given in Table 2, for $n_1 = 40$ and $n_2 = 26$, the observed average effects corresponding the GH and IGF-I treatments are obtained as: $\bar{X}_1 = 3846.05$,

Table 2. Serum concentrations of binding protein-3 after the treatment GH and IGF-I.

| Stage I data | 4507 | 4072 | 3036 | 2484 | 3540 | 3480 | 2055 |
|--------------|------|------|------|------|------|------|------|
| Stage I data | 4095 | 2315 | 1840 | 2483 | 2354 | 3178 | 3574 |
| Stage I data | 3196 | 2365 | 4136 | 3088 | 3464 | 5874 | 2929 |
| Stage I data | 3903 | 3367 | 2938 | 4142 | 4465 | 3967 | 4213 |
| Stage I data | 4321 | 4990 | 3622 | 6800 | 6185 | 4247 | 4450 |
| Stage I data | 4199 | 5390 | 5188 | 4788 | 4602 |      |      |

Serum concentrations of binding protein-3 after the use of IGF-I treatment.

| Stage II data | 3480 | 3515 | 4003 | 3667 | 4263 | 4797 | 2354 |
| Stage II data | 3570 | 3630 | 3666 | 2700 | 2782 | 3088 | 3405 |
| Stage II data | 3309 | 3444 | 2357 | 3831 | 2905 | 2888 | 2797 |
| Stage II data | 3083 | 3376 | 3464 | 4990 | 4590 | 2989 | 4081 |
| Stage II data | 4806 | 4435 | 3504 | 3529 | 4093 | 4114 | 4445 |
| Stage II data | 3622 | 5130 | 4784 | 4093 | 4852 |      |      |

Serum concentrations of binding protein-3 after the treatment GH.

| Stage III data | 3161 | 4942 | 3222 | 2699 | 3514 | 2963 | 3228 |
| Stage III data | 5995 | 3315 | 2919 | 3235 | 4379 | 5628 | 6152 |
| Stage III data | 4415 | 5251 | 3334 | 3910 | 2304 | 4721 | 3700 |
| Stage III data | 3228 | 2440 | 2698 | 5793 | 4926 |      |      |
Table 3. Various estimates of the selected treatment mean $\mu_S$.

| $\delta_M$ | $\delta_{BG}$ | $\delta_{BG}'$ | $\delta_0$ | $\delta_0'$ | $\delta_1^0$ | $\delta_1$ |
|------------|---------------|----------------|------------|-------------|--------------|------------|
| 3877.484   | 3860.262      | 3862.575       | 3846.05    | 3848.575    | 3850.142     | 3857.382   |

$X_2 = 3710.775$. Further, we obtain $X_S = 3846.05$, $X_{3,S} = 3710.775$, $Y = 3925.846$, $D_1 = -135.275$, $D_2 = 79.796$.

The various estimates of the selected treatment mean $\mu_S$ are tabulated in Table 3 below:

Therefore, on the basis of above analysis, we conclude that among the two treatments GH and IGF-I, GH is selected as the better treatment in Stage I and then it is carried forward to Stage II for further confirmatory analysis. After applying the treatment GH, if one prefers to use the estimator $\delta_M$ then it can be anticipated to have 3877.484 units serum concentration of binding protein-3. If one uses the estimators $\delta_{BG}'$ and $\delta_{BG}$ then the selected treatment GH is expected to have 3862.575 and 3860.262 units of serum concentrations of binding protein 3.

7. **Concluding remarks**

In case of single-stage sampling alone, estimation following selection of treatments is prone to bias, especially for normally distributed data. In the literature there are results, where it has been shown that no unbiased estimator of the selected treatment mean exists (see Putter and Rubinstein 1968, Vellaisamy 2009) and Masihuddin and Misra 2021). To overcome this issue, Cohen and Sackrowitz 1989, (Bowden and Glimm 2008 and (Robertson and Glimm 2019 have used the two-stage adaptive design approach and provided two-stage conditionally unbiased estimator.

Under the criterion of mean squared error, in this paper, we have addressed the problem of efficient estimation of the selected treatment mean under two-stage drop-the losers set up. Our main objective was to look for estimators of the selected treatment mean that perform better in terms of the mean squared error. In this direction we have shown that the maximum likelihood estimator (MLE), which is the weighted average of the first and second stage sample means (with weights being proportional to the corresponding sample sizes), is minimax and admissible. When MSE is the key criterion, we recommend the MLE $\delta_M$ for its use in practice. We also provide some alternative estimators to $\delta_M$, like $\delta_1$ and $\delta_{BG}'$ that also perform well in terms of MSE, under certain configuration of sample sizes.

Under the two-stage DLD, a sufficient condition for inadmissibility of an arbitrary location and permutation equivariant estimator has been derived. As a consequence of which the two-stage UMVCUE, proposed by Bowden and Glimm 2008, is, shown to be inadmissible and a better estimator is obtained. For estimation of the selected treatment mean in case of unknown variance $\sigma^2$, we are still trying to get improved estimators over the UMVCUE. We will address the problem of estimating the selected treatment mean in case of the unknown variance, separately in one of our future projects. The results of this paper can be extended to the case of unequal variances in a straight forward manner. We did not consider unequal variances as under unequal variances the natural selection rule (which chooses the treatment with the largest sample mean as the best treatment) is known to be non-optimal, especially in the case of more than two normal populations or in the non-normal setting (even with two treatments). In certain situations the natural selection rule may even perform worse than the no-data rule, which chooses one of the treatments at random without taking the available data into account (refer Misra and Dhariyal 1994).

Some of the results (like admissibility and minimaxity of the naive estimator) of this paper can be adapted to the case, where a control treatment $\tau_0$, characterized by $N(\mu_0, \sigma^2)$ distribution, is included in the design and it is of interest to estimate the treatment difference $\Delta_S = \mu_S - \mu_0$. Let $X_0$ and $Y_0$ be the sample means of the control arm’s responses at stages I and II, respectively. Several other variations of these designs are also possible, such as (i) trial continues to the second stage only if $X_S - X_0 \geq b$, where $b$ is a specified constant called the futility boundary (see Kimani et al. 2013);
(ii) control treatment is studied only at the second stage

The techniques used in Section 4 can also be extended to such designs, with appropriate changes. We plan to take up these studies in our future research.

The proposed approach in this paper can be easily extended to non-Gaussian data but it will be interesting to extend the results of this paper to data involving more than two treatment groups. Moreover, in a setting of \( k(\geq 2) \) treatments, if more than two best treatments are selected in stage I then it would be interesting to investigate how our methods can be modified to estimate the selected treatment effect. We will make attempts in these directions in our future works.

Acknowledgments

The authors would like to thank the two anonymous referees and the Editor of the journal for their comments and suggestions that have greatly improved this manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The research of the first author is supported by Department of Science & Technology (DST)- INSPIRE, Govt. of India.

ORCID

Masihuddin http://orcid.org/0000-0002-7771-8446

Neeraj Misra http://orcid.org/0000-0002-8412-1904

References

Abdalghani, O., M. Arshad, K. Meena, and A. Pathak. 2021. Adjusted Bias and Risk for Estimating Treatment Effect after Selection with an Application in Idiopathic Osteoporosis. Milton Park, Abingdon, Oxon: CRC Press.

Bahadur, R. R., and L. A. Goodman. 1952. Impartial decision rules and sufficient statistics. *Annals of Mathematical Statistics* 23 (4):553–562. doi:10.1214/aoms/1177729334.

Bauer, P., F. Bretz, V. Dragalin, F. König, and G. Wassmer. 2016. Twenty-five years of confirmatory adaptive designs: Opportunities and pitfalls. *Statistics in Medicine* 35 (3):325–347. doi:10.1002/sim.6472.

Bauer, P., and M. Kieser. 1999. Combining different phases in the development of medical treatments within a single trial. *Statistics in Medicine* 18 (14):1833–1848. doi:10.1002/(SICI)1097-0258(19990730)18:14<1833:AID-SIM221>3.0.CO;2-3.

Bowden, J., and E. Glimm. 2008. Unbiased estimation of selected treatment means in two-stage trials. *Biometrical Journal: Journal of Mathematical Methods in Biosciences* 50 (4):515–527. doi:10.1002/bimj.200810442.

Brewster, J. -F., and J. Zidek. 1974. Improving on equivariant estimators. *Annals of Statistics* 2 (1):21–38. doi:10.1214/aos/1176342610.

Chow, S. -C., and M. Chang. 2006. Adaptive design methods in clinical trials. Boca Raton, Florida: Chapman and Hall/ CRC.

Cohen, A., and H. B. Sackrowitz. 1989. Two stage conditionally unbiased estimators of the selected mean. *Statistics & Probability Letters* 8 (3):273–278. doi:10.1016/0167-7152(89)90133-8.

Dahiya, R. C. 1974. Estimation of the mean of the selected population. *Journal of the American Statistical Association* 69 (345):226–230. doi:10.1080/01621459.1974.10480159.

Daniel, W. W., and C. L. Cross. 2018. *Biostatistics: A Foundation for Analysis in the Health Sciences*. Wiley, USA: Wiley.

Eaton, M. L. 1967. Some optimum properties of ranking procedures. *Annals of Mathematical Statistics* 38 (1):124–137. doi:10.1214/aoms/1177699063.

Hwang, J. T. 1993. Empirical Bayes estimation for the means of the selected populations. *Sankhyā Series A* 55 (2):285–304.

Kimani, P. K., S. Todd, and N. Stallard. 2013. Conditionally unbiased estimation in phase ii/iii clinical trials with early stopping for futility. *Statistics in Medicine* 32 (17):2893–2910. doi:10.1002/sim.5757.
Lu, X., A. Sun, and S. S. Wu. 2013. On estimating the mean of the selected normal population in two-stage adaptive designs. *Journal of Statistical Planning and Inference* 143 (7):1215–1220. doi:10.1016/j.jspi.2013.01.011.

Mashiuddin, and N. Misra. 2021. Equivariant estimation following selection from two normal populations having common unknown variance. *Statistics* 55 (6):1407–1438. doi:10.1080/02331888.2021.2006660.

Misra, N., and I. D. Dhariyal. 1994. Non-minimaxity of natural decision rules under heteroscedasticity. *Statistics and Decisions* 12 (1):79–89. doi:10.1524/sdtr.1994.12.1.79.

Putter, J., and D. Rubinstein. 1968. Technical report No. 165: On estimating the mean of a selected population. *University of Wisconsin Statistics Department, Wisconsin*.

Robertson, D. S., B. Choodari-Oskooei, M. Dimairo, L. Flight, P. Pallmann, and T. Jaki. 2023. Point estimation for adaptive trial designs i: A methodological review. *Statistics in Medicine* 42 (2):122–145. doi:10.1002/sim.9605.

Robertson, D. S., and E. Glimm. 2019. Conditionally unbiased estimation in the normal setting with unknown variances. *Communications in Statistics-Theory and Methods* 48 (3):616–627. doi:10.1080/03610926.2017.1417429.

Sackrowitz, H., and E. Samuel-Cahn. 1986. Evaluating the chosen population: A Bayes and minimax approach. *Lecture Notes-Monograph Series* 386–399.

Sampson, A. R., and M. W. Sill. 2005. Drop-the-losers design: Normal case. *Biometrical Journal: Journal of Mathematical Methods in Biosciences* 47 (3):257–268. doi:10.1002/bimj.200410119.

Sill, M. W. (2000). *Estimation and inference for drop-the-losers designs*. PhD thesis, University of Pittsburgh.

Sill, M. W., and A. R. Sampson. 2007. Extension of a two-stage conditionally unbiased estimator of the selected population to the bivariate normal case. *Communications in Statistics—Theory and Methods* 36 (4):801–813. doi:10.1080/03610920601034072.

Tappin, L. 1992. Unbiased estimation of the parameter of a selected binomial population. *Communications in Statistics-Theory and Methods* 21 (4):1067–1083. doi:10.1080/03610929208830831.

Thall, P. F., R. Simon, and S. S. Ellenberg. 1988. Two-stage selection and testing designs for comparative clinical trials. *Biometrika* 75 (2):303–310. doi:10.1093/biomet/75.2.303.

Thall, P. F., R. Simon, and S. S. Ellenberg. 1989. A two-stage design for choosing among several experimental treatments and a control in clinical trials. *Biometrics* 45 (2):537–547. doi:10.2307/2531495.

Vellaaisamy, P. 2009. A note on unbiased estimation following selection. *Statistical Methodology* 6 (4):389–396. doi:10.1016/j.stamet.2008.12.001.

Wu, S. S., W. Wang, and M. C. Yang. 2010. Interval estimation for drop-the-losers designs. *Biometrika* 97 (2):405–418. doi:10.1093/biomet/asq003.
Appendix

Proof of Lemma 2.2: (i) Let us introduce the auxiliary random variables $Y_1$ and $Y_2$ such that $X_1, X_2, Y_1$ and $Y_2$ are mutually independent and $Y_i \sim N \left( \mu_i, \frac{\sigma^2}{n_i} \right), i = 1, 2.$ Let $V_1 = \frac{\sqrt{n_1}}{\sigma} (X_1 - \mu_1), \ V_2 = \frac{\sqrt{n_2}}{\sigma} (X_2 - \mu_2), \ V_3 = \frac{\sqrt{n_3}}{\sigma} (Y_1 - \mu_1), \ V_4 = \frac{\sqrt{n_4}}{\sigma} (Y_2 - \mu_2)$ and $w = \frac{n_1}{n_1 + n_2}.$ Then, the c.d.f. of

$$U = \frac{n_1 X_1 + n_2 Y}{n_1 + n_2} - \mu_5 = wX_1 + (1 - w)Y - \mu_5$$

is

$$F_{U \mid \mu}(u)$$

$$= P_{\mu} \left( wX_1 + (1 - w)Y - \mu_5 \leq u \right)$$

$$= P_{\mu} \left( X_1 < X_2, wX_2 + (1 - w)Y_2 - \mu_2 \leq u \right) + P_{\mu} \left( X_1, wX_1 + (1 - w)Y_1 - \mu_1 \leq u \right)$$

$$= h_\mu(\mu_1, \mu_2) + h_\mu(\mu_2, \mu_1), -\infty < u < \infty.$$ 

(7.1)

Due to symmetry, we may assume that $\mu_1 \leq \mu_2,$ so that $\mu_2 - \mu_1 = \theta.$ We have, for $-\infty < u < \infty$

$$h_\mu(\mu_1, \mu_2) = P_{\mu} \left( X_1 < X_2, wX_2 + (1 - w)Y_2 - \mu_2 \leq u \right)$$

$$= P_{\mu} \left( V_1 - V_2 < \frac{\sqrt{n_1}}{\sigma} (\mu_2 - \mu_1), \ \frac{w\sigma}{\sqrt{n_1}} V_2 + \frac{(1 - w)\sigma}{\sqrt{n_2}} V_4 \leq u \right)$$

$$= P_{\mu} \left( B_1 \leq \frac{\sqrt{n_1}}{\sigma} \theta, B_2 \leq u \right), \ \theta \geq 0,$$ 

(7.2)

where, $B_1 = V_1 - V_2 \sim N(0, 2), \ B_2 = \frac{w\sigma}{\sqrt{n_1}} V_2 + \frac{(1 - w)\sigma}{\sqrt{n_2}} V_4 \sim N(0, \sigma^2),$ and

$$\begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \sim N_2 \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \ \begin{bmatrix} 2 \frac{w\sigma}{\sqrt{n_1}} - \frac{w\sigma}{\sqrt{n_2}} & \frac{w\sigma}{\sqrt{n_1}} \frac{w\sigma}{\sqrt{n_2}} \\ \frac{w\sigma}{\sqrt{n_1}} & \frac{w\sigma}{\sqrt{n_2}} \end{bmatrix} \right).$$

Then, the conditional distribution of $B_1$ given that $B_2 = t(t \in \mathbb{R}),$ follows $N \left( -\frac{1}{\sigma^2} \frac{t}{\sqrt{n_1}}, 1 + \rho \right)$ where, $\rho = \frac{n_2}{n_1 + n_2}.$ Therefore, (7.2) becomes

$$h_\mu(\mu_1, \mu_2) = \int_{-\infty}^{\mu} \Phi \left( \frac{\sqrt{n_1} (t + \theta)}{\sigma \sqrt{1 + \rho}} \right) \frac{1}{\sigma_2} \phi \left( \frac{t}{\sigma_2} \right) dt, -\infty < u < \infty.$$ 

(7.3)

and, by symmetry, we get

$$h_{\mu}(\mu_2, \mu_1) = \int_{-\infty}^{\mu} \Phi \left( \frac{\sqrt{n_1} (t - \theta)}{\sigma \sqrt{1 + \rho}} \right) \frac{1}{\sigma_2} \phi \left( \frac{t}{\sigma_2} \right) dt, -\infty < u < \infty.$$ 

(7.4)

Consequently, using (7.1), (7.3) and (7.4), the c.d.f. of $U$ is

$$F_{U \mid \mu}(u) = \int_{-\infty}^{\mu} \left[ \Phi \left( \frac{\sqrt{n_1} (u + \theta)}{\sigma \sqrt{1 + \rho}} \right) + \Phi \left( \frac{\sqrt{n_1} (u - \theta)}{\sigma \sqrt{1 + \rho}} \right) \right] \frac{1}{\sigma_2} \phi \left( \frac{u}{\sigma_2} \right) du, -\infty < u < \infty.$$ 

(7.5)

Hence, the assertion follows.

(ii) Let $A = \frac{\sqrt{n_1} \theta}{\sigma \sqrt{1 + \rho}}$ and $B = \frac{\sqrt{n_2} \theta}{\sigma \sqrt{1 + \rho}}.$ For any $\mu \in \Theta,$ using Lemma 2.1(iii), we have,

$$E_{\mu}(U^2) = \int_{-\infty}^{\infty} \frac{u^2}{\sigma_2} \left[ \Phi \left( \frac{\sqrt{n_1} (u + \theta)}{\sigma \sqrt{1 + \rho}} \right) + \Phi \left( \frac{\sqrt{n_1} (u - \theta)}{\sigma \sqrt{1 + \rho}} \right) \right] \phi \left( \frac{u}{\sigma_2} \right) du$$

(7.6)
Similarly, $\mu, d_1 \Rightarrow \mu, 2$ so by (2) $h \leq k$.

Lemma 1, by $h(k)$.

Let $\mathbb{P}, \mathbb{V}, V_1, V_2, V_3$ and $V_4$ be as defined in the proof of Lemma 2.2. Also, due to symmetry, we may assume that $\mu_1 < \mu_2$, so that $\theta = \mu_2 - \mu_1$. Then, for sufficiently small $h > 0, k > 0$

\[ N_1(h, k) = \mathbb{P}(Z \leq z, d_1 - h < D_1 \leq d_1, d_2 - k < D_2 \leq d_2) \]

\[ \mathbb{P}_h(Z \leq z, d_1 - h < X_3 - X_2 \leq d_1, d_2 - k < Y - X_3 \leq d_2) \]

\[ \mathbb{P}_h(X \geq X_1, X_2 - \mu_1 \leq z, d_1 - h < X_2 - X_1 \leq d_1, d_2 - k < Y_1 - X_1 \leq d_2) \]

\[ \mathbb{P}_h(X_2 - \mu_2 \leq z, d_1 - h < X_1 \leq d_1, d_2 - k < Y_2 - X_2 \leq d_2) \]

\[ \mathbb{P}_h(V_1 \leq \frac{\sqrt{\mu_1}}{\sigma} \leq z, V_2 + \frac{\sqrt{\mu_1}}{\sigma} (d_1 - h - \theta) < V_2 \leq V_1 + \frac{\sqrt{\mu_1}}{\sigma} (d_1 - \theta), \]

\[ \frac{\sqrt{\mu_2}}{\sigma} \left( \frac{\sigma}{\sqrt{\mu_1}} V_1 + d_2 - k \right) \]

\[ \mathbb{P}_h(V_1 \leq \frac{\sqrt{\mu_1}}{\sigma} \leq z, V_2 + \frac{\sqrt{\mu_1}}{\sigma} (d_1 - h + \theta) < V_2 \leq V_1 + \frac{\sqrt{\mu_1}}{\sigma} (d_1 - \theta), \]

\[ \frac{\sqrt{\mu_2}}{\sigma} \left( \frac{\sigma}{\sqrt{\mu_1}} V_2 + d_2 - k \right) \]

\[ \int_{-\infty}^{+\infty} \left[ \Phi \left( v + \frac{\sqrt{\mu_1}}{\sigma} (d_1 - \theta) \right) - \Phi \left( v + \frac{\sqrt{\mu_1}}{\sigma} (d_1 - h - \theta) \right) \right] \]

\[ \times \left[ \Phi \left( \frac{\sqrt{\mu_2}}{\sigma} \left( \frac{\sigma}{\sqrt{\mu_1}} v + d_2 \right) \right) - \Phi \left( \frac{\sqrt{\mu_2}}{\sigma} \left( \frac{\sigma}{\sqrt{\mu_1}} v + d_2 - k \right) \right) \right] \phi(v) dv 

\[ + \int_{-\infty}^{+\infty} \left[ \Phi \left( v + \frac{\sqrt{\mu_1}}{\sigma} (d_1 + \theta) \right) - \Phi \left( v + \frac{\sqrt{\mu_1}}{\sigma} (d_1 - h + \theta) \right) \right] \]

\[ \times \left[ \Phi \left( \frac{\sqrt{\mu_2}}{\sigma} \left( \frac{\sigma}{\sqrt{\mu_1}} v + d_2 \right) \right) - \Phi \left( \frac{\sqrt{\mu_2}}{\sigma} \left( \frac{\sigma}{\sqrt{\mu_1}} v + d_2 - k \right) \right) \right] \phi(v) dv. \] (7.7)

Similarly,
\[ N_x(h, k) \]
\[ = \int_{-\infty}^{\infty} \left[ \Phi \left( t + \frac{\sqrt{n_1}}{\sigma} (d_1 - \theta) \right) - \Phi \left( t + \frac{\sqrt{n_1}}{\sigma} (d_1 - h - \theta) \right) \right] \]
\[ \times \left[ \Phi \left( \frac{\sqrt{n_2}}{\sigma} \left( \frac{\sigma}{\sqrt{n_1}} t + d_2 \right) \right) - \Phi \left( \frac{\sqrt{n_2}}{\sigma} \left( \frac{\sigma}{\sqrt{n_1}} t + d_2 - k \right) \right) \right] \phi(t) dt \]
\[ + \int_{-\infty}^{\infty} \left[ \Phi \left( t + \frac{\sqrt{n_1}}{\sigma} (d_1 + \theta) \right) - \Phi \left( t + \frac{\sqrt{n_1}}{\sigma} (d_1 - h + \theta) \right) \right] \]
\[ \times \left[ \Phi \left( \frac{\sqrt{n_2}}{\sigma} \left( \frac{\sigma}{\sqrt{n_1}} t + d_2 \right) \right) - \Phi \left( \frac{\sqrt{n_2}}{\sigma} \left( \frac{\sigma}{\sqrt{n_1}} t + d_2 - k \right) \right) \right] \phi(t) dt. \] (7.8)

Using (7.6)-(7.8), and the L' Hopital rule, we get for fixed \( d_1 \in (-\infty, 0] \) and \( d_2 \in \mathbb{R} \)

\[ F_{x, y}(z|d_1, d_2) \]
\[ = \int_{-\infty}^{\infty} \left[ \phi \left( \frac{\sqrt{n_1}}{\sigma} (z + d_1 - \theta) \right) + \phi \left( \frac{\sqrt{n_1}}{\sigma} (z + d_1 + \theta) \right) \right] \phi \left( \frac{\sqrt{n_2}}{\sigma} (z + d_2) \right) \phi \left( \frac{\sqrt{n_1}}{\sigma} z \right) \phi(t) dt \]
\[ = \int_{-\infty}^{\infty} \left[ \phi \left( \frac{\sqrt{n_1}}{\sigma} (t + d_1 - \theta) \right) + \phi \left( \frac{\sqrt{n_1}}{\sigma} (t + d_1 + \theta) \right) \right] \phi \left( \frac{\sqrt{n_2}}{\sigma} (t + d_2) \right) \phi \left( \frac{\sqrt{n_1}}{\sigma} t \right) dt \]

Consequently, for any fixed \( d_1 \in (-\infty, 0] \) and \( d_2 \in \mathbb{R} \), the conditional p.d.f. of \( Z \) given \( (D_1, D_2) = (d_1, d_2) \), is given by

\[ f_{x, y}(z|d_1, d_2) \]
\[ = e^{-\frac{n_1}{2(2n_1 + n_2)}} \frac{(n_1 + n_2)(d_1 - n_2d_2)\theta}{2(2n_1 + n_2)} \frac{\sqrt{n_1 + n_2}}{\sigma} \phi \left( \frac{\sqrt{n_1 + n_2}}{\sigma} \left( z + n_1d_1 - n_2d_2 - n_1 \theta \right) \right) \]
\[ + e^{-\frac{n_1}{2(2n_1 + n_2)}} \frac{(n_1 + n_2)(d_1 - n_2d_2)\theta}{2(2n_1 + n_2)} \frac{\sqrt{n_1 + n_2}}{\sigma} \phi \left( \frac{\sqrt{n_1 + n_2}}{\sigma} \left( z + n_1d_1 + n_2d_2 + n_1 \theta \right) \right) \]
\[ = e^{-\frac{n_1}{2(2n_1 + n_2)}} \frac{(n_1 + n_2)(d_1 - n_2d_2)\theta}{2(2n_1 + n_2)} \frac{\sqrt{n_1 + n_2}}{\sigma} \phi \left( \frac{\sqrt{n_1 + n_2}}{\sigma} \left( z + n_1d_1 + n_2d_2 \right) \right) \]
\[ - e^{-\frac{n_1}{2(2n_1 + n_2)}} \frac{(n_1 + n_2)(d_1 - n_2d_2)\theta}{2(2n_1 + n_2)} \frac{\sqrt{n_1 + n_2}}{\sigma} \phi \left( \frac{\sqrt{n_1 + n_2}}{\sigma} \left( z + n_1d_1 - n_2d_2 \right) \right) \]
\[ - e^{-\frac{n_1}{2(2n_1 + n_2)}} \frac{(n_1 + n_2)(d_1 - n_2d_2)\theta}{2(2n_1 + n_2)} \frac{\sqrt{n_1 + n_2}}{\sigma} \phi \left( \frac{\sqrt{n_1 + n_2}}{\sigma} \left( z + n_1d_1 + n_2d_2 \right) \right) \]
\[ - e^{-\frac{n_1}{2(2n_1 + n_2)}} \frac{(n_1 + n_2)(d_1 - n_2d_2)\theta}{2(2n_1 + n_2)} \frac{\sqrt{n_1 + n_2}}{\sigma} \phi \left( \frac{\sqrt{n_1 + n_2}}{\sigma} \left( z + n_1d_1 - n_2d_2 \right) \right) \], \(-\infty < z < \infty\).

(b) Using (a), for any fixed \( d_1 \in (-\infty, 0] \) and \( d_2 \in \mathbb{R} \)

\[ \mathbb{E}_x(Z|(D_1, D_2) = (d_1, d_2)) \]
\[ = \frac{n_1 \theta}{2n_1 + n_2} \left\{ 1 - e^{-\frac{n_1}{2(2n_1 + n_2)}} \frac{(n_1 + n_2)(d_1 - n_2d_2)\theta}{2(2n_1 + n_2)} \frac{\sqrt{n_1 + n_2}}{\sigma} \phi \left( \frac{\sqrt{n_1 + n_2}}{\sigma} \left( z + n_1d_1 + n_2d_2 \right) \right) \right\} - n_1d_1 + n_2d_2 \]
\[ \quad \times \left\{ 1 + e^{-\frac{n_1}{2(2n_1 + n_2)}} \frac{(n_1 + n_2)(d_1 - n_2d_2)\theta}{2(2n_1 + n_2)} \frac{\sqrt{n_1 + n_2}}{\sigma} \phi \left( \frac{\sqrt{n_1 + n_2}}{\sigma} \left( z + n_1d_1 - n_2d_2 \right) \right) \right\} - n_1d_1 + n_2d_2 \]