mathematics

MDPI

Article

A Purely Sequential Selection and Its Acceleration for the Best Treatment Under a New Formulation of the Indifference Zone: Asymptotic Theory and Simulations

Nitis Mukhopadhyay

Department of Statistics, University of Connecticut, Austin Building U-4120, 215 Glenbrook Road, Storrs, CT 06269-4120, USA; nitis.mukhopadhyay@uconn.edu

Abstract

A traditional *indifference zone* (IZ) approach selects the best treatment, with a preset high probability of correct selection when the best one is much better than the second-best in line. Customarily, the effect size is the difference between the best (largest mean) and the second-best treatment under consideration from $k(\geq 2)$ normal populations with unknown means and equal but unknown variance σ^2. We argue that an effect size could involve unknown means, as well as unknown variance σ^2, because it is sometimes unrealistic to assess the effect size and preset it in the absence of σ^2. Under a new IZ formulation, we rebuild both purely sequential and accelerated sequential selection methodologies for the largest mean and associated *first-order* (f.o.) and *second-order* (s.o.) asymptotics. Such a breadth of rich asymptotic theory is supplemented by a thorough discussion of the performance of the two selection strategies from a set of simulations when the sample sizes are small, medium, or large.

Keywords: acceleration; best treatment selection; correct selection; data analysis; effect-size; indifference zone (IZ); least favorable configuration (LFC); multiple comparisons; preference zone (PZ); purely sequential; sample size determination; save resources; second-order approximations; sequential selection; simulations

MSC: 62L10; 62L12

1. Introduction

A serious debate about philosophies of selection and ranking began in the mid-1940s through the mid-1950s, breaking new fundamental ground and developing an approach that was markedly different from customary *analysis of variance* (ANOVA) methods. We may cite, as examples, Wald (1947) [1], Bahadur (1950) [2] and Bahadur and Robbins (1950) [3]. From this remarkably bright and energetic group of researchers, Bechhofer (1954) [4] and Gupta (1956) [5] independently developed two fundamentally different formulations, known as the *indifference zone* (IZ) approach [4] and the *subset selection* approach (Gupta 1956) [5].

These pioneering formulations changed the statistical landscape (Bechhofer, Dunnett and Sobel 1954a [6]; Dunnett and Sobel 1955 [7]), leading to steadfast growth that has continued beyond anyone's wildest imagination. Such valiant efforts enriched the discipline of multiple comparisons (Tukey 1991, 1993 [8,9]; Hochberg and Tamhane 1987 [10]; Hsu 1996 [11]). The combined literature is vast as well as deep, handling nearly every

Check for updates

Academic Editor: Manuel Alberto M. Ferreira

Received: 31 July 2026
Revised: 27 August 2026
Accepted: 8 September 2026
Published: 2 October 2026

Copyright: © 2026 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

conceivable topic across the whole breadth of statistical science, and producing with ground-breaking theory, methods and applications.

One can get a glimpse of the depth of this field by revisiting and comparing topics of interest from some of the citations included in the following list of major publications, highlighted chronologically: Gupta and Sobel (1957) [12]; Bechhofer, Kiefer and Sobel (1968) [13]; Robbins, Sobel and Starr (1968) [14]; Dudewicz and Dalal (1975) [15]; Gibbons, Olkin and Sobel (1977) [16]; Gupta and Panchapakesan (1979) [17]; Gupta and Huang (1981) [18]; Dudewicz and Koo (1981) [19]; Mukhopadhyay and Solanky (1994) [20]; Bechhofer, Santner and Goldsman (1995) [21]; Mukhopadhyay (2000 [22], chapter 13); and Mukhopadhyay and de Silva (2009 [23], chapter 14).

The present paper revisits Bechhofer's (1954) [4] pioneering IZ approach in selection and ranking methodologies by returning to and refining the problem of identifying the best treatment from a set of $k(\geq 2)$ competing treatments with desirable properties. In order to be specific, we begin at the origin by briefly laying out the original formulation from Bechhofer.

1.1. A Glimpse of Bechhofer's Original Formulation

We suppose that we have a population Π_i corresponding to the responses obtained from the i^{th} treatment under which a sequence of observations $\{X_{ij}; j \geq 1\}$ arrives successively, or in groups, as *independent and identically distributed* (i.i.d.) $N(\mu_i, \sigma^2)$ random variables, $i = 1, ..., k(\geq 2)$. We assume that the responses from $\Pi_1, ..., \Pi_k$ arrive independently across the board and that all k treatment means $\mu_1, ..., \mu_k$ as well as the common variance σ^2 remain unknown, where the parameter spaces are

$$\boldsymbol{\mu} \equiv (\mu_1, ..., \mu_k) \in R^k \text{ and } \sigma \in R^+ = (0, \infty).$$

One may interpret $\boldsymbol{\mu}$ and σ as a $(k+1)$-dimensional parameter vector $\boldsymbol{\theta} \equiv (\boldsymbol{\mu}, \sigma)$ with its domain space given by $R^k \times R^+$.

Having recorded n observations $\{X_{ij}; j = 1, ..., n(\geq 2)\}$ obtained from Π_i under a fixed-sample-size sampling design, we denote

$$\begin{aligned} \text{The } i^{\text{th}} \text{ sample mean:} \quad & \overline{X}_{in} \equiv n^{-1}\Sigma_{j=1}^{n} X_{ij}; \\ \text{The } i^{\text{th}} \text{ sample variance:} \quad & S_{in}^2 \equiv (n-1)^{-1}\Sigma_{j=1}^{n}(X_{ij} - \overline{X}_{in})^2; \\ \text{The pooled sample variance:} \quad & U_n^2 \equiv k^{-1}\Sigma_{i=1}^{k} S_{in}^2; \end{aligned} \tag{1}$$

with $i = 1, ..., k$.

We record the sample mean $\overline{X}_{in}$ and the sample variance S_{in}^2 from $\Pi_i, i = 1, ..., k$, obtained from the full observed data:

$$\boldsymbol{X}_n \equiv \{X_{ij}; i = 1, ..., k(\geq 2); j = 1, ..., n\} \text{ for } n = 2, 3, ...$$

We note that we expect no ties when considering the observed treatment mean values $\overline{X}_{in}, i = 1, ..., k$, since the responses are continuous variables. Throughout this discourse, we denote

$$\phi(z) = (\sqrt{2\pi})^{-1} \exp\{-\tfrac{1}{2}z^2\} \text{ and } \Phi(z) = \textstyle\int_{-\infty}^{z} \phi(x)dx, z \in R, \tag{2}$$

respectively, for the *probability density function* (p.d.f.) and the *distribution function* (d.f.) of a standard normal random variable.

Next, suppose that $\mu_{[1]} \leq \mu_{[2]} \leq ... \leq \mu_{[k-1]} \leq \mu_{[k]}$ respectively denote the ordered treatment mean parameters. We may express

$$P_{\Pi_i}\{X_{i1} \geq x\} = 1 - \Phi\left(\frac{x-\mu_i}{\sigma}\right) = a_i(x), \text{ say, for any fixed } x \in R, \tag{3}$$

$i = 1, ..., k$. Now, let us focus on any single particular x value and maximize $a_i(x)$ with respect to its associated location parameter, across $i = 1, ..., k$. Obviously, we have

$$\frac{\partial}{\partial\theta}\left\{1 - \Phi\left(\frac{x-\theta}{\sigma}\right)\right\} = \frac{1}{\sigma}\phi\left(\frac{x-\theta}{\sigma}\right) > 0,$$

showing that the $1 - \Phi\left(\frac{x-\theta}{\sigma}\right)$ function (in θ) is $\uparrow$ in $\theta(\in R)$. That is, $P_{\Pi_i}\{X_{i1} \geq x\}$ would coincide with $\max_{1\leq l\leq k} P_{\Pi_l}\{X_{l1} \geq x\}$ for any fixed $x \in R$ *if and only if* Π_i's mean μ_i happens to be exactly $\mu_{[k]}$, the largest treatment mean. Figure 1 presents this sentiment pictorially when $k = 4$. Clearly, for all x values within our consideration, the curve corresponding to $a_4(x)$ associated with $\mu_4 (= 6.8 = \mu_{[4]})$ clearly stays on top of the other curves corresponding to $a_1(x), a_2(x), a_3(x)$.

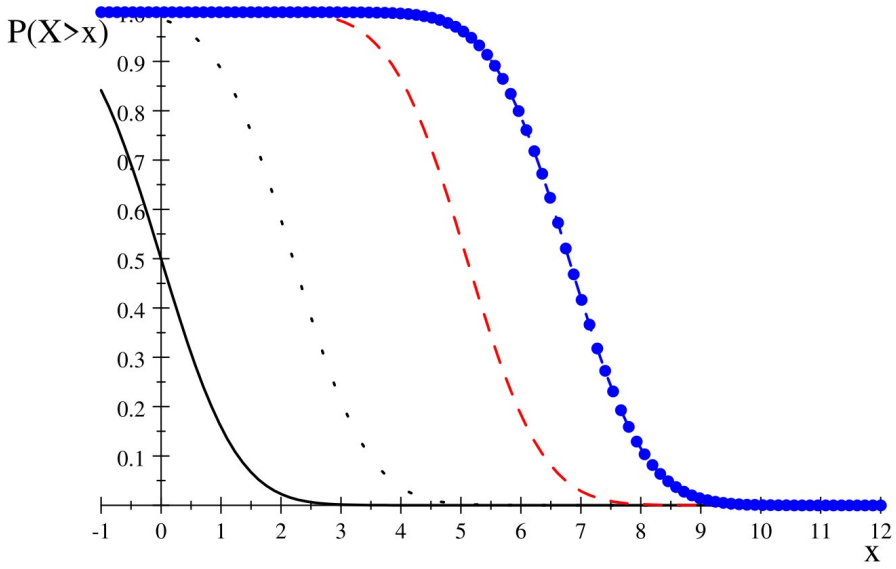


Figure 1. Plots of $a_i(x)$ from (3) associated with $\Pi_i \equiv N(\mu_i, \sigma^2), i = 1, ..., k(= 4)$, under fixed values $\mu_1 = 0, \mu_2 = 2.2, \mu_3 = 5.1$, and $\mu_4 = 6.8$ respectively. Going from left to right, highlights: $a_1(x)$ with a solid curve, $a_2(x)$ with a dotted curve, $a_3(x)$ with a dashed curve, and $a_4(x)$ with a dotted-dashed curve.

If we put a different spin on this by assuming that a larger sample mean is more significant than a smaller sample mean, then the "best" treatment or population ought to be defined as the one associated with the largest treatment mean, $\mu_{[k]}$. For obvious reasons, however, we may not pay much attention to a dual problem of selecting the "worst" treatment, namely the one associated with the smallest mean parameter, $\mu_{[1]}$.

Indeed, one may **define**:

$$\text{a population } \Pi_i, \text{ i.e., the } i^{\text{th}} \text{ treatment, is the } \textit{best} \text{ one if } \mu_i = \mu_{[k]}, \tag{4}$$

which led Bechhofer (1954) [4] to propose the following *fixed-sample-size* (FSS) **decision rule**:

$$Q_n\text{: Declare that } \Pi_i, \text{ the } i^{\text{th}} \text{ treatment, is the } \textit{best} \text{ if and only if } \overline{X}_{in} = \max_{1\leq j\leq k} \overline{X}_{jn}, \tag{5}$$

based on all recorded responses $\boldsymbol{X}_n$.

One may let (i) $\Pi_{(i)}$ stand for the population associated with the i^{th} ordered mean $\mu_{[i]}$ and (ii) $\overline{X}_{(in)}$ denote the sample mean from $\Pi_{(i)}$ without assuming any apriori knowledge regarding the correspondences between $(\Pi_1, ..., \Pi_k)$ and $(\Pi_{(1)}, ..., \Pi_{(k)})$. Under the customary FSS selection rule Q_n from (4), one's final pick of the best treatment based on k observed sample means would be *correct* if and only if $\overline{X}_{(kn)}$ exceeds $\overline{X}_{(in)}$, obtained respectively from $\Pi_{(k)}$ and $\Pi_{(i)}, \forall i = 1, ..., k-1$. One frequently refers to Q_n as a "natural" selection rule perhaps since both maximum likelihood rule and the Bayes rule coincide with Q_n for every fixed n. In this literature, the abbreviation CS commonly refers to *correct selection*.

When one focuses on the whole parameter space R^k for $\boldsymbol{\mu}$, Equation (8) in Section 2 will subsequently emphasize the following important point: One would *not* be able to conclude that the associated $P_n(CS \mid Q_n)$ may reach some arbitrarily preset high target level $P^* \in (k^{-1}, 1)$ considering all $\boldsymbol{\mu}$ in R^k. Bechhofer (1954) [4] forcefully argued that an experimenter should be mindful of this situation, thereby identifying the best treatment when the best treatment is way ahead of the next best treatment, with a guarantee that $P_n(CS \mid Q_n) \geq P^*$, or close to P^*, with a preset large value of $P^* \in (k^{-1}, 1)$.

That led Bechhofer (1954) [4] to propose the original notion of a *preference-zone* (PZ_{Orig}), defined as the complimentary parameter subspace $R^k -$ IZ_{Orig} after appropriately formulating his original IZ, namely IZ_{Orig}. We add a brief discourse in Section 2. Then, we precisely address the motivation and specification of a new class of IZ and PZ, respectively denoted by IZ_{New} and PZ_{New}, in Section 3. We make it abundantly clear how Bechhofer's (1954) [4] classical (IZ_{Orig},PZ_{Orig}) formulation would occupy an integral part of our enlarged new class of specifications (IZ_{New},PZ_{New}) under a microscope.

Is This Research Innovative?

Bechhofer (1954) [4] introduced selection methodologies by proposing his seminal approach by addressing situations when $\mu_{[k]} - \mu_{[k-1]} \geq \Delta \equiv \delta^*$, involving a preassigned number $\delta^* (> 0)$. Pretending that larger differences continue to be more significant, the originality or novelty in this work is that we introduce comparing $\mu_{[k]} - \mu_{[k-1]}$ with $\Delta \equiv \delta^* \sigma^\beta$, $0 \leq \beta < 1$.

For example, when $\sigma = 5$, a smaller magnitude of $\mu_{[k]} - \mu_{[k-1]}$ may be significant, but instead if σ was truly 8, then one would expect a larger magnitude of $\mu_{[k]} - \mu_{[k-1]}$ to be significant as indicated by a larger effect-size Δ. Tables 4 and 5, compiled from simulations, exhibit this interesting interplay between the μ's and σ while one compares the associated Δ-values across Blocks (i)–(iii).

This constitutes an innovative idea with practical merits giving rise to an opportunity to build a solid theory around it. We emphasize that as a pragmatic consideration, we do not shut the door to Bechhofer's classic and original approach allowing a possible choice $\beta = 0$, that is using $\Delta \equiv \delta^*$, if an experimenter so desires.

1.2. A Layout

In what follows, we show a full layout of all sections and sub-sections.

1. Introduction with Historical Notes
   - 1.1. A Glimpse of Bechhofer's Original Formulation
   - (a) Is This Research Innovative?
   - 1.2. A Layout
2. Preliminaries on Fixed-Sample-Size (FSS) Selection Rule
3. Formulation of (PZ_{New}, IZ_{New}) Approach
   - 3.1. ($\text{PZ}_{\text{Orig}}(\delta^*)$,$\text{IZ}_{\text{Orig}}(\delta^*)$) Approach Is Included in (10) and (11)

3.2. Following Up on $P_n(CS \mid Q_n)$
3.3. Optimal Fixed Sample Size C
3.4. Why Restricting $\beta \in [0,1)$?
3.5. Identifying the Requisite h_{k,P^*} Satisfying (15)
3.6. Expression of $P(CS)$ Under Multi-Stage Sampling
4. A Purely Sequential Selection Strategy
4.1. Asymptotic First-Order (F.O.) Properties
4.1.1. Behavior of $C^{-1}N_{\text{PurSeq}}$
4.1.2. Behavior of the Moments of $C^{-1}N_{\text{PurSeq}}$
4.1.3. Behavior of $P(CS \mid Q_N)$
4.2. Asymptotic Second-Order (S.O.) Properties
4.2.1. Elements of Sequential Nonlinear Renewal Theory
4.2.2. Key Asymptotics on Moments of N_{PurSeq}
i. Outlining Proofs for (45) Parts (i)–(iii)
4.2.3. Approximation of $P(CS \mid Q_N)$ Under the LFC
ii. Outlining a Proof of (51)
4.2.4. Verification of the Mapping from (28) to (40)
5. An Accelerated Sequential Selection Strategy
6. Summaries from Simulations
6.1. Fixing Initial Input
6.2. Running Simulations
6.3. An overview of Data Analyses
(a) Conclusions from Data Analyses
7. Action Items in the Horizon: Future Research

Funding
Data Availability Statement
Acknowledgements
Conflicts of Interest
References

2. Preliminaries on Fixed-Sample-Size (FSS) Selection Rule

Let us denote $\Delta_i = \mu_{[k]} - \mu_{[i]}$ and we recall that $\Pi_{(i)}$ stands for the population associated with the i^{th} ordered mean $\mu_{[i]}$ for $i = 1, ..., k-1$. Again, we make no apriori assumption regarding any mapping between the indices of $(\Pi_1, ..., \Pi_k)$ and those of $(\Pi_{(1)}, ..., \Pi_{(k)})$. In the context of $\Pi_{(j)}$, we also define the associated standardized random variables:

$$Z_{(jn)} = n^{1/2}\left(\overline{X}_{(jn)} - \mu_{[j]}\right)/\sigma, j = 1, ..., k, \tag{6}$$

which are i.i.d. standard normal random variables.

Then, under the customary *fixed-sample-size* (FSS) selection rule Q_n from (5), we can express the probability of correct selection $\forall \boldsymbol{\mu} \in R^k$ and $\forall \sigma \in R^+$ as follows:

$$\begin{aligned} & P_n(CS \mid Q_n) \\ & = P_n\left\{\overline{X}_{(kn)} > \overline{X}_{(in)}, i = 1, ..., k-1\right\} \\ & = P_n\left\{Z_{(kn)} > Z_{(in)} - n^{1/2}\Delta_i\sigma^{-1}, i = 1, ..., k-1\right\} \text{ with } \Delta_i = \mu_{[k]} - \mu_{[i]} \\ & = \int_{-\infty}^{\infty}\left\{\Pi_{i=1}^{k-1}\Phi(z + n^{1/2}\Delta_i\sigma^{-1})\right\}\phi(z)dz, \end{aligned} \tag{7}$$

so that we must have:

$$\inf_{\boldsymbol{\mu}\in R^k} P_n(CS \mid Q_n) = \int_{-\infty}^{\infty} \Phi^{k-1}(z)\phi(z)dz = k^{-1}, \tag{8}$$

since $\Delta_i \geq 0$ for all $i = 1, ..., k-1$ when $\boldsymbol{\mu} \in R^k$. Here, and elsewhere, such infimum in $\boldsymbol{\mu}$ holds pointwise in σ.

Clearly, the $\inf_{\boldsymbol{\mu}\in R^k}$ in (8) would be attained if $\boldsymbol{\mu}$ corresponds to:

$$\begin{gathered}\text{the } \textit{least favorable configuration} \text{ (LFC), that is, if } \boldsymbol{\mu} \in R^k \\ \text{satisfies: } \mu_1 = ... = \mu_k \Leftrightarrow \Delta_1 = ... = \Delta_{k-1} = 0.\end{gathered} \tag{9}$$

This LFC (9) indicates that all treatment means from $\Pi_1, ..., \Pi_k$ identically coincide with the best treatment. In other words, under the LFC, any treatment from $\Pi_1, ..., \Pi_k$ may be identified as the best treatment by incorporating some simple randomization which is the spirit of (8). That is, on the whole parameter space R^k for $\boldsymbol{\mu}$, one will *not* be able to claim that the associated $P(CS \mid Q_n)$ may reach a desirable and preset high value P^*, having fixed an arbitrary $P^* \in (k^{-1}, 1)$ before data collection.

Bechhofer (1954) [4] emphasized that it would be more practical to identify the best treatment with a guarantee that $P(CS \mid Q_n) \geq P^*$ or approximately so, having some arbitrarily preset large value $P^* \in (k^{-1}, 1)$. But, then, one must focus within an appropriate parameter sub-space of R^k making it worthwhile in practice to identify the best treatment when that best treatment stands way ahead of all other $k-1$ treatments, at a desired preset level P^* as $P(CS)$.

Bechhofer argued in favor of selection and ranking methodologies by proposing his seminal (PZ_{Orig},IZ_{Orig}) approach corresponding to situations when $\mu_{[k]} - \mu_{[k-1]} \geq \Delta$ or $\mu_{[k]} - \mu_{[k-1]} < \Delta$ respectively, involving a preassigned number $\Delta \equiv \delta^*(> 0)$. The prefixed number $\delta^*(> 0)$ is supposed to quantify an experimenter's minimum acceptable separation or the effect-size between the best treatment and the next best treatment. The point is this: One may not feel an urge to identify the best treatment when the next best treatment is within δ^*-closeness of the best treatment. In other words, when $\boldsymbol{\mu}$ came from IZ_{Orig}, one might rightfully feel indifferent or hesitant to declare definitively a winning treatment over the rest.

3. Formulation of (PZ_{New},IZ_{New}) Approach

A recent paper of Venkatesan and Mukhopadhyay (2025) [24] opened new directions in one specific sequential estimation problem for a multivariate normal mean vector. We take this opportunity to explore one of their main ideas and significantly expand its breadth of applications in classical selection problems. That will be apparent from the following logical constructs as we initially fix two preassigned numbers as input:

$$\delta^* > 0, \beta \in [0, 1), \text{ and let } \Delta \equiv \Delta(\delta^*, \beta, \sigma) = \delta^* \sigma^\beta. \tag{10}$$

This is the first instance where we have introduced $\Delta \equiv \Delta(\delta^*, \beta, \sigma)$. Obviously, Δ should have the same unit of measurement as that for the X-data. Our β is unit-free and it is held fixed. For example, if X-data is observed with the unit *seconds* (sec), then σ^β should have a unit $(\text{sec})^\beta$. That is, δ^* must have the unit $(\text{sec})^{1-\beta}$. On the other hand, if X-data happens to be midterm exam score, then X will have no associated unit tied with it. In that case, δ^* must be a positive real number that is unit-free. But, such considerations of associated units of measurement will not affect proposed data analyses or inference method.

Then, we incorporate a *new formulation*, abbreviated by (PZ_{New}, IZ_{New}), in the present selection problem:

$$\begin{aligned} &\textit{Preference-zone } PZ_{New} \equiv PZ_{New}(\delta^*, \beta, \sigma): \quad \Omega(\Delta) = \left\{ \boldsymbol{\mu} \in R^k : \mu_{[k]} - \mu_{[k-1]} \geq \Delta \right\}, \\ &\textit{Indifference-zone } IZ_{New} \equiv IZ_{New}(\delta^*, \beta, \sigma): \quad \Omega^c(\Delta) = R^k - \Omega(\Delta), \end{aligned} \tag{11}$$

with the effect-size Δ seen from (10). A "key idea" from Venkatesan and Mukhopadhyay (2025) [24] has led us to this choice of Δ that depends upon σ.

We should add that an appropriate choice of β may depend heavily on one's willingness and personal belief towards modeling a context in a specific statistical application in selection and ranking. The way (11) has laid out the preference-zone $PZ_{New}(\delta^*, \beta, \sigma)$, we are going to quantify a sense of how far ahead is the best treatment (with mean $\mu_{[k]}$) from the next best treatment (with mean $\mu_{[k-1]}$) by comparing:

$$\text{The effect-size } \mu_{[k]} - \mu_{[k-1]} \text{ with } \Delta = \delta^* \sigma^\beta \text{ defined via (10).}$$

A very important point to emphasize here is that the significance of $\mu_{[k]} - \mu_{[k-1]}$ is determined by comparing it with $\delta^* \sigma^\beta$. What it means is that we are willing to declare a treatment corresponding to $\mu_{[k]}$ as sufficiently worthwhile to be designated as a winner when $\mu_{[k]} - \mu_{[k-1]} \geq \Delta$. Otherwise, we record our indifferent attitude towards ranking these k treatments by not identifying any one from the list as the best one.

3.1. ($PZ_{Orig}(\delta^*)$, $IZ_{Orig}(\delta^*)$) *Approach Is Included in (10) and (11)*

Under this new formulation from (10) and (11), we assess how far ahead is the leading treatment associated with $\mu_{[k]}$ from the immediately following next treatment associated with $\mu_{[k-1]}$ by taking into account $\mu_{[k]}, \mu_{[k-1]}, \delta^*$ **as well as** how small or large σ might have been. In Bechhofer's (1954) [4] original ($PZ_{\mathbf{Orig}}$, $IZ_{\mathbf{Orig}}$) approach, one would have judged how far ahead is the leading treatment from the immediately following next treatment by taking into account **only** $\mu_{[k]}, \mu_{[k-1]}$ and δ^*, without paying any attention to how small or large σ was.

Conceptually, however, it makes sense to believe that a smaller (larger) effect size $\mu_{[k]} - \mu_{[k-1]}$ could imply significance when $\sigma(> 0)$ is perhaps relatively smaller (larger)! That is, Bechhofer's (1954) [4] original formulation, namely,

$$(PZ_{\mathbf{Orig}}(\delta^*), IZ_{\mathbf{Orig}}(\delta^*)) \text{ approach}$$

would readily coincide with our newly formulated

$$(PZ_{New}(\delta^*, \beta, \sigma), IZ_{New}(\delta^*, \beta, \sigma)) \text{ approach}$$

as a special sub-case when $\Delta(\delta^*, \beta, \sigma) \equiv \delta^*$ by fixing $\beta = 0$, thereby letting σ disappear completely from the set-up defined via (10) and (11). Under Bechhofer's (1954) [4] original ($PZ_{\mathbf{Orig}}(\delta^*)$, $IZ_{\mathbf{Orig}}(\delta^*)$) approach, a number of existing multi-stage selection and ranking methodologies may be reviewed from Mukhopadhyay and Solanky (1994 [20], chapter 3) and Mukhopadhyay and de Silva [23] (2009, chapter 14).

3.2. *Following Up on* $P_n(CS \mid Q_n)$

We recall that $\Delta_i = \mu_{[k]} - \mu_{[i]}$ with $i = 1, ..., k-1$ and $Z_{(jn)} \equiv n^{1/2}\left(\overline{X}_{(jn)} - \mu_{[j]}\right)/\sigma$ for $j = 1, ..., k$. These were incorporated in (6) and (7). Under $PZ_{New}(\delta^*, \beta, \sigma)$ formulation from (11), inside the preference-zone $\Omega(\Delta)$, we have $\Delta_i \geq \Delta(\delta^*, \beta, \sigma) = \delta^* \sigma^\beta$ for all $i = 1, ..., k-1$.

Next, under the selection rule Q_n from (5), we rewrite the probability of CS from (7) as follows:

$$\begin{aligned}\int_{-\infty}^{\infty}\left\{\Pi_{i=1}^{k-1}\Phi(z+n^{1/2}\Delta_i\sigma^{-1})\right\}\phi(z)dz &\geq \int_{-\infty}^{\infty}\left\{\Pi_{i=1}^{k-1}\Phi(z+n^{1/2}\delta^*\sigma^{\beta}\sigma^{-1})\right\}\phi(z)dz \\ &= \int_{-\infty}^{\infty}\left\{\Phi(z+n^{1/2}\delta^*\sigma^{\beta-1})\right\}^{k-1}\phi(z)dz,\end{aligned} \tag{12}$$

since $\Phi(.)$ is $\uparrow$ and $\Delta_i \geq \delta^*\sigma^{\beta}$ for all $\boldsymbol{\mu} \in \Omega(\Delta)$, $i = 1, ..., k-1$.

But, considering the *lower bound* of $P_n(CS \mid Q_n)$ for all $\boldsymbol{\mu} \in \Omega(\Delta)$ as expressed by (12), we also observe that it is attained pointwise when $\Delta_i = \delta^*\sigma^{\beta}$, $i = 1, ..., k-1$. Hence, we have the following conclusion:

$$\inf_{\boldsymbol{\mu}\in\Omega(\Delta)} P_n(CS \mid Q_n) = \int_{-\infty}^{\infty}\left\{\Phi(z+n^{1/2}\delta^*\sigma^{\beta-1})\right\}^{k-1}\phi(z)dz, \tag{13}$$

with the attainment of the stated infimum inside the preference-zone $\Omega(\Delta)$ if and only if $\boldsymbol{\mu}$ satisfies:

$$\begin{aligned}&\mu_{[1]} = ... = \mu_{[k-1]} = \mu_{[k]} - \delta^*\sigma^{\beta}\text{: designated as the } \textit{least favorable} \\ &\textit{configuration} \text{ (LFC) under } (\mathrm{PZ}_{\mathrm{New}}(\delta^*, \beta, \sigma), \mathrm{IZ}_{\mathrm{New}}(\delta^*, \beta, \sigma)).\end{aligned} \tag{14}$$

This LFC identifies a scenario where exactly $k-1$ treatments collapse at the location of the second best treatment whereas the second best treatment is exactly $\delta^*\sigma^{\beta}$-distance below the identified best treatment.

Now, considering the expression of the integral from (13), it is clear that we ought to determine the unique solution $h \equiv h_{k,P^*}(> 0)$ satisfying the following equation:

$$\int_{-\infty}^{\infty}\left\{\Phi(z+h_{k,P^*})\right\}^{k-1}\phi(z)dz = P^*, \tag{15}$$

where P^* is a fixed preset "high" number chosen from the interval $(k^{-1}, 1)$ by the practitioner involved. The customary expressions of $\phi(z), \Phi(z)$ may be recalled from (2). In practical applications, one may frequently prefer fixing $P^* = 0.90, 0.95$, or 0.99 as long as such chosen P^* sits inside the interval $(k^{-1}, 1)$.

3.3. Optimal Fixed Sample Size C

At this point, having fixed a suitable P^*, obviously the required minimum fixed sample size needs to be worked out. That leads us to determine the required optimal fixed sample size, had σ^2 been known:

$$\begin{aligned}&n = \text{the smallest (positive) integer such that } n^{1/2}\delta^*\sigma^{\beta-1} \geq h_{k,P^*}, \\ &\text{that is, the fixed sample size recorded from each treatment ought} \\ &\text{to be } n \geq h_{k,P^*}^2\sigma^{2(1-\beta)}/\delta^{*2} = C, \text{ say, where } C \equiv C(k, P^*, \delta^*, \beta, \sigma).\end{aligned} \tag{16}$$

We tacitly disregard that C may not be an integer.

However, having identified an expression $C(= n)$, if σ^2 were unknown, under $(\mathrm{PZ}_{\mathrm{New}}(\delta^*, \beta, \sigma), \mathrm{IZ}_{\mathrm{New}}(\delta^*, \beta, \sigma))$ approach, we would conclude:

$$\begin{aligned}&\text{(i) } P_n(CS \mid Q_n \text{ from (5)}) \geq P^* \ \forall \boldsymbol{\mu} \in \Omega(\Delta), \text{ and (ii) } \inf_{\boldsymbol{\mu}\in\Omega(\Delta)} P_n(CS \mid Q_n) = P^*, \\ &\text{while incorporating } \Omega(\Delta) \text{ from (11) with } \Delta \equiv \Delta(\delta^*, \beta, \sigma) = \delta^*\sigma^{\beta}.\end{aligned} \tag{17}$$

However, one realizes that the magnitude of C remains unknown since σ^2 is assumed unknown. We can easily improvise upon Dudewicz's (1971) [25] and Dantzig's (1940) [26] proofs to show that no fixed-sample-size selection methodology would satisfy the P^*-requirement stated in (17). Hence, purely sequential and accelerated sequential selection methodologies would provide essential ingredients in this field enriching theory, methods and practice. One may revisit Mukhopadhyay and Solanky (1994 [20] chapter 3), Mukhopadhyay and de Silva (2009 [23] chapter 14) or another suitable source for a review.

3.4. Why Restricting $\beta \in [0,1)$?

We postulated $\Delta \equiv \delta^*\sigma^\beta$ which led to an expression of $C(k, P^*, \delta^*, \beta, \sigma)$ in (16) involving $\sigma^{2(1-\beta)}$ in its numerator. Soon we would adapt suitably designed purely sequential and accelerated sequential selection methodologies by relying upon notions of f.o. and s.o. asymptotics as well as nonlinear renewal theory in sequential analysis (Woodroofe 1982 [27]; Ghosh et al. 1997 [28], chapter 2). That would require smooth handling of intricate boundary crossing problems when the expression of C involves a positive power of σ in the numerator. Thus, we included $0 < 1-\beta < 1$, or equivalently, $0 < \beta < 1$.

In Section 3.1, we explained how the new approach $(\mathrm{PZ}_{\mathrm{New}}(\delta^*, \beta, \sigma), \mathrm{IZ}_{\mathrm{New}}(\delta^*, \beta, \sigma))$ arising from (11) would reduce to Bechhofer's (1954) [4] original approach $(\mathrm{PZ}_{\mathbf{Orig}}(\delta^*), \mathrm{IZ}_{\mathbf{Orig}}(\delta^*))$ when $\beta = 0$. Since Bechhofer's fundamental approach has stood as a landmark among all methodologies in this field, we must embrace that benchmark. Hence, we included the choice $\beta = 0$, however, we do not necessarily recommend fixing $\beta = 0$. The new formulation has allowed us to argue in favor of a more practical approach by demanding that a notion of significance of an effect size should take into account the magnitudes of both $\mu_{[k]} - \mu_{[k-1]}$ and $\delta^*\sigma^\beta$.

We should highlight another point in this regard that is especially relevant: We may want to compare, in theory, the optimal fixed sample sizes required by the approach $(\mathrm{PZ}_{\mathrm{New}}(\delta^*, \beta, \sigma), \mathrm{IZ}_{\mathrm{New}}(\delta^*, \beta, \sigma))$ and by Bechhofer's (1954) [4] original approach $(\mathrm{PZ}_{\mathbf{Orig}}(\delta^*), \mathrm{IZ}_{\mathbf{Orig}}(\delta^*))$, respectively. Thus, we offer a requisite ratio of the associated sample sizes and propose considering:

$$\eta \equiv \eta(\beta, \sigma) = \frac{C(k, P^*, \delta^*, \beta, \sigma \mid \mathrm{New})}{C(k, P^*, \delta^*, 0, \sigma \mid \mathrm{Bechhofer})} = \frac{h^2_{k,P^*}\sigma^{2(1-\beta)}/\delta^{*2}}{h^2_{k,P^*}\sigma^2/\delta^{*2}} = \sigma^{-2\beta}. \tag{18}$$

Obviously, $\eta(\beta, \sigma) \equiv 1$ if (i) $\sigma = 1$ and $\beta \in [0,1)$, (ii) $\sigma \neq 1$ but $\beta = 0$. Also, $\eta(\beta, \sigma) > 1$ if $0 < \sigma < 1$ and $\beta \in (0,1)$ whereas $0 < \eta(\beta, \sigma) < 1$ if $\sigma > 1$ and $\beta \in (0,1)$. In other words, comparatively the new (or original) formulation would potentially lead to undersampling (or oversampling) when $\sigma > 1$ (or $\sigma < 1$) for all $\beta \in (0,1)$. These sentiments may be summarized as follows:

$$\begin{aligned} &\text{If } 0 < \sigma < 1: \quad \text{Oversampling } \%(\equiv \text{OS}\%) \quad = 100 \times \{\eta(\beta, \sigma) - 1\}\text{: Figure 2,} \\ &\text{If } 1 < \sigma < \infty: \quad \text{Undersampling } \%(\equiv \text{US}\%) \quad = 100 \times \{1 - \eta(\beta, \sigma)\}\text{: Figure 3,} \end{aligned} \tag{19}$$

with $\eta(\beta, \sigma)$ defined by (18). Figures 2 and 3 respectively illustrate behaviors of OS% and US% for four selected choices of β while considering reasonable spreads of σ values.

In order to offer additional help on fixing some possible β value in practice, we also show a range of numerical values of the OS% and US% corresponding to four selected choices of $\beta = \frac{1}{8}, \frac{1}{4}, \frac{1}{2}$ and $\frac{3}{4}$ with reasonable spreads of σ values on both sides of 1. Upon glancing over Figures 2 and 3 and the entries in Table 1, we may empirically suggest the following guidelines: β could be chosen approximately around $\frac{1}{8}$ to $\frac{1}{12}$ (or even smaller) so that (i) the OS% would not be too bothersome when $0.7 \leq \sigma < 1$, but (ii) the US% can be substantial if $\sigma > 1$. That way, a practical choice of appropriate β may not entirely depend

on the exact magnitude of σ. On the other hand, if a practitioner convincingly argues that $0 < \sigma < 0.7$, that is σ may be deemed very small, then one could fall back on Bechhofer's (1954) [4] classical methodology (with $\beta = 0$), as needed.

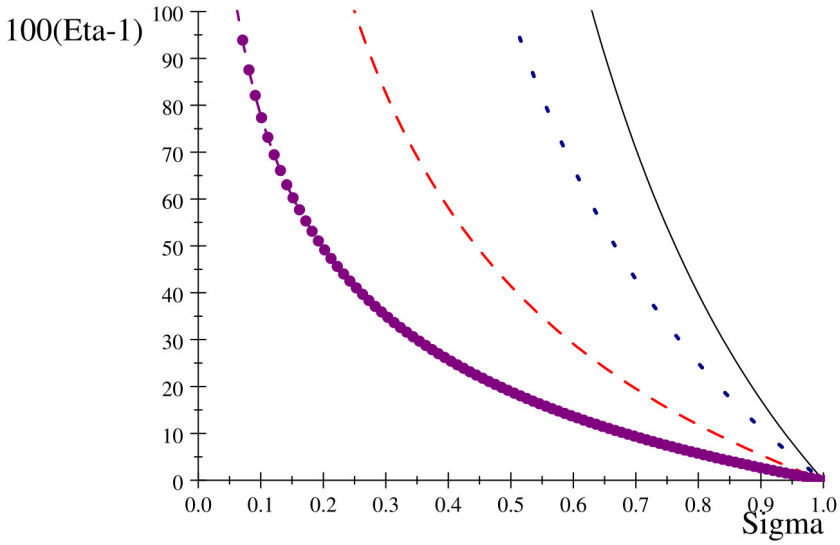


Figure 2. Behaviors of percentage oversampling from (19) due to the new approach compared with Bechhofer's classical approach for a selected set of reasonable values of $\sigma \in (0, 1]$. Going from bottom to top: the dot-dashed curve, dashed curve, dotted curve, and solid curve correspond to $\beta = \frac{1}{8}, \frac{1}{4}, \frac{1}{2}$ and $\frac{3}{4}$ respectively.

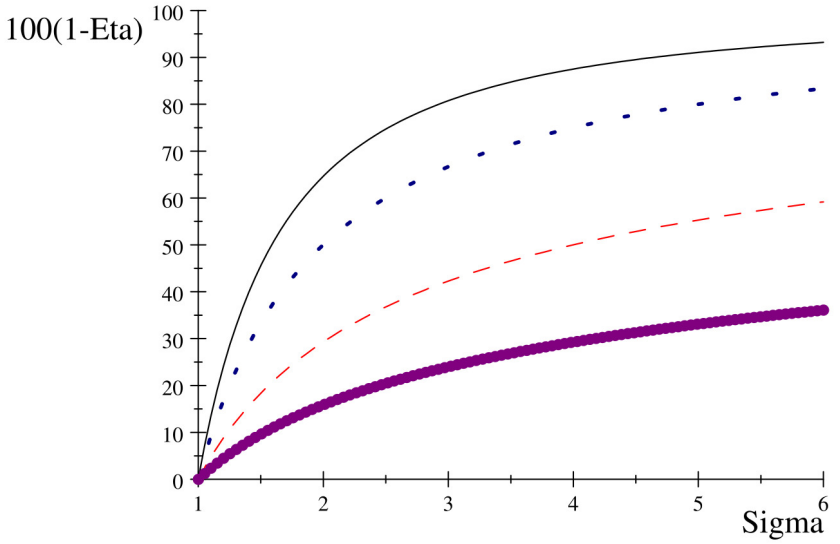


Figure 3. Behaviors of percentage undersampling [see (19)] due to the new approach compared with Bechhofer's classical approach for a selected set of reasonable values of $\sigma > 1$. Going from bottom to top: the dot-dashed curve, dashed curve, , dotted curve, and solid curve correspond to $\beta = \frac{1}{8}, \frac{1}{4}, \frac{1}{2}$ and $\frac{3}{4}$ respectively.

That said, in a large majority of applications, one may expect an unknown σ which substantially exceeds 1. If that sounds reasonable in a practical scenario, then β could be chosen approximately around $\frac{1}{8}$ to $\frac{1}{12}$ leading to substantial savings in the average sample sizes compared with Bechhofer's (1954) [4] classical formulation. These are simply suggestions or guidelines. One may discover more on this issue through simulations as well as real data analysis.

Table 1. Selected oversampling% $100 \times (\sigma^{-2\beta} - 1)$ or undersampling% $100 \times (1 - \sigma^{-2\beta})$ due to ($PZ_{New}(\delta^*, \beta, \sigma)$,$IZ_{New}(\delta^*, \beta, \sigma)$) when compared with ($PZ_{\mathbf{Orig}}(\delta^*)$,$IZ_{\mathbf{Orig}}(\delta^*)$) as $\sigma(>0) \uparrow 1$ or $\sigma(>1) \uparrow 5$ respectively in the context of (18).

As $\sigma(>0) \uparrow 1$, Percentage Oversampling: $100(\sigma^{-2\beta} - 1) \downarrow$					**As $\sigma(>1) \uparrow 5$, Percentage Undersampling $100(1 - \sigma^{-2\beta}) \uparrow$**				
	β					β			
σ	$\frac{3}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	σ	$\frac{3}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$
0.1	3062.3	900.0	216.2	77.8	1	0.00	00.00	0.00	0.00
0.3	508.6	233.3	82.6	35.1	2	64.65	50.00	29.29	15.91
0.5	182.8	100.0	41.4	18.9	3	80.76	66.67	42.27	24.02
0.7	70.8	42.9	19.5	9.3	4	87.50	75.00	50.00	29.29
0.9	17.1	11.1	5.4	2.7	5	91.06	80.00	55.28	33.14

3.5. Identifying the Requisite h_{k,P^*} Satisfying (15)

Having fixed k and P^*, the associated h_{k,P^*} values satisfying the requirement laid out by (15) were originally made available in tabular forms by Bechhofer (1954) [4], Gupta (1956,1963) [5,29], Gupta et al. (1973) [30], and Gibbons et al. (1977 [16], Table A.1, p. 400: Their notation "τ" coincides with our h_{k,P^*}). One may however prefer using Table A.1 from Gibbons et al. (1977) [16] perhaps because (i) its entries are mildly more accurate than those in other highlighted tables and (ii) it includes practical methods of interpolation to guide users when some input values k, P^* may not be available in those tables.

Alternatively, there are other possibilities to identify h_{k,P^*} values indirectly from some other available resources for a wide range of choices of k and P^*. In that vein, let us begin with $Z_1, ..., Z_k$ which are i.i.d. $N(0,1)$ random variables along with $U_i \equiv (Z_i - Z_k)/\sqrt{2}$, $i = 1, ..., k-1$. Suppose that we create a column vector $\boldsymbol{U}_{\overline{k-1} \times 1}$where $\boldsymbol{U}^{'}$ stands for the row vector $(U_1, ..., U_{k-1})$. It follows that

$$\boldsymbol{U} \sim N_{k-1}\left(\mathbf{0}, \Sigma_{\overline{k-1} \times \overline{k-1}}\right) \text{ with } \Sigma = \frac{1}{2} I_{\overline{k-1} \times \overline{k-1}} + \frac{1}{2} \mathbf{1}_{\overline{k-1} \times 1} \mathbf{1}^{'}_{1 \times \overline{k-1}} \text{ and } \mathbf{1}^{'} = (1, ..., 1), \tag{20}$$

since every linear function of $\boldsymbol{U}$ is distributed as a univariate normal random variable (Rao 1973 [31], p. 518). The random vector $\boldsymbol{U}$ is frequently described as a $(k-1)$-dimensional multivariate normal variable composed of standardized normal variables with equi-correlation coefficient $\frac{1}{2}$ between every pair U_i, U_j, for all $i \neq j = 1, ..., k-1$.

Now, we may alternatively view the earlier integral Equation (15) as follows:

$$\begin{aligned} P^* = \int_{-\infty}^{\infty} \{\Phi(z + h_{k,P^*})\}^{k-1} \phi(z) dz &= P\left\{Z_i < Z_k + h_{k,P^*}, i = 1, ..., k-1\right\} \\ &= P\left\{U_i < h_{k,P^*}/\sqrt{2}, i = 1, ..., k-1\right\}. \end{aligned} \tag{21}$$

Accurate upper 10% and 5% equi-percentage points for $\boldsymbol{U}$ can be found from Table A in Tong (1990 [32], p. 230) where one may begin by identifying Tong's (i) "n" with $k-1$, (ii) "γ" with P^*, and then his table entries of "c" values with our h_{k,P^*}.

In theory, (21) guides us to utilize Tong's (1990) [32] c-values, necessarily multiplied by $\sqrt{2}$, and that must provide Gibbons et al.'s (1977) [16] "τ" values successively matching with h_{k,P^*} values. Since (i) Gibbons et al.'s (1977) [16] Table A.1 or Tong's (1990) [32] Table A may not be readily available to many users, (ii) matching specifications from such different sources may not be exactly a trivial exercise, and (iii) publicly available software *R* may not be as freely available as one may believe considering many corners of the globe, we feel

compelled to highlight a short Table 2 comparing $\sqrt{2}c$ with h_{k,P^*} right here for immediate completeness. Table 2 speaks loudly for itself showing how closely the entries from this table's columns 2–3 stack up to those from its columns 7–8.

Table 2. Exploring the $\tau \equiv h_{k,P^*}$ values from Gibbons et al.'s (1977 [16], Table A.1) together with Tong's (1990 [32], Table A) "$c \equiv c_{n,\gamma}$" values.

	Source: Gibbons et al.'s (1977 [16], Table A.1): $\tau \equiv h = h_{k,P^*}$			**Source: Tong's (1990 [32], Table A): with $n = k-1$, $\gamma = P^*, \rho = 0.5$, $c \equiv c_{n,\gamma} = h_{k,P^*}/\sqrt{2}$**		**Do Columns 2–3 and 7–8 Look Useful? Compare h_{k,P^*}, $\sqrt{2}c_{n,\gamma}$ (20): $h_{k,P^*} = \sqrt{2}c_{n,\gamma}$**	
	$P^* = 0.90$	$P^* = 0.95$		$\gamma = 0.90$	$\gamma = 0.95$	$P^* = 0.90$	$P^* = 0.95$
k	h_{k,P^*}	h_{k,P^*}	n	$c_{n,\gamma}$	$c_{n,\gamma}$	$\sqrt{2}c_{n,\gamma}$	$\sqrt{2}c_{n,\gamma}$
2	1.8124	2.3262	1	1.2820	1.6449	1.8130	2.3262
3	2.2302	2.7101	2	1.5770	1.9163	2.2302	2.7101
5	2.5997	3.0552	4	1.8383	2.1603	2.5997	3.0551
10	2.9829	3.4182	9	2.1092	2.4170	2.9829	3.4182

3.6. Expression of $P(CS)$ Under Multi-Stage Sampling

Let N be an appropriate stopping time, an observable positive integer valued random variable, associated with a generic multi-stage sampling strategy. In the spirits of (4) and (5), we may describe the *terminal dataset* and the associated *decision rule* respectively as follows:

$$\text{Responses } \{X_{i1}, ..., X_{iN}\} \text{ arising from } \Pi_i, i = 1, ..., k, \text{ along with a selection rule } Q_N\text{: Declare that the } i^{\text{th}} \text{ treatment } \Pi_i \text{ is the } \textit{best} \text{ if and only if } \overline{X}_{iN} = \max_{1 \le j \le k} \overline{X}_{jN}. \tag{22}$$

We denote (**a**) an indicator variable $I(A)$ associated with a random event A which takes the value 1 (or 0) on the set A (or A^c) respectively, (**b**) $\boldsymbol{Y}_n \equiv (\overline{X}_{1n}, ..., \overline{X}_{kn})$, $\boldsymbol{T}_{in} \equiv (X_{i2} - X_{i1}, ..., X_{in} - X_{i1})$, $i = 1, ..., k$, and $\boldsymbol{T}_n \equiv (\boldsymbol{T}_{1n}, ..., \boldsymbol{T}_{kn})$ on the set (or event) $[N = n]$ for $n \ge 2$, (**c**) $\boldsymbol{\Delta} \equiv (\Delta_1, ..., \Delta_{k-1})$ with $\Delta_j \ge \delta^* \sigma^\beta$ for all $j = 1, ..., k-1$ when $\boldsymbol{\mu} \in \Omega(\Delta)$, the preference-zone, and (**d**) $\boldsymbol{U} \perp \boldsymbol{V}$ to highlight that the two random variables $\boldsymbol{U}, \boldsymbol{V}$ are independent of each other.

Theorem 1. *Let N be a stopping time under a generic multi-stage sampling strategy satisfying the following sufficient conditions:*

(i) $P\{N < \infty\} = 1$;

(ii) $I(N = n)$ *depends only on the statistic $\boldsymbol{T}_n$ obtained from n recorded responses arising from the i^{th} treatment $\Pi_i, i = 1, ..., k$, for all fixed $n \ge 2$.*

Then, under the selection rule Q_N in (22) and the formulation $(PZ_{New}(\delta^*, \beta, \sigma), IZ_{New}(\delta^*, \beta, \sigma))$ from (10) and (11), for every fixed σ, the following conclusion holds:

$$\inf_{\boldsymbol{\mu} \in \Omega(\Delta)} P(CS \mid Q_N) = E_\sigma \left[\int_{-\infty}^{\infty} \left\{ \Phi(z + N^{1/2} \delta^* \sigma^{-(1-\beta)}) \right\}^{k-1} \phi(z) dz \right]. \tag{23}$$

Additionally, the infimum in (23) is attained inside the preference-zone $\Omega(\Delta)$ if and only if we work under the configuration:

$$\mu_{[1]} = ... = \mu_{[k-1]} = \mu_{[k]} - \delta^* \sigma^\beta, \textit{ referred to as the LFC.} \tag{24}$$

Proof. Let $m(\geq 2)$ be a pilot sample size that culminates in a generic fashion to a terminal sample size N from each Π_i, $i = 1, ..., k$. Assumption (i) guarantees that the terminal data $\{X_{i1}, ..., X_{iN}\}$ from $\Pi_i, i = 1, ..., k$, and $\boldsymbol{W}_N \equiv (\overline{X}_{1N}, ..., \overline{X}_{kN})$, a vector of k sample means, are all genuinely meaningful (i.e., measurable).

Now, we begin by holding $\sigma = \sigma_0(> 0)$ fixed. Then, for any fixed $n(\geq m)$, $\boldsymbol{T}_n$ is a location invariant statistic for $\boldsymbol{\mu}$, that is $\boldsymbol{T}_n$ is *ancillary* for the parameter vector $\boldsymbol{\mu}$. Additionally, the statistic $\boldsymbol{Y}_n$ is complete and sufficient for $\boldsymbol{\mu}$. Then, Basu's (1955) [33] theorem would apply to let us conclude immediately that $\boldsymbol{Y}_n \perp \boldsymbol{T}_n$ for every fixed $n(\geq m)$ when $\sigma = \sigma_0$. The Assumption (ii) will guarantee that two statistics $\boldsymbol{Y}_n$ and $I(N = n)$ must be independent (i.e., $\boldsymbol{Y}_n \perp I(N = n)$) for each fixed $n(\geq m)$, for all $\boldsymbol{\mu}$ and *any* fixed σ_0. So, the two statistics $(\overline{X}_{1n}, ..., \overline{X}_{kn})$ and $I(N = n)$ must be independent for every fixed, but otherwise arbitrary, $n(\geq m)$, $\forall \boldsymbol{\mu}$ and $\forall \sigma^2$.

Hence, for all fixed $\boldsymbol{\mu} \in R^k$ and fixed $\sigma \in R^+$, under the selection rule Q_N from (22), we can express the probability of correct selection as follows:

$$\begin{aligned} & P_{\boldsymbol{\mu},\sigma}\{CS\} \\ & = \Sigma_{n=m}^{\infty} P_n\left\{\overline{X}_{(kn)} > \overline{X}_{(in)}, i = 1, ..., k-1 \middle| N = n\right\} P_\sigma\{N = n\} \\ & = \Sigma_{n=m}^{\infty} P_n\left\{\overline{X}_{(kn)} > \overline{X}_{(in)}, i = 1, ..., k-1\right\} P_\sigma\{N = n\}, \\ & \qquad \because \boldsymbol{Y}_n \equiv (\overline{X}_{1n}, ..., \overline{X}_{kn}) \perp I(N = n),\ \forall n \geq m \\ & = \Sigma_{n=m}^{\infty} a_n(\boldsymbol{\mu}, \sigma) P_\sigma\{N = n\}, \text{ say,} \end{aligned} \tag{25}$$

where we have denoted:

$$a_n(\boldsymbol{\mu}, \sigma) = \int_{-\infty}^{\infty} \left\{\Pi_{i=1}^{k-1} \Phi(z + n^{1/2}\Delta_i \sigma^{-1})\right\} \phi(z) dz \text{ with } \Delta_i = \mu_{[k]} - \mu_{[i]}, \tag{26}$$

utilizing (12).

Now, in the spirits of (12) and (13), for each fixed $n \geq m$, we must have:

$$\inf_{\boldsymbol{\mu} \in \Omega(\Delta)} a_n(\boldsymbol{\mu}, \sigma) = \int_{-\infty}^{\infty} \left\{\Phi(z + n^{1/2}\delta^* \sigma^{\beta-1})\right\}^{k-1} \phi(z) dz. \tag{27}$$

We may emphasize that the $\inf_{\boldsymbol{\mu} \in \Omega(\Delta)}$ operator in (27) would be attained under the same relevant LFC stated in (24) for every fixed $n \geq m$, in view of (14). That is, the LFC does not change with changing n and this remains true for every fixed $n \geq m$.

Now, the proof of (23) and (24) is complete. □

Remark 1. *We may reiterate by emphasizing that Assumption (ii) implies that the probability distribution of N may involve* ***only*** *the parameter σ^2, but certainly not $\boldsymbol{\mu}$, for all fixed $n \geq 2$. That is, under the assumptions stated in Theorem 1, an expression of $P_{\boldsymbol{\mu},\sigma}(N = n)$ could (or would) involve the scale parameter σ^2, and of course known β, but not be affected by the location parameter vector $\boldsymbol{\mu}(\in R^k)$ at all.*

4. A Purely Sequential Selection Strategy

The magnitude of the optimal fixed-sample-size $C(k, P^*, \delta^*, 0, \sigma)$ from (16) had σ been known, that is assuming $\beta = 0$, remains unknown and no fixed-sample-size method would work (Dudewicz 1971) [25]. Hence, specialized purely sequential, accelerated sequential, three-stage, and other multi-stage selection methodologies were vigorously developed under the classical IZ formulation ($\text{PZ}_{\textbf{Orig}}(\delta^*)$,$\text{IZ}_{\textbf{Orig}}(\delta^*)$). Bechhofer et al.

(1954a) [6] developed a classical two-stage selection methodology in the spirits of Stein (1945, 1949) [34,35].

Robbins et al. (1968) [14] and Bechhofer et al. (1968) [13] proposed purely sequential selection rules when $\beta = 0$. The accelerated sequential and three-stage selection rules subsequently came along, followed by other methodologies enriching this field since mid 1980s by laying down the groundwork for numerous asymptotic f.o. as well as s.o. approximations. One may refer to Mukhopadhyay and Solanky (1994 [20], chapter 3).

In our quest for accurate estimation of the optimal fixed-sample-size $C(k, P^*, \delta^*, \beta, \sigma)$ from (16), assuming $0 \leq \beta < 1$, we must incorporate appropriate randomized schemes of essential sampling strategies. Section 3.6 developed an expression of $P(CS)$ and identified the associated LFC (Theorem 1) under a generic multi-stage sampling scheme satisfying a broad set of sufficient conditions (i)–(ii). Now, we direct our focus on introducing an explicitly laid out purely sequential selection rule along with a range of desirable asymptotic considerations.

We begin with pilot observations $\{X_{ij}; j = 1, ..., m(\geq 2)\}$ from Π_i of size m which give rise to $\overline{X}_{im}, S^2_{im}$ according to (1), $i = 1, ..., k$, followed by the pooled sample variance, U^2_m. The pilot data would be followed by recording one additional observation at-a-time from each Π_i according to the stopping rule:

$$\begin{gathered} N_{\text{PurSeq}} \equiv N_{\delta^*} = N = \inf\left\{ n \geq m : n \geq h^2_{k,P^*} \delta^{*-2} U^{2(1-\beta)}_n \right\}, \delta^* > 0, 0 \leq \beta < 1, \\ \text{along with the terminal selection rule } Q_N \text{ from (22) by declaring } \Pi_i \\ \text{as the best treatment if and only if } \overline{X}_{iN} = \max_{1 \leq j \leq k} \overline{X}_{jN}. \end{gathered} \tag{28}$$

The motivation behind setting up our stopping time N_{PurSeq} as in (28) may be explained as follows: Having observed the pilot dataset $\{X_{ij}; j = 1, ..., m(\geq 2)\}, i = 1, ..., k$, we estimate $C(k, P^*, \delta^*, \beta, \sigma)$ with $\widehat{C}_m = h^2_{k,P^*} \delta^{*-2} U^{2(1-\beta)}_m$. Now, if $m \geq \widehat{C}_m$ is observed, then sampling will terminate right here, and then (28) would give us the final sample size $N_{\text{PurSeq}} = m$ from each Π_i.

But, if $m < \widehat{C}_m$ is observed, then sampling would continue by recording one new data point from each Π_i leading to the augmented dataset $\{X_{ij}; j = 1, ..., m, m+1\}, i = 1, ..., k$, while reestimating $C(k, P^*, \delta^*, \beta, \sigma)$ from this updated data with $\widehat{C}_{m+1} = h^2_{k,P^*} \delta^{*-2} U^{2(1-\beta)}_{m+1}$. Now, if $m + 1 \geq \widehat{C}_{m+1}$ is observed, then sampling will terminate here, and then (28) would give us the final sample size $N_{\text{PurSeq}} = m + 1$ from each Π_i.

However, if $m + 1 < \widehat{C}_{m+1}$ is observed, then sampling would continue by recording one new data point from each Π_i leading to the further updated dataset $\{X_{ij}; j = 1, ..., m, m+1, m+2\}, i = 1, ..., k$, while reestimating $C(k, P^*, \delta^*, \beta, \sigma)$ from this data with $\widehat{C}_{m+2} = h^2_{k,P^*} \delta^{*-2} U^{2(1-\beta)}_{m+2}$. This process would continue up to and including an integer $n(\geq m)$ when n exceeds the corresponding estimate $\widehat{C}_n$ of C for the *first time*.

We would use the notation $N_{\text{PurSeq}}, N_{\delta^*}, N$ interchangeably.

The stated sufficient conditions from Theorem 1 clearly hold under this purely sequential methodology. At termination, we would have available the final updated dataset:

$$\{m, N_{\text{PurSeq}}, X_{i1}, ..., X_{im}, ..., X_{iN_{\text{PurSeq}}} \text{ from } \Pi_i, i = 1, ..., k\}, \tag{29}$$

which would then be implemented with the customary selection rule Q_N. That is, we would declare Π_i as the best treatment if and only if its sample mean is the largest one considering competing sample means arising from *all other* $\Pi_j, j \neq i$.

4.1. Asymptotic First-Order (F.O.) Properties

We may add that all asymptotic f.o. properties described in this subsection holds when the pilot size $m \geq 2$. Under the purely sequential sampling strategy (28), we can obviously claim that $N_{\delta_1^*} \geq N_{\delta_2^*}$ w.p.1 for any fixed $0 < \delta_1^* < \delta_2^*$. That is, N_{PurSeq} is non-decreasing as a function of δ^* which allows us to claim that $\lim_{\delta^* \to 0} N_{\delta^*} = \infty$ w.p.1 as we hold $\boldsymbol{\mu}, \sigma^2, \beta, P^*$ fixed. One may also look over the classic paper of Chow and Robbins (1965) [36].

We should emphasize that in a practical problem, $\delta^*(> 0)$ is also held fixed, however we may conceptualize $\delta^* \to 0$ to help us move towards the asymptotics. In other words, we may treat δ^* converging to 0 purely as a mathematical device to make the sample size N_{PurSeq} approach ∞. Equivalently, we could have argued in favor of analogous asymptotics in the limit as $\sigma \to \infty$ or $P^* \to 1$ or $C \to \infty$. We just prefer to continue presenting the requisite asymptotics as $\delta^* \to 0$.

4.1.1. Behavior of $C^{-1}N_{\text{PurSeq}}$

We hold $\boldsymbol{\mu}, \sigma^2, \beta, P^*$ fixed and then we have as $\delta^* \to 0$:

$$\sigma^{-2}U^2_{N_{\delta^*}} \overset{P_{\boldsymbol{\mu},\sigma}}{\to} 1, \sigma^{-2}U^2_{N_{\delta^*}-1} \overset{P_{\boldsymbol{\mu},\sigma}}{\to} 1, \text{and } mC^{-1} \to 0. \tag{30}$$

Next, we may express the following *basic inequality* (w.p.1) directly from the boundary stopping criterion stated in (28) as:

$$h^2_{k,P^*}\delta^{*-2}U^{2(1-\beta)}_{N_{\delta^*}} \leq N_{\delta^*} \leq m + h^2_{k,P^*}\delta^{*-2}U^{2(1-\beta)}_{N_{\delta^*}-1}I(N_d > m), \tag{31}$$

and then divide all sides of (31) by C coming from (16) to claim (w.p.1):

$$\sigma^{-2(1-\beta)}U^{2(1-\beta)}_{N_{\delta^*}} \leq N_{\delta^*}C^{-1} \leq mC^{-1} + \sigma^{-2(1-\beta)}U^{2(1-\beta)}_{N_{\delta^*}-1}. \tag{32}$$

In view of (30) and (32) combined, we see clearly that $\lim_{\delta^* \to 0} N_{\delta^*}C^{-1}$ exists, in probability, and then,

$$\lim_{\delta^* \to 0} N_{\delta^*}C^{-1} = 1 \text{ in probability.} \tag{33}$$

In other words, for small δ^*, the terminal sample size N_{δ^*} will stay in a close proximity of C with high probability.

4.1.2. Behavior of the Moments of $C^{-1}N_{\text{PurSeq}}$

Again we hold $\boldsymbol{\mu}, \sigma^2, \beta, P^*$ fixed and then consider small enough δ^* such that $mC^{-1} \leq 1$. In what follows, we successively express a number of updated upper bounds (w.p.1) for the right-hand side of the basic inequality seen from (31):

$$\begin{aligned} &NC^{-1} \\ &\leq 1 + \sigma^{-2(1-\beta)}U^{2(1-\beta)}_{N-1}I(N > m) \\ &\leq 1 + \sigma^{-2(1-\beta)}(N-2)^{-(1-\beta)}I(N > m)\left\{\Sigma_{i=1}^k \Sigma_{j=1}^{N-1}(X_{ij} - \overline{X}_{iN-1})^2\right\}^{1-\beta} \\ &\leq 1 + \sigma^{-2(1-\beta)}\left[\left\{N(N-2)^{-1}\right\}^{1-\beta}I(N > m)\right]\left\{\Sigma_{i=1}^k N^{-1}\Sigma_{j=1}^N (X_{ij} - \mu_i)^2\right\}^{1-\beta}. \end{aligned} \tag{34}$$

Ghosh and Mukhopadhyay (1976) [37] included analogous helpful bounds in some sequential estimation problems.

Next, we may denote:

$$L_i \equiv \sup_{n \geq 2}\left\{n^{-1}\Sigma_{j=1}^n (X_{ij} - \mu_i)^2\right\}, i = 1, ..., k, \tag{35}$$

and also observe that $N(N-2)^{-1}I(N>m) \leq 3$ holds w.p.1 since $m \geq 2$. Then, from the last step in (34) combined with (35), we obtain:

$$NC^{-1} \leq 1 + 3^{1-\beta}\sigma^{-2(1-\beta)}\left\{\Sigma_{i=1}^{k} L_i\right\}^{1-\beta} \text{ w.p.1.} \tag{36}$$

By the ergodic theorem of Wiener (1939) [38], the expected values of all positive powers of L_i must be finite, $i = 1, ..., k$, since positive powers of $|X|$'s are all finite. That is, the positive powers of NC^{-1} must all be *uniformly integrable* (u.i.). Next, we incorporate the Lebesgue dominated convergence theorem in (36) to conclude:

$$\lim_{\delta^* \to 0} E_{\boldsymbol{\mu},\sigma}\left[\left\{NC^{-1}\right\}^{\omega}\right] = E_{\boldsymbol{\mu},\sigma}\left[\lim_{\delta^* \to 0}\left\{NC^{-1}\right\}^{\omega}\right] = 1 \text{ for all fixed } \omega(>0). \tag{37}$$

The conclusion from (37) corresponding to $t = 1$ is referred to as the *asymptotic f.o. efficiency property* in the sense of Ghosh and Mukhopadhyay (1981) [39]. In practice, what this means is that the terminal sample size N_{δ^*} will stay in a close proximity of $C \equiv C_{\delta^*}$ in mean (L_1-norm) when δ^* (or C_{δ^*}) is small (or large).

Remark 2. *The uniform integrability argument incorporated in (36) and (37) also shows that $E_{\boldsymbol{\mu},\sigma}[N^{\omega}]$ is finite for any fixed $\omega > 0$ when $\boldsymbol{\mu}, \sigma^2, \beta, P^*$, and δ^* are held fixed.*

4.1.3. Behavior of $P(CS \mid Q_N)$

Going back to (23) and (24), under the LFC $\mu_{[1]} = ... = \mu_{[k-1]} = \mu_{[k]} - \delta^*\sigma^{\beta}$, we may rewrite the main conclusion from Theorem 1 as follows:

$$\begin{gathered} P(CS \mid Q_N) = E_{\sigma}\left[F(h_{k,P^*}^2 N/C)\right] \text{ where we define the function:} \\ F(y) = \int_{-\infty}^{\infty}\left\{\Phi(z + y^{1/2})\right\}^{k-1}\phi(z)dz, y > 0. \end{gathered} \tag{38}$$

Clearly, for all $y > 0$, we can claim:

$$F(y) = |F(y)| \leq \int_{-\infty}^{\infty}\left|\Phi(z + y^{1/2})\right|^{k-1}|\phi(z)|dz \leq \int_{-\infty}^{\infty}\phi(z)dz = 1,$$

that is $F(y), y > 0$ is a bounded function.

Then, we appeal to (38) and apply the Lebesgue dominated convergence theorem to obtain:

$$\lim_{\delta^* \to 0} P(CS \mid Q_N) = E_{\sigma}\left[F(h_{k,P^*}^2 \lim_{\delta^* \to 0}\{N/C\})\right] = F\left(h_{k,P^*}^2\right) = P^*, \tag{39}$$

on account of the explicit determination of h_{k,P^*} as shown in (15). The conclusion (39) is often described as *asymptotic consistency property* in the sense of Chow and Robbins (1965) [36] or *asymptotic f.o. consistency property* in the sense of Mukhopadhyay and Solanky (1994 [20], chapter 3, p. 81).

4.2. Asymptotic Second-Order (S.O.) Properties

In order to fully appreciate a majority of asymptotic s.o. approximations under the purely sequential stopping time, we rely upon original nonlinear renewal theoretic analysis from Woodroofe (1977) [40] and Lai and Siegmund (1977,1979) [41,42]. The associated technicalities may also be reviewed from Woodroofe (1982) [27], Siegmund (1985) [43], Mukhopadhyay and Solanky (1994 [20], Section 2.4.2), Ghosh et al. (1997 [28], Section 2.9), and Mukhopadhyay and de Silva (2009 [23], Section A.4, Appendix A).

First, we may rewrite the stopping variable N_{PurSeq} from (28) *equivalently* as $G+1$ w.p.1 by denoting:

$$\begin{gathered} G \equiv \inf\Big\{ n \geq m-1 : kC^{-1/(1-\beta)} n^{(2-\beta)(1-\beta)^{-1}} L(n) \geq \Sigma_{i=1}^{n} W_i \Big\} \text{ where} \\ L(n) \equiv (1+n^{-1})^{1/(1-\beta)} = 1 + L_0 n^{-1} + o(n^{-1}) \text{ with } L_0 = (1-\beta)^{-1} \\ \text{and additionally, } W_1, W_2, \ldots \overset{\text{i.i.d.}}{\sim} \chi_k^2 \text{ random variables.} \end{gathered} \tag{40}$$

A brief justification of this one-to-one mapping from N to G is included in Section 4.2.4. That said, all positive moments of W are clearly finite.

4.2.1. Elements of Sequential Nonlinear Renewal Theory

First, we line-up G from (40) with the initial nonlinear renewal theoretic representation from Woodroofe (1977,1982) [27,40], emphasized later in Mukhopadhyay and de Silva (1994 [20], (A.4.1), p. 447). Thereby, we readily identify the following entities from their equations (A.4.2)–(A.4.5):

$$\begin{gathered} \delta = (2-\beta)(1-\beta)^{-1} (> 1), h^* = kC^{-1/(1-\beta)}, \theta = E[W_1] = k, \tau^2 = V[W_1] = 2k, \\ \beta^* = (\delta - 1)^{-1} = 1 - \beta, n_0^* = (\theta h^{*-1})^{\beta^*} = C, q = \beta^{*2} \tau^2 \theta^{-2} = k^{-1}(1-\beta)^2, \\ m_0 = m - 1, \text{ and also } r = \tfrac{1}{2}k. \end{gathered} \tag{41}$$

The last bit in (41), that is $r = \frac{1}{2}k$, is not hard to check. For all fixed $u > 0$, we can get an upper bound for the distribution function of W_1 as follows:

$$\begin{aligned} & P\{W_1 \leq u\} \\ &= \{2^{k/2}\Gamma(\tfrac{1}{2}k)\}^{-1} \int_0^u \exp(-w/2) w^{(k-2)/2} dw \\ &\leq \{2^{k/2}\Gamma(\tfrac{1}{2}k)\}^{-1} \int_0^u w^{(k-2)/2} dw, \because 0 < \exp(-w/2) < 1, w \in (0,u) \\ &= Bu^r, \text{with } B = \{2^{k/2}\Gamma(\tfrac{1}{2}k+1)\}^{-1} \text{ and } r = \tfrac{1}{2}k. \end{aligned}$$

Obviously, $B(>0)$ and $r(>0)$ do not involve $u(>0)$. We may remark that B does not play a key role, but r does in determining the sufficient conditions on m for the results subsequently shown in (43) parts (i)–(iv).

Next, we utilize a number of related bits and pieces from Woodroofe (1977, 1982) [27,40] or (A.4.4)–(A.4.5) in Mukhopadhyay and de Silva (2009 [23], p. 447) to express:

$$\begin{aligned} &\text{(i)} \quad D \equiv D(k,\beta) &&= \textstyle\sum_{n=1}^{\infty} n^{-1} E\Big[\max\Big\{0, \chi_{nk}^2 - \frac{(2-\beta)}{(1-\beta)} nk\Big\}\Big]; \\ &\text{(ii)} \quad \nu \equiv \nu(k,\beta) &&= \Big[\tfrac{k}{2(1-\beta)} + (1-\beta)\Big] - D(k,\beta); \\ &\text{(iii)} \quad \kappa \equiv \kappa(k,\beta) &&= \tfrac{(1-\beta)}{k}(\nu - 2 + \beta) - 1. \end{aligned} \tag{42}$$

Numerical evaluation of $D \equiv D(k,\beta)$ would follow along the lines of Mukhopadhyay and Zhang (2023 [44], Section 5.1). For selected choices of k and β, one could identify $D(k,\beta)$ values from Mukhopadhyay and Zhang (2023, 2024) [44,45] and also from Venkatesan and Mukhopadhyay (2025 [24], Table 1).

4.2.2. Key Asymptotics on Moments of N_{PurSeq}

We begin by summarizing a set of fundamental s.o. asymptotic (as $\delta^* \to 0$) results in connection with G coming from (40):

$$\begin{aligned}
&\text{(i)} \quad E_{\mu,\sigma}[G] = C + \kappa + o(1) \text{ if } m > 1 + 2(1-\beta)k^{-1};\\
&\text{(ii)} \quad G^* \equiv \frac{G - C}{C^{1/2}} \xrightarrow{\mathcal{L}} N\left(0,\ 2k^{-1}(1-\beta)^2\right) \quad \text{if } m > 1;\\
&\text{(iii)} \quad P_{\mu,\sigma}\{G \le \varepsilon C\} = O\left(C^{-(m-1)k/(2-2\beta)}\right),\ 0 < \varepsilon < 1 \text{ fixed, if } m > 1; \text{ and}\\
&\text{(iv)} \quad |G^*|^{\omega} \text{ is } \textit{uniformly integrable} \text{ (u.i.), } \omega > 0 \text{ fixed, if } m > 1 + (1-\beta)k^{-1}\omega;
\end{aligned} \tag{43}$$

quoted from Woodroofe (1977, 1982) [27,40]. We also refer to Theorems A.4.1–A.4.2 stated in Mukhopadhyay and de Silva (2009 [23], pp. 447–448). One may readily check from the original sources that all required sufficient conditions hold.

One may note that the asymptotic (in $\mathcal{L}$) normality result for the standardized stopping time G^* quoted in (43) part (ii) would follow independently by invoking Ghosh and Mukhopadhyay's (1975) [46] theorem. One may refer to the well-known Ghosh-Mukhopadhyay theorem from Mukhopadhyay and Solanky (1994 [20], Theorem 2.4.1, pp. 41–43) plus other essential details.

Since we know that $N_{\text{PurSeq}} = G + 1$ w.p.1, (43) part (i) immediately leads to the following sharper (s.o.) conclusion as $\delta^* \to 0$ [39]:

$$\begin{aligned}
&E_{\mu,\sigma}\left[N_{\text{PurSeq}} - C\right] = E_{\mu,\sigma}[G - C] + 1 = (1-\beta)k^{-1}(\nu - 2 + \beta) + o(1) \text{ if}\\
&m > 1 + 2(1-\beta)k^{-1} \text{ which is referred to as } \textit{asymptotic s.o. efficiency}\\
&\text{property (Ghosh and Mukhopadhyay 1981).}
\end{aligned} \tag{44}$$

Recall the *asymptotic f.o. efficiency property of* N_{PurSeq} from (37) which meant that

$$\lim_{\delta^* \to 0} E_{\mu,\sigma}\left[C^{-1}N_{\text{PurSeq}}\right] = 1.$$

But, can $E_{\mu,\sigma}\left[N_{\text{PurSeq}}\right]$ be far away from C in the limit? The concept of asymptotic s.o. efficiency property of N_{PurSeq} described by (44) settles that issue. Indeed $E_{\mu,\sigma}\left[N_{\text{PurSeq}} - C\right]$ will hang very tightly around a fixed real number $\kappa + 1$ or equivalently

$$(1-\beta)k^{-1}(\nu - 2 + \beta) \text{ when } \delta^* \to 0.$$

Obviously, asymptotic s.o. efficiency property would imply asymptotic f.o. efficiency property, but its converse does not necessarily hold. Ghosh and Mukhopadhyay (1981) [39] had settled that fundamental issue.

Now, we apply the remaining s.o. properties from (43) parts (ii)–(iv) to claim the following results when $\delta^* \to 0$:

$$\begin{aligned}
&\text{(i)} \quad N^* \equiv \frac{N_{\text{PurSeq}} - C}{C^{1/2}} \xrightarrow{\mathcal{L}} N\left(0,\ 2k^{-1}(1-\beta)^2\right) \quad \text{if } m > 1;\\
&\text{(ii)} \quad P_{\mu,\sigma}\left\{N_{\text{PurSeq}} \le \varepsilon C\right\} = O\left(C^{-(m-1)k/(2-2\beta)}\right),\ 0 < \varepsilon < 1 \text{ fixed, if } m > 1;\\
&\text{(iii)} \quad |N^*|^2 \text{ is u.i. if } m > 1 + 2(1-\beta)k^{-1}.
\end{aligned} \tag{45}$$

Outlining Proofs for (45) Parts (i)–(iii)

Proof. For clarity and completeness, we selectively record some of the key arguments only.

In part (i), we express (w.p.1):

$$N^* \equiv \frac{G - C + 1}{C^{1/2}} = \frac{G - C}{C^{1/2}} + C^{-1/2} = G^* + C^{-1/2}, \tag{46}$$

and then appeal to (43) part (ii) combined with Slutsky's Theorem (since $C^{-1/2} \to 0$).

In part (ii), one notes right away:

$$P_{\boldsymbol{\mu},\sigma}\{N_{\text{PurSeq}} \leq \varepsilon C\} = P_{\boldsymbol{\mu},\sigma}\{G+1 \leq \varepsilon C\} \leq P_{\boldsymbol{\mu},\sigma}\{G \leq \varepsilon C\},$$

and then appeal to (43) part (iii). Incidentally, part (ii) can also be verified directly in the spirit of Lemma 7.2.1 from Ghosh et al. (1997 [28], Equation (7.2.19), pp. 215–216).

In part (iii), for small $\delta^*(<\delta_0)$ with $\delta_0 > 0$ such that $C > 1$, we incorporate (46) as follows:

$$|N^*|^2 \leq \left\{\left|\frac{G-C}{C^{1/2}}\right| + C^{-1/2}\right\}^2 = \Sigma_{\omega=0}^2 \binom{2}{\omega} |G^*|^\omega C^{-(2-\omega)/2} \leq \Sigma_{\omega=0}^2 \binom{2}{\omega} |G^*|^\omega, \quad (47)$$

and then apply (43) part (iv) successively when $\omega = 1, 2$. Obviously, $|G^*|$ is u.i. if $|G^*|^2$ is u.i., and hence the sufficient condition on m would correspond to the case when $\omega = 2$. Now, (45) part (iii) follows. □

4.2.3. Approximation of $P(CS \mid Q_N)$ Under the LFC

Under the LFC, namely $\mu_{[1]} = ... = \mu_{[k-1]} = \mu_{[k]} - \delta^*\sigma^\beta$, we return to (38) which expressed $P(CS)$ as $E_\sigma\left[F(h_{k,P^*}^2 N_{\text{PurSeq}}/C)\right]$. Let us denote the first two derivatives of $F(y), y > 0$ as:

$$F^{(1)}(y) = \frac{\partial}{\partial y}F(y),\ F^{(2)}(y) = \frac{\partial}{\partial y}F^{(1)}(y) = \frac{\partial^2}{\partial y^2}F(y),\ y > 0,$$

and then begin by expanding $F(y)$ around $y = h_{k,P^*}^2$ in a Taylor series as follows:

$$\begin{gathered} F(y) = F(h_{k,P^*}^2) + (y - h_{k,P^*}^2)F^{(1)}(h_{k,P^*}^2) + \tfrac{1}{2}(y - h_{k,P^*}^2)^2 F^{(2)}(u^*) \text{ involving an} \\ \text{appropriate intermediate number } u^* \text{ between } y \text{ and } h_{k,P^*}^2, y > 0. \end{gathered} \quad (48)$$

But $F(h_{k,P^*}^2) = P^*$ by the specific choice of h_{k,P^*} that was made explicit in (15).

Next, we recall the standardized stopping variable N^* from (45) part (i) and then, (48) would immediately lead to the following expression involving an appropriately chosen intermediate random variable $U^* \equiv U_{\delta^*}^*$ between $h_{k,P^*}^2 N_{\text{PurSeq}}/C$ and h_{k,P^*}^2:

$$\begin{gathered} E_\sigma\left[F(h_{k,P^*}^2 N_{\text{PurSeq}}/C)\right] = P^* + h_{k,P^*}^2 C^{-1}\left\{F^{(1)}(h_{k,P^*}^2)A + \tfrac{1}{2}h_{k,P^*}^2 B\right\} \text{ where} \\ \text{we denote: } A \equiv E_\sigma[N - C] \text{ and } B \equiv E_\sigma\left[N^{*2}F^{(2)}(U_{\delta^*}^*)\right]. \end{gathered} \quad (49)$$

From (44), we reiterate that

$$A \equiv E_\sigma[N - C] = (1-\beta)k^{-1}(\nu - 2 + \beta) + o(1) \text{ if } m > 1 + 2(1-\beta)k^{-1}. \quad (50)$$

Now, we make another explicit **claim**:

$$\begin{gathered} B \equiv E_\sigma\left[N^{*2}F^{(2)}(U_{\delta^*}^*)\right] = 2k^{-1}(1-\beta)^2 F^{(2)}(h_{k,P^*}^2) + o(1) \\ \text{if } m > 1 + 5(1-\beta)k^{-1}, \end{gathered} \quad (51)$$

which is proved shortly in Section Outlining a Proof of (51).

For all fixed $\boldsymbol{\mu}, \sigma^2, \beta$ and P^*, we combine (50) and (51) with (49) to obtain the following asymptotic s.o. expansion of $P(CS \mid Q_N)$ under the LFC as $\delta^* \to 0$:

$$E_\sigma\Big[F(h^2_{k,P^*}N_{\text{PurSeq}}/C)\Big] = P^* + (1-\beta)k^{-1}h^2_{k,P^*}C^{-1}\Big\{(\nu-2+\beta)F^{(1)}(h^2_{k,P^*}) + h^2_{k,P^*}(1-\beta)F^{(2)}(h^2_{k,P^*})\Big\} + o(C^{-1}) \text{ if } m > 1+5(1-\beta)k^{-1}. \tag{52}$$

Obviously, the conclusion stated in (52) will be totally justified upon verifying the claim (51).

Outlining a Proof of (51)

Proof. Let us selectively supply some of the key arguments only. In view of (45) parts (i), (iii) and (33), a validation of (51) will follow upon showing that $N^{*2}F^{(2)}(U^*_{\delta^*})$ is u.i. for small enough δ^* under some appropriate condition on m. We begin by noting that for all $y > 0$, we easily verify:

$$\Big|F^{(2)}(y)\Big| \le a_1 y^{-1/2} + a_2 y^{-1} + a_3 y^{-3/2} \text{ with appropriate fixed positive real numbers: } a_1, a_2 \text{ and } a_3, y > 0. \tag{53}$$

Next, we proceed to handle the term $B \equiv E_\sigma\Big[N^{*2}F^{(2)}(U^*_{\delta^*})\Big]$ and express (w.p.1) $N^{*2}F^{(2)}(U^*_{\delta^*})$ as:

$$N^{*2}F^{(2)}(U^*_{\delta^*})I(N > \tfrac{1}{2}C) + N^{*2}F^{(2)}(U^*_{\delta^*})I(N \le \tfrac{1}{2}C) = B_1 + B_2, \text{ say.}$$

Now, on the set $[NC^{-1} > \frac{1}{2}]$, the intermediate random variable $U^*_{\delta^*}$ between $h^2_{k,P^*}N_{\text{PurSeq}}/C(> \frac{1}{2}h^2_{k,P^*})$ and h^2_{k,P^*} must satisfy $U^*_{\delta^*} > \frac{1}{2}h^2_{k,P^*}$ so that we obtain (w.p.1):

$$|B_1| = \Big|N^{*2}F^{(2)}(U^*_{\delta^*})I(N > \tfrac{1}{2}C)\Big| \le bN^{*2} \tag{54}$$

where we may denote: $b = a_1(2h^{-2})^{1/2} + a_2(2h^{-2}) + a_3(2h^{-2})^{3/2}$.

Also, (45) part (iii) and (54) combined would imply:

$$B_1 \equiv N^{*2}F^{(2)}(U^*_{\delta^*})I(N > \tfrac{1}{2}C) \text{ is u.i. if } m > 1+2(1-\beta)k^{-1}. \tag{55}$$

Obviously, $I(N > \frac{1}{2}C)$ converges in probability to 1 in view of (33). But, additionally B_1 converges to $2k^{-1}(1-\beta)^2F^{(2)}(h^2_{k,P^*})\chi^2_1$ in distribution in view of (45) part (i). Thus, we can conclude:

$$E_\sigma\Big[B_1\Big] = 2k^{-1}(1-\beta)^2F^{(2)}(h^2_{k,P^*}) + o(1) \text{ if } m > 1+2(1-\beta)k^{-1}. \tag{56}$$

Now, on the set $[NC^{-1} \le \frac{1}{2}]$, the intermediate random variable $U^*_{\delta^*}$ between $h^2_{k,P^*}N_{\text{PurSeq}}/C(\le \frac{1}{2}h^2_{k,P^*})$ and h^2_{k,P^*} must satisfy $U^*_{\delta^*} > h^2_{k,P^*}N_{\text{PurSeq}}/C$ so that we obtain:

$$\begin{aligned}
&\Big|E_\sigma[B_2]\Big| \\
&= E_\sigma\Big[N^{*2}\Big|F^{(2)}(U^*_{\delta^*})\Big|I(N \le \tfrac{1}{2}C)\Big] \\
&\le h^2_{k,P^*}\Big\{a_1(m/C)^{-1/2} + a_2(m/C)^{-1} + a_3(m/C)^{-3/2}\Big\}E_\sigma\Big[N^{*2}I(N \le \tfrac{1}{2}C)\Big] \\
&\le h^2_{k,P^*}O(C^{3/2})O(C)P_\sigma\Big[N \le \tfrac{1}{2}C\Big], \text{since } N^{*2} \le \tfrac{5}{4}C \text{ when } N \le \tfrac{1}{2}C \\
&= O(C^{5/2})O\Big(C^{-(m-1)k/(2-2\beta)}\Big), \text{in view of (45) part (ii)} \\
&= o(1),
\end{aligned} \tag{57}$$

if $m > 1+5k^{-1}(1-\beta)$. In other words, we have:

$$E_\sigma\Big[B_2\Big] = o(1) \text{ if } m > 1+5(1-\beta)k^{-1}. \tag{58}$$

Now, a combination of (56) and (58) completes the verification of (51). □

4.2.4. Verification of the Mapping from (28) to (40)

Proof. We begin to rewrite N_{PurSeq} from (28) as:

$$\begin{aligned} &\inf\left\{n \geq m : n \geq h_{k,P^*}^2 \delta^{*-2} U_n^{2(1-\beta)}\right\} \\ &= \inf\left\{n \geq m : \left(n\delta^{*2} h_{k,P^*}^{-2}\right)^{1/(1-\beta)} \geq U_n^2\right\} \\ &= \inf\left\{n \geq m : k\left(n\delta^{*2} h_{k,P^*}^{-2}\right)^{1/(1-\beta)} \geq \Sigma_{i=1}^k S_{in}^2\right\}, \end{aligned} \tag{59}$$

where S_{in}^2, defined by (1), was the sample variance from n observations on the i^{th} treatment, $i = 1, ..., k$.

Then, we can employ Helmert's (1876) [47] classic orthogonal transformation (see also, Mukhopadhyay 2000 [22], pp. 197–201; (4.4.9)) to obtain a set of i.i.d. $N(0, \sigma^2)$ random variables $Y_{i1}, Y_{i2}, ..., Y_{in-1}$ arising from the i^{th} treatment so that have:

$$\begin{aligned} &(n-1)S_{in}^2\sigma^{-2} = \Sigma_{j=1}^{n-1} Y_{ij}^2, i = 1, ..., k \Rightarrow (n-1)\sigma^{-2}\Sigma_{i=1}^k S_{in}^2 = \Sigma_{j=1}^{n-1} W_i, \text{ by} \\ &\text{rearranging the } Y\text{'s and denoting } W_i = \Sigma_{r=1}^k Y_{ri}^2, i = 1, ..., n-1. \end{aligned} \tag{60}$$

Clearly, $\{Y_{ij}^2; i = 1, ..., k; j = 1, ..., n-1\}$ are i.i.d. random variables with a common χ_1^2 distribution, implying that $W_1, ..., W_{n-1}$ are i.i.d. random variables with a common χ_k^2 distribution. Now, combining (59) and (60), N is equivalently expressed as:

$$\begin{aligned} N = \inf\left\{n \geq m : k(C)^{-1/(1-\beta)} n^{1/(1-\beta)}(n-1) \geq \Sigma_{j=1}^{n-1} W_i\right\}, \\ \text{where } W_i\text{'s are } \overset{\text{i.i.d.}}{\sim} \chi_k^2, \end{aligned} \tag{61}$$

since $\left(\delta^{*2} h_{k,P^*}^{-2}\right)^{1/(1-\beta)} = \sigma^2 C^{-1/(1-\beta)}$.

We substitute $n' = n - 1$, thereby (61) leads to:

$$\begin{aligned} &G \\ &= \inf\left\{n' \geq m - 1 : kC^{-1/(1-\beta)}(n'+1)^{1/(1-\beta)} n' \geq \Sigma_{j=1}^{n'} W_i\right\} \\ &= \inf\left\{n \geq m - 1 : kC^{-1/(1-\beta)}(n+1)^{1/(1-\beta)} n \geq \Sigma_{j=1}^{n} W_i\right\} \\ &= \inf\left\{n \geq m - 1 : kC^{-1/(1-\beta)} n^{(2-\beta)/(1-\beta))}(1+n^{-1})^{1/(1-\beta)} \geq \Sigma_{j=1}^{n} W_i\right\} \end{aligned} \tag{62}$$

with $L(n) = (1+n^{-1})^{1/(1-\beta)} = 1 + (1-\beta)^{-1}n^{-1} + o(n^{-1})$. Clearly, $n' = n - 1$ so that (62) implies $G = N - 1$, that is, $N = G + 1$ w.p.1, verifying (40). □

5. An Accelerated Sequential Selection Strategy

The core intents originally laid down by Hall (1983) [48] were developed further in a unified fashion by Mukhopadhyay and Solanky (1994) [20]. The improved approach of acceleration that we would embrace now was developed by Mukhopadhyay (1996) [49]. More recently, however, Mukhopadhyay and Zhang (2023, 2024) [44,45] have taken analogs of acceleration to newer heights delivering asymptotic *s.o. expansions* rather that asymptotic *s.o. bounds*.

In the spirits of how we initialized a purely sequential rule (28), we begin an accelerated sequential rule with pilot observations $\{X_{ij}; j = 1, ..., m(\geq 2)\}$ from Π_i, followed by recording one additional observation at-a-time from each Π_i according to a stopping rule:

$$R_{\text{PurSeq}} \equiv R_{\delta^*} = R = \inf\Big\{n \geq m : n \geq q^{-1}h^2_{k,P^*}\delta^{*-2}U_n^{2(1-\beta)}\Big\}, \delta^* > 0, \\ 0 \leq \beta < 1, \text{ where } q(\geq 2) \text{ is a fixed and preset integer.} \tag{63}$$

Intuitively, the stopping variable R_{PurSeq} from (63) would estimate

$$C_0 \equiv q^{-1}C(k, P^*, \delta^*, \beta, \sigma) = q^{-1}C, \text{ say, a preset fraction of } C, \tag{64}$$

purely sequentially. In other words, we proceed purely sequentially in the first part of gathering data. After R_{PurSeq} comes to a stop drawing observations, we define:

$$N_{\text{AccSeq}} \equiv N = qR_{\text{PurSeq}}. \tag{65}$$

We gather additional new observations $\Big\{X_{iR_{\text{AccSeq}}+1}, ..., X_{iN_{\text{AccSeq}}}\Big\}$ of size $N_{\text{AccSeq}} - R_{\text{PurSeq}} \equiv q(R-1)$, that is we record $(q-1)R_{\text{PurSeq}}$ new observations from each Π_i in a *single batch independently* of the random variables:

$$\Big\{X_{i1}, ..., X_{im}, X_{im+1}, ..., X_{iR_{\text{PurSeq}}} \text{ from } \Pi_i\Big\}, i = 1, ..., k.$$

In applications one customarily picks $q = 2$, that is one would run the first half of a sampling process purely sequentially, followed by a single batch of sampling.

This accelerated methodology would save approximately $100(1 - q^{-1})\%$ sampling operations leading to operational simplicity with substantial cost saving by gaining significant logistical convenience. Upon terminating the combined sampling strategy of initial part consisting of purely sequential sampling per (63) followed by batch sampling per (64), one will have available the full terminal dataset:

$$\{m, R_{\text{PurSeq}}, N_{\text{AccSeq}}, X_{i1}, ..., X_{im}, X_{im+1}, ..., X_{iR_{\text{PurSeq}}}, X_{iR_{\text{AccSeq}}+1}, \\ ..., X_{iN_{\text{AccSeq}}} \text{ from } \Pi_i, i = 1, ..., k\}, \text{ of size } N_{\text{AccSeq}}(\equiv N), \tag{66}$$

which would then be accompanied with the selection rule Q_N. That is, we would declare Π_i as the best treatment if and only if $\overline{X}_{iN} = \max_{1 \leq j \leq k} \overline{X}_{jN}$.

The asymptotic f.o. properties of this accelerated sequential selection strategy would be the same as those summarized via (30), (33), (37), and (39) seen from Section 4.1. Next, the asymptotic s.o. analysis from Section 4.2 would remain valid by rewriting (w.p.1) R_{PurSeq} as $G + 1$ with G defined just like that in (40) by replacing C with $C_0 \equiv q^{-1}C$. The expressions from (42) would also remain valid with obvious minor updating as needed.

That said, the previous conclusions from (44) and (45) will remain valid respectively as follows (as $\delta^* \to 0$) with essential updating:

$$E_{\boldsymbol{\mu},\sigma}\big[R_{\text{PurSeq}} - C_0\big] = (1-\beta)k^{-1}(\nu - 2 + \beta) + o(1) \text{ if } m > 1 + 2(1-\beta)k^{-1}, \tag{67}$$

and

$$\text{(i)} \quad R^* \equiv \frac{R_{\text{PurSeq}} - C_0}{C_0^{1/2}} \xrightarrow{\mathcal{L}} N\big(0,\ 2k^{-1}(1-\beta)^2\big) \quad \text{if } m > 1; \\ \text{(ii)} \quad P_{\boldsymbol{\mu},\sigma}\big\{R_{\text{PurSeq}} \leq \varepsilon C_0\big\} = O\Big(C_0^{-(m-1)k/(2-2\beta)}\Big) \text{ with } 0 < \varepsilon < 1 \text{ fixed, if } m > 1; \\ \text{(iii)} \quad |R^*|^b \text{ is uniformly integrable (u.i.), } b > 0, \text{ if } m > 1 + (1-\beta)k^{-1}b. \tag{68}$$

Since $N_{\text{AccSeq}} \equiv qR_{\text{PurSeq}}$, we immediately obtain as $\delta^* \to 0$:

$$\begin{aligned} E_{\boldsymbol{\mu},\sigma}\left[N_{\text{AccSeq}} - C\right] &= E_{\boldsymbol{\mu},\sigma}\left[qR_{\text{PurSeq}} - qC_0\right] = qE_{\boldsymbol{\mu},\sigma}\left[R_{\text{PurSeq}} - C_0\right] \\ &= q(1-\beta)k^{-1}(\nu - 2 + \beta) + o(1) \text{ if } m > 1 + 2(1-\beta)k^{-1}. \end{aligned} \tag{69}$$

Equation (69) proves asymptotic s.o. efficiency property of the accelerated sequential strategy. Comparing (69) with (44), we readily see that the accelerated sequential strategy provides a more biased sample size. However, this extra bias is customarily *not very large* because the term $(1-\beta)k^{-1}(\nu - 2 + \beta)$ is often rather small. Comparing Tables obtained from simulations, one should feel assured that any such bias is so small that one may not even realize whether or not this may be a serious issue.

Next, we briefly verify first:

$$N^* \equiv \frac{N_{\text{AccSeq}} - C}{C^{1/2}} = \frac{qR_{\text{PurSeq}} - qC_0}{q^{1/2}C_0^{1/2}} = q^{1/2}R^* \xrightarrow{\mathcal{L}} N\Big(0,\ 2qk^{-1}(1-\beta)^2\Big) \quad \text{if } m > 1, \tag{70}$$

which parallels (45) part (i). We note right away that the asymptotic variance of N_{AccSeq} from (70) exceeds that of N_{PurSeq} from (45). But, that goes hand-in-hand with our earlier observation of having a slightly increased bias mentioned after (69) in estimation of C.

With fixed $\varepsilon \in (0,1)$, we can also claim:

$$P_{\boldsymbol{\mu},\sigma}\{N_{\text{AccSeq}} \le \varepsilon C\} = P_{\boldsymbol{\mu},\sigma}\{R_{\text{PurSeq}} \le \varepsilon C_0\} = O(C^{-(m-1)p/(2-2\beta)}), \tag{71}$$

which parallels (45) part (ii). Again, since $N^* \equiv q^{1/2}R^*$, it follows immediately that N^{*2} would be u.i. when $m > 1 + 2(1-\beta)k^{-1}$.

Having in place all essential f.o. as well as s.o. asymptotic ingredients developed in (63)–(71), we can fairly easily provide asymptotic s.o. approximations for $P(CS \mid Q_N)$ under the LFC in the spirit of (52). Obviously, that would require only some minor adjustments in view of mild changes seen in (69) and (70) compared with similar claims proved previously in (44) and (45). For brevity, we leave out further details.

6. Summaries from Simulations

Now that we have developed a number of crucial theoretical properties for two sampling strategies, we are in a position to implement the selection methodologies via simulations with 5000($\equiv B$) independent replications incorporating each configuration.

6.1. Fixing Initial Input

We hope to gauge how the two strategies may perform when the optimal sample sizes are small (e.g., $C < 100$), moderate (e.g., $100 \le C < 200$) or large (e.g., $C \ge 200$). Under both strategies, pseudo random samples of same size were generated independently from $k = 5$ populations, as needed, with Π_i standing for a $N(2, 25)$ distribution, $i = 1, 2, 3, 4$, and Π_5 corresponding to a $N(2 + \delta^*, 25)$ where we fixed the effect-size $\delta^* = 0.6, 1.0, 1.3, 1.5$ successively.

That is, the reported simulations were all run under the LFC with Π_5 being the *best* population. We also fixed in advance $P^* = 0.95$ which led to $h_{5,0.95} \equiv 3.0552$ (from Table 2). The optimal sample size C came from (16) after fixing $\beta = 0, \frac{1}{8}, \frac{1}{12}$. The next display briefly shows the combinations of (β, δ^*) leading to a range of C-values under our consideration:

	δ^*			
	1.5	1.3	1.0	0.6
$\beta = 0$	103.71	138.08	233.36	648.21
$\beta = \frac{1}{12}$	79.31	105.59	178.45	433.49
$\beta = \frac{1}{8}$	69.36	92.34	156.05	495.70

(72)

We show C-values up to 2-decimal places in (72). In a real application, an optimal fixed sample size may be interpreted as the smallest integer $\geq C$. In survey sampling methods, that is a long-standing custom.

6.2. Running Simulations

An executable computer program *Seq14.exe* from Mukhopadhyay and de Silva's (2009 [23], chapter 14, pp. 384–392) text helped us to run the required simulations when $\beta = 0$ corresponding to the purely sequential methodology under Bechhofer's (1954) [4] original approach ($\mathrm{PZ}_{\mathbf{Orig}}(\delta^*)$,$\mathrm{IZ}_{\mathbf{Orig}}(\delta^*)$). We draw attention to some basic notations that are formally explained in Table 3.

Table 3. Basic notations and parameter configurations leading to the reported summaries in Tables 4 and 5 under a fixed number of replications $(B \equiv 5000)$ across all simulations.

$n_b, b = 1, ..., B$:	sample size in the b^{th} run;
$\overline{n} = B^{-1}\Sigma_{b=1}^{B} n_b$:	estimate of C or $E[N]$;
$s_{\overline{n}} = \left\{ \left(B^2 - B\right)^{-1} \Sigma_{b=1}^{B} (n_b - \overline{n})^2 \right\}^{1/2}$:	standard error (s.e.) of $\overline{n}$;
p_b:	indicator (0 or 1) of correctly identifying $\mu_{[k]}$ in the b^{th} run;
$\overline{p} = B^{-1}\Sigma_{b=1}^{B} p_b$:	estimate of P^* or $P(CS)$;
$s_{\overline{p}} = \{\overline{p}(1-\overline{p})/B\}^{1/2}$:	s.e. of $\overline{p}$;
$J_1 = [(\overline{n} - C) \pm 2s_{\overline{n}}]$:	two standard deviation (≈95%) confidence interval for $E[N - C]$
$J_2 = [\overline{p} \pm 2s_{\overline{p}}]$:	two standard deviation (≈95%) confidence interval for P^* or $P(CS)$
A_1:	indicator (0 or 1), 1 validating the s.o. efficiency from simulations;
A_2:	indicator (0 or 1), 1 validating that P^* or $P(CS)$ is included in J_2;
Blocks (ii), (iii):	estimated % savings in the ave sample size when $\beta > 0$ case is compared with $\beta = 0$;
$\overline{n}_{(j)}$:	ave sample size (col 8) from block j under each fixed δ^*, expressed generically without introducing further sub-labels, $j = i, ii, iii$.

Parameter Configuration: The LFC. We generated $5(= k)$ independent normal populations as follows: Let Π_i represent a $N(2, \sigma^2)$ population, $i = 1, 2, 3, 4$, and Π_5 stand for a $N(2 + \delta^*, \sigma^2)$ population with the same $\sigma(= 5)$ in all of them. In other words, Π_5 stands for the *best* treatment. That is, all summaries reported from simulations in Tables 4 and 5 were generated under the *least favorable configuration* (LFC, defined via (14) and (24)) with a preset *effect size*, $\delta^* = 0.6, 1.0, 1.3, 1.5$, incorporated successively.

Table 4. A summary from simulations of the purely sequential approach when $k = 5$ and $P^* = 0.95$: $m = 10, h_{5,0.95} \equiv 3.0552, \sigma = 5, J_1 = [(\overline{n} - C) \pm 2s_{\overline{n}}]$, $J_2 = [\overline{p} \pm 2s_{\overline{p}}]$ and indicator variables $A_1 = I(\kappa + 1 \in J_1), A_2 = I(P^* \in J_2)$.

δ^*	C	$\overline{n}$ $s_{\overline{n}}$	$\overline{n} - C$ $\overline{n}/C$	$\overline{p}$ $s_{\overline{p}}$	J_1 J_2	A_1 A_2	$100\left[1 - \left(\overline{n}_{(ii)}/\overline{n}_{(i)}\right)\right]$% or $100\left[1 - \left(\overline{n}_{(iii)}/\overline{n}_{(i)}\right)\right]$% as % Savings
Block (i). Stopping rule (28): Bechhofer's original formulation (PZ_{Orig}, IZ_{Orig}): $\beta = 0$ and $\kappa + 1 \equiv 0.249$ from (73)							
1.5	103.71	103.78 0.091075	0.07 1.0007	0.9514 0.0030410	$[-0.11215, 0.25215]$ $[0.94532, 0.95748]$	1 1	N/A
			$\Delta \equiv \delta^*, n_{\min} = 83, n_{\max} = 130$				
1.3	138.08	138.12 0.10578	0.04 1.0003	0.9504 0.0030705	$[-0.17156, 0.25156]$ $[0.94426, 0.95654]$	1 1	N/A
			$\Delta \equiv \delta^*, n_{\min} = 111, n_{\max} = 163$				
1.0	233.36	233.45 0.13718	0.09 1.0004	0.9488 0.0031170	$[-0.18436, 0.36436]$ $[0.94257, 0.95503]$	1 1	N/A
			$\Delta \equiv \delta^*, n_{\min} = 197, n_{\max} = 272$				
0.6	648.21	648.31 0.22670	0.10 1.0002	0.9491 0.0031083	$[-0.11156, 0.31156]$ $[0.94288, 0.95532]$	1 1	N/A
			$\Delta \equiv \delta^*, n_{\min} = 588, n_{\max} = 712$				
Block (ii). Stopping rule (28): new formulation (PZ_{New}, IZ_{New}): $\beta = \frac{1}{8}$ and $\kappa + 1 \equiv 0.293$ from (73)							
1.5	69.36	69.62 0.076085	0.262 1.0038	0.9482 0.0031342	$[0.10983, 0.41417]$ $[0.94193, 0.95447]$	1 1	$100\left[1 - \left(\overline{n}_{(ii)}/\overline{n}_{(i)}\right)\right]$% 32.916%
			$\Delta \equiv \delta^* \sigma^\beta = 1.8343, n_{\min} = 50, n_{\max} = 87$				
1.3	92.34	92.68 0.087681	0.340 1.0037	0.9516 0.003035	$[0.16464, 0.51536]$ $[0.94553, 0.95767]$	1 1	32.899%
			$\Delta \equiv \delta^* \sigma^\beta = 1.5897, n_{\min} = 71, n_{\max} = 114$				
1.0	156.05	156.44 0.11300	0.390 1.0025	0.9533 0.0029839	$[0.164, 0.616]$ $[0.94733, 0.95927]$	1 1	32.988%
			$\Delta \equiv \delta^* \sigma^\beta = 1.2228, n_{\min} = 125, n_{\max} = 188$				
0.6	433.49	434.07 0.18625	0.580 1.0013	0.9558 0.0029068	$[-1.4226, 2.5826]$ $[0.94999, 0.96161]$	1 1	33.046%
			$\Delta \equiv \delta^* \sigma^\beta = 0.73371, n_{\min} = 383, n_{\max} = 483$				
Block (iii). Stopping rule (28): under new formulation (PZ_{New}, IZ_{New}): $\beta = \frac{1}{12}$ and $\kappa + 1 \equiv 0.279$ from (73)							
1.5	79.31	79.40 0.093818	0.087 1.0011	0.9487 0.0031199	$[-0.10064, 0.27464]$ $[0.94246, 0.95494]$	0 1	$100\left[1 - \left(\overline{n}_{(iii)}/\overline{n}_{(i)}\right)\right]$% 23.492%
			$\Delta \equiv \delta^* \sigma^\beta = 1.7153, n_{\min} = 57, n_{\max} = 98$				
1.3	105.59	105.85 0.093197	0.260 1.0025	0.9523 0.0030141	$[0.073606, 0.44639]$ $[0.94627, 0.95833]$	1 1	23.364%
			$\Delta \equiv \delta^* \sigma^\beta = 1.4866, n_{\min} = 82, n_{\max} = 134$				
1.0	178.45	178.61 0.12021	0.16 1.0009	0.9534 0.0029809	$[-0.08042, 0.40042]$ $[0.94744, 0.95936]$	1 1	23.491%
			$\Delta \equiv \delta^* \sigma^\beta = 1.1435, n_{\min} = 148, n_{\max} = 208$				
0.6	495.70	496.14 0.20011	0.44 1.0009	0.9473 0.0031598	$[0.03978, 0.84022]$ $[0.94098, 0.95362]$	1 1	23.472%
			$\Delta \equiv \delta^* \sigma^\beta = 0.68612, n_{\min} = 447, n_{\max} = 549$				

Table 5. A summary from simulations of the accelerated sequential approach when $k = 5, q = 2$ and $P^* = 0.95$: $m = 10, h_{5,0.95} \equiv 3.0552, \sigma = 5, J_1 = [(\overline{n} - C) \pm 2s_{\overline{n}}]$, $J_2 = [\overline{p} \pm 2s_{\overline{p}}]$ and indicator variables $A_1 = I(2\kappa + 2 \in J_1), A_2 = I(P^* \in J_2)$.

δ^*	C	$\overline{n}$ $s_{\overline{n}}$	$\overline{n} - C$ $\overline{n}/C$	$\overline{p}$ $s_{\overline{p}}$	J_1 J_2	A_1 A_2	$100\left[1 - \left(\overline{n}_{(ii)}/\overline{n}_{(i)}\right)\right]\%$ or $100\left[1 - \left(\overline{n}_{(iii)}/\overline{n}_{(i)}\right)\right]\%$ as % Savings
Block (i). Stopping rule (63)–(65): original formulation (PZ_{Orig}, IZ_{Orig}): $\beta = 0$ and $2\kappa + 2 \equiv 0.498$ from (73)							
1.5	103.71	103.94 0.14001	0.23 1.0022	0.9512 0.0030469	$[-0.05002, 0.51002]$ $[0.94511, 0.95729]$	1 1	N/A
			$\Delta \equiv \delta^*, n_{\min} = 72, n_{\max} = 140$				
1.3	138.08	138.46 0.15076	0.38 1.0028	0.9493 0.0031026	$[0.07848, 0.68152]$ $[0.94309, 0.95551]$	1 1	N/A
			$\Delta \equiv \delta^*, n_{\min} = 94, n_{\max} = 174$				
1.0	233.36	234.14 0.19771	0.78 1.0004	0.9488 0.003117	$[0.38458, 1.1754]$ $[0.94257, 0.95503]$	1 1	N/A
			$\Delta \equiv \delta^*, n_{\min} = 180, r_{\max} = 290$				
0.6	648.21	648.68 0.32583	0.47 1.0007	0.9508 0.0030587	$[-0.18166, 1.1217]$ $[0.94468, 0.95692]$	1 1	N/A
			$\Delta \equiv \delta^*, n_{\min} = 558, n_{\max} = 726$				
Block (ii). Stopping rule (63)–(65): new formulation (PZ_{New}, IZ_{New}): $\beta = \frac{1}{8}$ and $2\kappa + 2 \equiv 0.587$ from (73)							
1.5	69.36	69.68 0.10720	0.32 1.0046	0.9482 0.0030939	$[0.1056, 0.5344]$ $[0.94341, 0.95579]$	0 1	$100\left[1 - \left(\overline{n}_{(ii)}/\overline{n}_{(i)}\right)\right]\%$ 32.961%
			$\Delta \equiv \delta^* \sigma^\beta = 1.8343, n_{\min} = 38, n_{\max} = 98$				
1.3	92.34	92.82 0.12615	0.48 1.0052	0.9527 0.0030021	$[0.2277, 0.7323]$ $[0.94670, 0.9587]$	1 1	32.798%
			$\Delta \equiv \delta^* \sigma^\beta = 1.5897, n_{\min} = 58, n_{\max} = 124$				
1.0	156.05	156.70 0.16094	0.65 1.0042	0.9547 0.002941	$[0.32812, 0.97188]$ $[0.94882, 0.96058]$	1 1	33.074%
			$\Delta \equiv \delta^* \sigma^\beta = 1.2228, n_{\min} = 114, n_{\max} = 190$				
0.6	433.49	433.82 0.26361	0.33 1.0008	0.9558 0.0029068	$[-0.19722, 0.85722]$ $[0.94999, 0.96161]$	1 1	33.123%
			$\Delta \equiv \delta^* \sigma^\beta = 0.73371, n_{\min} = 354, n_{\max} = 506$				
Block (iii). Stopping rule (63)–(65): under new formulation (PZ_{New}, IZ_{New}): $\beta = \frac{1}{12}$ and $2\kappa + 2 \equiv 0.558$ from (73)							
1.5	79.31	79.86 0.11823	0.55 1.0069	0.9497 0.0030910	$[0.31354, 0.78646]$ $[0.94352, 0.95588]$	1 1	$100\left[1 - \left(\overline{n}_{(iii)}/\overline{n}_{(i)}\right)\right]\%$ 23.167%
			$\Delta \equiv \delta^* \sigma^\beta = 1.7153, n_{\min} = 48, n_{\max} = 106$				
1.3	105.59	106.06 0.13435	0.47 1.0045	0.9557 0.0029099	$[0.2013, 0.7387]$ $[0.94988, 0.96152]$	1 1	23.400%
			$\Delta \equiv \delta^* \sigma^\beta = 1.4866, n_{\min} = 74, n_{\max} = 140$				
1.0	178.45	179.08 0.17253	0.63 1.0035	0.9546 0.0029441	$[0.28494, 0.97506]$ $[0.94871, 0.96049]$	1 1	23.516%
			$\Delta \equiv \delta^* \sigma^\beta = 1.1435, n_{\min} = 126, n_{\max} = 218$				
0.6	495.70	496.20 0.28030	0.50 1.0010	0.9516 0.003035	$[-0.0606, 1.0606]$ $[0.94553, 0.95767]$	1 1	23.506%
			$\Delta \equiv \delta^* \sigma^\beta = 0.68612, n_{\min} = 428, n_{\max} = 568$				

Moving forward, in Block (i) of Table 4 (with $\beta = 0$) the illustrated output came from such runs. However, we mildly modified the same program to work appropriately when $\beta = \frac{1}{8}, \frac{1}{12}$ with output illustrated in Blocks (ii)–(iii) in Table 4. Table 5 was compiled with the help of a similarly modified program under acceleration with $\beta = 0, \frac{1}{8}, \frac{1}{12}$. These computer programs were originally developed using R 4.0.0 (R Core Team 2020) [50], available for downloading from the publisher's website: https://www.routledge.com/Sequential-Methods-and-Their-Applications/Mukhopadhyay-deSilva/p/book/9780367386535, accessed on 7 September 2026, as one adopts the textbook of Mukhopadhyay and de Silva (2009) [23]. On the other hand, producing one's own R codes independently is not complicated either.

We ran simulations with $m = 10, 15, 20$ as well as $P^* = 0.90, 0.95$ but overall performances of the purely sequential selection strategy (28) and the accelerated sequential selection strategy (63) with $q = 2, 3$ did not fluctuate substantially in any sense. It is easy to feel a sense of stability in the choice of rather small m when we realize that updating the sample size sequentially a number of times in succession would have nulled any possible impact from a particular choice of m. The analogous features when $P^* = 0.90$ appeared rather similar to what we saw when $P^* = 0.95$. Hence, for brevity, our Table 4 highlights performances of the strategy (28) when $m = 10$, $P^* = 0.95$ whereas Table 5 emphasizes summaries from the strategy (63) when $m = 10$, $P^* = 0.95$ and $q = 2$.

Soon, we would move towards validating the s.o. expansion of $E[N - C]$ via data analyses under both selection strategies as laid out via (44) and (69) respectively for small, moderate and large sample sizes. In order to briefly do so, we needed the numerical values of $\kappa \equiv \kappa(k, \beta)$ when $k = 5$ and $\beta = 0, \frac{1}{12}, \frac{1}{8}$. Indeed we gave an expression for $\kappa(k, \beta)$ in (42) part (iii). Professor Swathi Venkatesan kindly provided the required κ-values correct up to 2-decimal places but we display them up to 4-dcimals as follows:

β :	0	$\frac{1}{12}$	$\frac{1}{8}$
$\kappa(5, \beta)$:	-0.7510	-0.7208	-0.7067

(73)

A sampling strategy is implemented under a fixed configuration as appropriate. The b^{th} replication was run by beginning with m pilot observations and then terminating the procedure at the sample size $N = n_b$, with $b = 1, ..., B$, $B = 5000$. Refer to Table 3 for a major set of notations displayed in Tables 4 and 5. They summarize the average simulated performances of the selection problem under the purely sequential (Table 4) and accelerated sequential (Table 5) strategies, respectively.

6.3. An Overview of Data Analyses

Each block in Tables 4 and 5 shows δ^* (col 1), C (col 2), the estimated values $\overline{n}$ and its estimated standard error (s.e.) $s_{\overline{n}}$ (col 3: up and down), the $\overline{n}/C$ and $\overline{n} - C$ values (col 4: up and down), the estimated values $\overline{p}$ and its s.e. $s_{\overline{p}}$ (col 5: up and down), two s.e. approximate 95% confidence intervals J_1 for $E[N - C]$ and J_2 for $P(CS)$ (col 6: up and down), indicator values A_1, A_2 (1 or 0) that check if the s.o. term for $E[N - C]$ or P^* is included in J_1, J_2 (col 7: up and down).

Block (i) corresponds to Bechhofer's (1954) [4] original approach ($\text{PZ}_{\textbf{Orig}}(\delta^*)$,$\text{IZ}_{\textbf{Orig}}(\delta^*)$), that is when $\beta = 0$. Blocks (ii) and (iii) show analogous output when $\beta = \frac{1}{8}$ and $\frac{1}{12}$ respectively. The last column (col 8) in Blocks (ii) and (iii) show the % savings in the average sample size ($\overline{n}$) corresponding to $\beta = \frac{1}{8}$ and $\frac{1}{12}$ compared to the average sample size ($\overline{n}$) corresponding to $\beta = 0$ highlighted in Block (i). Obviously, col 8 in Block (i) shows N/A (not applicable).

We include two precise examples highlighting how entries in col 8 are calculated:

Example 1 (Table 4, Block (ii), Col 8, 2nd Entry)**.** *We evaluate the % savings in the average sample size as follows:*

$$100\left[1-\left(\overline{n}_{(ii)}/\overline{n}_{(i)}\right)\right]=100(1-(92.68/138.12))=32.899.$$

Note that $\overline{n}_{(i)}$ *corresponding to* $\beta=0$ *and* $\overline{n}_{(ii)}$ *corresponding to* $\beta=\frac{1}{8}$ *come from Table 4.*

Example 2 (Table 5, Block (iii), Col 8, 3rd Entry)**.** *We evaluate the % savings in the average sample size as follows:*

$$100\left[1-\left(\overline{n}_{(iii)}/\overline{n}_{(i)}\right)\right]=100(1-(179.08/234.14))=23.516.$$

Note that $\overline{n}_{(i)}$ *corresponding to* $\beta=0$ *and* $\overline{n}_{(iii)}$ *corresponding to* $\beta=\frac{1}{12}$ *come from Table 5.*

Conclusions from Data Analyses

The following itemized commentaries address both Tables 4 and 5:

1. Comparing cols 2–3, we observe that $\overline{n}$-values stay very close to relevant C-values all across the board.
2. Col 4 (bottom) shows that $\overline{n}/C$-values stay very close to relevant 1 all across the board. We view this as data validation of the asymptotic f.o. efficiency property (33) even when C is rather small or medium.
3. Col 4 (top) shows that $\overline{n}-C$-values stay in a close proximity of $\kappa+1$ (Table 4) or $2(\kappa+1)$ (Table 5) as needed to validate asymptotic s.o. efficiency property. This becomes clear after revisiting (44) or (69) as a situation may call for. Considering the customary 2-standard deviation confidence interval J_1 for $E[N-C]$ in col 6 (top), we observe quickly that nearly all A_1 (col 7 top) values are 1 even when C is rather small or medium. One A_1-value each in Tables 4 and 5 went against that feature for small C, but that is certainly not out of line because of built-in randomness in simulated data.
4. Comparing col 5 (top) with the fixed preset value $P^*=0.95$, we get a feeling that the $\overline{p}$-values estimate $P(CS)$ tightly around P^* for C-values all across the board. More formally, considering the customary 2-standard deviation confidence interval J_2 for $P(CS)$ in col 6 (bottom), we observe quickly that all A_2 (col 7 bottom) values are 1 even when C is rather small or medium. These provide reasonable data validation in favor of asymptotic f.o. consistency property (39).
5. In col 8, we have emphasized that the % savings in the average sample sizes under the new formulation compared with those required by the original formulation of Bechhofer (1954) [4] can be substantial, amounting to nearly 33% and 23.5% when $\beta=\frac{1}{8},\frac{1}{12}$ respectively. Such % savings rates came from comparing **sample estimates** obtained via simulations.
6. In these specific contexts, however, could one guess such high rates of % savings theoretically? Recall the expression for $\eta(\beta,\sigma))$ from (18) and then, the population value of % savings may be approached as follows by evaluating:

$$\begin{gathered}100(1-\eta(\beta,\sigma))=100(1-\sigma^{-2\beta})=100(1-5^{-2\beta})\text{ which reduces}\\ \text{to 33.126\% and 23.528\% savings as }\textbf{population estimates}\text{ upon}\\ \text{fixing }\sigma=5\text{ and }\beta=\tfrac{1}{8},\tfrac{1}{12}\text{respectively.}\end{gathered}\tag{74}$$

   We are encouraged by observing the close proximity between the simulated data validation (col 8) and the population guideline seen from (19) and (74).
7. In order to run simulations, one must input a numerical value of σ, and hence it will be feasible to look at the magnitude expected theoretical savings % as shown in (74)

for a few β-values near 0 before running any simulation. A broad set of guideline may suggest selecting a β-value near 0 that leads to expected theoretical savings around 20% to 30% as long as such goal applies in practice. Too much savings on the other hand could substantially reduce the sample size by a large margin making an associated terminal selection unstable. Everyone involved must work together deciding what may be considered good or not-so-good in a field experiment.

8. The overall performances illustrated in Table 4 are incredibly close to those reported in Table 5. With $q = 2$, the accelerated sequential strategy (63)–(65) requires nearly 50% of the time, effort, resources, logistics, and data analyses compared with purely sequential strategy (28) to arrive at a terminal conclusion. A natural question arises: Which methodology should one implement in practice? A resolution is simple: If both methodologies are implementable, then an experimenter may lean towards acceleration.

7. Action Items in the Horizon: Future Research

We have introduced a practical way to modify the existing formulation of an IZ approach for selecting the best treatment thereby offering a large tent backed by a theoretical structure holding tightly a purely sequential methodology and its acceleration leading to important f.o. and s.o. asymptotic approximations. Nonlinear renewal theory have been emphasized in every turn. Ensuing data analyses look very appealing. We will feel humbled if this research energizes other interesting advances combining good measures of substantial theory as well as applications. In what follows, we suggest some specific ideas for future research activities:

1. Purely sequential and accelerated sequential methodologies are associated with s.o. properties. The literature on other multi-stage and modified sequential methodologies are rich, enjoying s.o. as well as higher-order asymptotic properties, growing independently outside of nonlinear renewal theory. These should be brought in to bear upon this present research. Some very large-scale and well-designed simulations as well as real data applications would highlight the importance of such multi-stage methodologies in practice.
2. Admittedly, this paper has been devoted to ranking means out of k independent normal treatments having equal variances. Why do we not go beyond ranking the means? Also, how about tackling other probability models such as exponential or negative exponential? We expect them to be fruitful by broadening with interesting research under analogous new and practical IZ formulations. Why could one not push distribution-free asymptotics?
3. Eventually, one could explore estimation problems of ranked parameter(s) or handle estimation problems after selection and ranking. High power computing and/or big data environments have certainly opened new challenging directions with many possible ways to discover other kinds of innovative research.

Indeed possibilities are endless.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: The Guest Editors, Takeshi Emura and Jun Hu, requested me to submit a paper for the consideration of possible publication in a special issue. I thank them for giving me this opportunity. My former Ph.D. advisee and collaborator, Swathi Venkatesan, enthusiastically shared with me the numerical values of $\kappa(5, \beta)$ corresponding to $\beta = 0, \frac{1}{12}, \frac{1}{8}$ which I displayed in (73) with her expressed consent. I warmly thank her for the kind consent. I remain grateful to two anonymous reviewers for helpful and detailed feedback which surely improved presentation. I also

appreciated comments from the Editor and the Assistant Editor. I thank them all for their continued encouragement and support.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Wald, A. *Sequential Analysis*; Wiley: New York, NY, USA, 1947.
2. Bahadur, R.R. On a Problem in the Theory of k Populations. *Ann. Math. Stat.* **1950**, *21*, 362–375 [CrossRef]
3. Bahadur, R.R.; Robbins, H. The Problem of the Greater Mean. *Ann. Math. Stat.* **1950**, *21*, 469–487 [CrossRef]
4. Bechhofer, R.E. A Single-Sample Multiple Decision Procedure for Ranking Means of Normal Populations with Known Variances. *Ann. Math. Stat.* **1954**, *25*, 16–39. [CrossRef]
5. Gupta, S.S. On a Decision Rule for a Problem in Ranking Means. Ph.D. Thesis, Institute of Statistics, University of North Carolina, Chapel Hill, NC, USA, 1956.
6. Bechhofer, R.E.; Dunnett, C.W.; Sobel, M. A Two-Sample Multiple Decision Procedure for Ranking Means of Normal Populations with a Common Unknown Variance. *Biometrika* **1954**, *41*, 170–176 [CrossRef]
7. Dunnett, C.W.; Sobel, M. Approximations to the Probability Integral and Certain Percentage Points of a Multivariate Analogue of Student's *t*-Distribution. *Biometrika* **1955**, *42*, 258–267. [CrossRef]
8. Tukey, J.W. The Philosophy of Multiple Comparisons. *Stat. Sci.* **1991**, *6*, 100–116. [CrossRef]
9. Tukey, J.W. Where Should Multiple Comparisons Go Next? In *Multiple Comparisons, Selection, and Applications in Biometry*; Charlie Dunnett 70th Birthday Festschrift; Hoppe, F.M., Ed.; Dekker: New York, NY, USA, 1993; pp. 187–207.
10. Hochberg, Y.; Tamhane, A.C. *Multiple Comparison Procedures*; Wiley: New York, NY, USA, 1987.
11. Hsu, J. *Multiple Comparisons: Theory and Methods*; Chapman and Hall/CRC: New York, NY, USA, 1996.
12. Gupta, S.S.; Sobel, M. On a Statistic Which Arises in Selection and Ranking Problems. *Ann. Math. Stat.* **1957**, *28*, 957–967. [CrossRef]
13. Bechhofer, R.E.; Kiefer, J.C.; Sobel, M. *Sequential Identification and Ranking Procedures*; University of Chicago Press: Chicago, IL, USA, 1968.
14. Robbins, H.; Sobel, M.; Starr, N. A Sequential Procedure for Selecting the Largest of k Means. *Ann. Math. Stat.* **1968**, *39*, 88–92. [CrossRef]
15. Dudewicz, E.J.; Dalal, S.R. Allocation of Observations in Ranking and Selection with Unequal Variances. *Sankhyā Ser. B* **1975**, *37*, 28–78.
16. Gibbons, J.D.; Olkin, I.; Sobel, M. *Selecting and Ordering Populations: A New Statistical Methodology*; Reprinted in 1999 by SIAM: Philadelphia, PA, USA; Wiley: New York, NY, USA, 1977.
17. Gupta, S.S.; Panchapakesan, S. *Multiple Decision Procedures: Theory and Methodology of Selecting and Ranking Populations*; Reprinted in 2002 by SIAM: Philadelphia, PA, USA; Wiley: New York, NY, USA, 1979.
18. Gupta, S.S.; Huang, D.-Y. *Multiple Statistical Decision Theory: Recent Developments*; Lecture Notes in Statistics 6; Springer: New York, NY, USA, 1981; p. 6.
19. Dudewicz, E.J.; Koo, J.O. *The Complete Categorized Guide to Statistical Selection and Ranking Procedures*; American Sciences Press: Columbus, OH, USA, 1981.
20. Mukhopadhyay, N.; Solanky, T.K.S. *Multistage Selection and Ranking Procedures: Second-Order Asymptotics*; Dekker: New York, NY, USA, 1994.
21. Bechhofer, R.E.; Santner, T.J.; Goldsman, D. *Design and Analysis of Experiments for Statistical Selection, Screening, and Multiple Comparisons*; Wiley: New York, NY, USA, 1995.
22. Mukhopadhyay, N. *Probability and Statistical Inference*; Available from Taylor & Francis; Dekker: New York, NY, USA, 2000.
23. Mukhopadhyay, N.; de Silva, B.M. *Sequential Methods and Their Applications*; Chapman and Hall/CRC: New York, NY, USA, 2009.
24. Venkatesan, S.; Mukhopadhyay, N. New Multistage Formulations of Minimum Risk Fixed-Size Confidence Region (MRFSCR) Problems for Estimating a Multivariate Normal Mean with Illustrations, Simulations and Data Analysis. *Commun. Stat.-Theory Methods* **2025**, *54*, 7230–7271. [CrossRef]
25. Dudewicz, E.J. Non-Existence of a Single-Sample Selection Procedure Whose P(CS) Is Independent of the Variance. *S. Afr. Stat. J.* **1971**, *5*, 37–39.
26. Dantzig, G.B. On the Non-Existence of Tests of Student's Hypothesis Having Power Functions Independent of σ. *Ann. Math. Stat.* **1940**, *11*, 186–192. [CrossRef]
27. Woodroofe, M. *Nonlinear Renewal Theory in Sequential Analysis*; CBMS Lecture Notes #39; SIAM: Philadelphia, PA, USA, 1982.
28. Ghosh, M.; Mukhopadhyay, N.; Sen, P.K. *Sequential Estimation*; Wiley: New York, NY, USA, 1997.
29. Gupta, S.S. Probability Integrals of Multivariate Normal and Multivariate *t*. *Ann. Math. Stat.* **1963**, *34*, 792–828. [CrossRef]

30. Gupta, S.S.; Nagel, K.; Panchapakesan, S. On the Order Statistics from Equally Correlated Normal Random Variables. *Biometrika* **1973**, *60*, 403–413. [CrossRef]
31. Rao, C.R. *Linear Statistical Inference*, 2nd ed.; Wiley: New York, NY, USA, 1973.
32. Tong, Y.L. *The Multivariate Normal Distribution*; Springer: New York, NY, USA, 1990.
33. Basu, D. On Statistics Independent of a Complete Sufficient Statistic. *Sankhyā* **1955**, *15*, 377–380
34. Stein, C. A Two Sample Test for a Linear Hypothesis Whose Power Is Independent of the Variance. *Ann. Math. Stat.* **1945**, *16*, 243–258. [CrossRef]
35. Stein, C. Some Problems in Sequential Estimation (abstract). *Econometrica* **1949**, *17*, 77–78.
36. Chow, Y.S.; Robbins, H. On the Asymptotic Theory of Fixed-Width Sequential Confidence Intervals for the Mean. *Ann. Math. Stat.* **1965**, *36*, 457–462. [CrossRef]
37. Ghosh, M.; Mukhopadhyay, N. On Two Fundamental Problems of Sequential Estimation. *Sankhyā Ser. B* **1976**, *38*, 203–218.
38. Wiener, N. The Ergodic Theorem. *Duke Math. J.* **1939**, *5*, 1–18. [CrossRef]
39. Ghosh, M.; Mukhopadhyay, N. Consistency and Asymptotic Efficiency of Two-Stage and Sequential Estimation Procedures. *Sankhyā Ser. A* **1981**, *43*, 220–227.
40. Woodroofe, M. Second Order Approximations for Sequential Point and Interval Estimation. *Ann. Stat.* **1977**, *5*, 984–995. [CrossRef]
41. Lai, T.L.; Siegmund, D. A Nonlinear Renewal Theory with Applications to Sequential Analysis I. *Ann. Stat.* **1977**, *5*, 946–954. [CrossRef]
42. Lai, T.L.; Siegmund, D. A Nonlinear Renewal Theory with Applications to Sequential Analysis II. *Ann. Stat.* **1979**, *7*, 60–76. [CrossRef]
43. Siegmund, D. *Sequential Analysis*; Springer: New York, NY, USA, 1985.
44. Mukhopadhyay, N.; Zhang, B. Theory of New Second-Order Expansions for the Moments of $100\rho\%$ Accelerated Sequential Stopping Times in Normal Mean Estimation Problems When $0 < \rho < 1$ Is Arbitrary. *Jap. J. Stat. Data Sci.* **2023**, *6*, 57–101. [CrossRef]
45. Mukhopadhyay, N.; Zhang, B. Theory and Practice of second-order expansions for moments of $100\rho\%$ Accelerated Sequential Stopping Times in Parametric and Nonparametric Estimation with Arbitrary Fractional ρ. *Seq. Anal.* **2024**, *43*, 179–210. [CrossRef]
46. Ghosh, M.; Mukhopadhyay, N. *Asymptotic Normality of Stopping Times in Sequential Analysis*; Unpublished Manuscript; Math-Stat Division, Research and Training School, Indian Statistical Institute: Calcutta, India, 1975.
47. Helmert, F.R. Die Genauigkeit der Formel von Peters zur Berechnung des wahrscheinlichen Beobachtungs-fehlers direkter Beo-bachtungen gleicher Genauigkeit. *Astron. Nachrichten* **1876**, *88*, 113–131. [CrossRef]
48. Hall, P. Sequential Estimation Saving Sampling Operations. *J. Roy. Stat. Soc. Ser. B* **1983**, *45*, 219–223. [CrossRef]
49. Mukhopadhyay, N. An Alternative Formulation of Accelerated Sequential Procedures with Applications to Parametric and Non-parametric Estimation. *Seq. Anal.* **1996**, *15*, 253–269. [CrossRef]
50. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020. Available online: https://www.R-project.org/ (accessed on 30 July 2026).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.