

Article

Adaptive Elimination in Curtailed Subset Selection: A Refined Procedure

Yifang Zhang * and Pinyuen Chen 

Department of Mathematics, Syracuse University, Syracuse, NY 13244, USA; pinchen@syr.edu

* Correspondence: yzhan377@syr.edu

Abstract

Subset selection from binomial populations is a classical problem in ranking and selection. In our earlier work, the curtailed procedure R_c introduced early stopping into this framework by terminating sampling once a current leader could no longer be overtaken. However, when several strong populations remain close, early stopping may be difficult to trigger and clearly inferior populations may continue to consume observations. To address this limitation, we propose an adaptive elimination refinement of R_c that permanently removes a population once its best possible future trajectory can no longer reach the current selection boundary. We study the additional error caused purely by elimination, defined as the probability that a population, especially the best population, is eliminated even though it would have been retained under the original curtailed procedure. Conservative bounds, approximation results, and simulation studies show that this elimination-induced risk remains very small, while the refined procedure yields noticeable reductions in sampling effort.

Keywords: subset selection; adaptive elimination; binomial populations; sequential sampling**MSC:** 62F07; 62L05; 60G40

1. Introduction

Subset selection from k binomial populations is a classical problem in ranking and selection. Its objective is to identify a subset that contains the population with the largest success probability. This formulation is especially relevant in screening experiments, where several promising candidates may need to be retained for further evaluation rather than forcing a single final choice. In the binomial setting, the fixed size procedure proposed by Gupta and Sobel remains the foundational reference because it provides a transparent selection rule for selecting populations close to the observed leader while guaranteeing a preassigned probability of correct selection under the worst parameter setting [1]. However, the fixed size structure does not use accumulating evidence during the experiment. In many modern experimental settings, it may be undesirable to continue sampling every population to the full design size after the data already indicate that some candidates are no longer competitive.

This concern is closely related to the broader motivation behind adaptive experimental designs. In clinical trials and other resource constrained experiments, interim information is often used to guide planned design modifications, such as early stopping, dropping poorly performing arms, changing allocation rules, or adding new treatment arms under



Academic Editor: Xuejun Wang

Received: 13 April 2026

Revised: 16 June 2026

Accepted: 8 July 2026

Published: 13 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

prespecified rules [2–5]. Multi-arm adaptive trials provide a useful example. In the STAMPEDE multi-arm multi-stage trial, research arms could stop accrual early if they failed to meet prespecified activity criteria at interim stages [6,7]. More generally, adaptive allocation methods for multi-arm trials are motivated by the idea that sampling effort should not continue to be spent equally on arms that are unlikely to be useful [8]. Although the present paper studies a different mathematical framework, namely binomial subset selection under a probability of correct selection requirement, the same operational issue arises: can sampling be reduced for populations that have already become noncompetitive?

In our earlier work, we introduced the curtailed subset selection procedure R_C to address one part of this issue in the binomial subset selection framework [9]. The main contribution of R_C was to incorporate an early stopping rule into subset selection. Sampling terminates once a current leader can no longer be overtaken, even under the most favorable remaining outcomes for its competitors. Simulation studies showed that this refinement can reduce the expected sample size $E(N)$ across several parameter settings while maintaining the desired probability of correct selection. At the same time, those results also revealed an important limitation. When several competitive populations remain close to one another, the stopping rule may be difficult to trigger. As a result, clearly poor populations may continue to be sampled even when their accumulated performance already suggests that they have little chance of entering the final selected subset.

This limitation motivates the adaptive elimination rule studied in the present paper. The proposed rule is not intended to replace the curtailed procedure R_C , nor does it change the subset selection objective. Instead, it acts as an in-processing refinement. At each stage, a population is permanently removed from further sampling if, even after receiving successes in all remaining observations, it still cannot reach the current selection boundary. Thus, adaptive elimination is distinct from early stopping. The stopping rule terminates the entire experiment once a leading population has an insurmountable advantage, whereas adaptive elimination only reduces the active set while sampling continues among the remaining populations. Its purpose is to improve the operational efficiency of R_C in settings where the stopping rule is not yet available.

The adaptive elimination rule also introduces a theoretical issue that must be separated from the ordinary probability of wrong selection. In the randomized curtailed procedure, populations exactly on the selection boundary may still be included with a calibrated probability x . Therefore, eliminating a population that can only reach this boundary may, in rare cases, remove a population that would otherwise have been retained under the original randomized curtailed procedure. The relevant additional error is not the full probability of wrong selection, but the marginal error caused by the elimination step itself. We therefore focus on the event that the best population is eliminated by the adaptive rule even though it would have been retained under the corresponding curtailed procedure without elimination.

The contributions of this paper are as follows. First, we formally introduce an adaptive elimination refinement of the curtailed subset selection procedure R_C for binomial populations. Second, we analyze the additional error induced solely by the elimination rule. For the focused case $k = 3$, we derive discrete analytical bounds and normal approximations for the elimination-induced error probability. For larger k , we provide a structural approximation supported by simulation rather than claiming a fully general theorem. Third, through simulation studies, we compare the original randomized curtailed procedure with its adaptive elimination refinement. The results show that adaptive elimination can reduce sampling effort while maintaining the target probability of correct selection in the configurations considered. These findings support adaptive elimination as a useful refinement of R_C , while also clarifying the limits of the current theoretical analysis.

2. Curtailed Procedure with Adaptive Elimination

2.1. Definitions and Notation

Let $\Pi_1, \Pi_2, \dots, \Pi_k$ denote the k binomial populations with corresponding unknown probabilities of success $p_1, p_2, \dots, p_k$. Let $\mathbf{p} = (p_1, p_2, \dots, p_k)$ be the true configuration, where $\mathbf{p} \in \Omega = [0, 1]^k$. Let n be the common maximum number of observations available from each population, and let N denote the total number of observations used in the experiment. Let $m \leq n$ denote the current sampling stage, and let $z_{i,m}$ denote the cumulative number of successes from population Π_i by stage m .

Because adaptive elimination removes populations during the experiment, we let $\mathcal{A}_m$ denote the set of active, non-eliminated populations after the elimination step at stage m . Before any elimination occurs, $\mathcal{A}_0 = \{1, 2, \dots, k\}$. For a given active set, define

$$z_m^1 = \max_{i \in \mathcal{A}_m} z_{i,m}$$

to be the largest observed success count among active populations at stage m . When no ambiguity arises, we refer to z_m^1 simply as the current largest success count.

Let P_{CS} denote the probability of correct selection, where a correct selection means that the final selected subset contains the true best population. Let P^* denote the required probability of correct selection, and let d be the smallest non-negative integer such that

$$P(CS \mid d) \geq P^* \quad \text{for all } \mathbf{p} \in \Omega.$$

The parameter d serves as the selection threshold in the curtailed procedure and is used both in the adaptive elimination rule and in the final selection rule. Let $x \in [0, 1]$ denote the randomized inclusion probability for populations that fall exactly on the selection boundary.

For reference, Table 1 summarizes the main notation and abbreviations used in the procedure.

Table 1. Main notation and abbreviations used in the adaptive curtailed procedure.

Notation	Meaning
R_C	Curtailed subset selection procedure
AE	Adaptive elimination
P_{CS}	Probability of correct selection
P^*	Required probability of correct selection
$E(N)$	Expected total number of observations
$E(S)$	Expected size of the selected subset
d	Selection threshold
x	Randomized inclusion probability at the boundary
$\mathcal{A}_m$	Active set after the elimination step at stage m
$z_{i,m}$	Cumulative successes from population Π_i by stage m
z_m^1	Largest success count among active populations at stage m

2.2. The Curtailed Procedure R_C with Adaptive Elimination

The adaptive elimination version of R_C proceeds in the following order at each stage: first, sample the active populations; second, update the cumulative success counts; third, remove populations that can no longer reach the current selection boundary; and fourth, check whether the stopping rule is satisfied among the remaining active populations. Once sampling terminates, the randomized selection rule is applied to the remaining active populations. In summary, it consists of the following four rules:

Sampling Rule. At each stage m , one observation is taken from each population Π_i with $i \in \mathcal{A}_{m-1}$, and the cumulative success count $z_{i,m}$ is updated. Populations that have already been eliminated are not sampled further.

Adaptive Elimination Rule. After the observations at stage m are collected, each active population is checked against the current selection boundary. If, for some $j \in \mathcal{A}_{m-1}$,

$$z_{j,m} + (n - m) \leq z_m^1 - d,$$

then population Π_j is immediately and permanently eliminated from further sampling. Equivalently, even if Π_j succeeds in every remaining observation, it cannot exceed the current boundary required for retention under the final selection rule.

The use of “ $\leq$ ” rather than “ $<$ ” is intentional. Under the randomized curtailed procedure, a population exactly on the boundary $z_m^1 - d$ may still be included with probability x if it remains active until the final selection step. The adaptive elimination rule treats such populations as noncompetitive for sampling purposes. This improves sampling efficiency, but it also creates the additional elimination-induced error studied in Section 3.

Stopping Rule. After the elimination step, stop sampling at the first stage m for which there exists a population Π_i among the remaining active populations such that

$$z_{i,m} \geq z_{j,m} + n - m \quad \text{for all } j \in \mathcal{A}_m, j \neq i.$$

This rule terminates the entire experiment only when an active population can no longer be overtaken by any other active population.

Selecting Rule. Once sampling stops at stage m , the final subset is determined from the remaining active populations, namely those indexed by $\mathcal{A}_m$, as follows:

- Populations satisfying

$$z_{j,m} \geq z_m^1 - (d - 1), \quad j \in \mathcal{A}_m,$$

are included in the final subset with certainty.

- Populations satisfying

$$z_{j,m} = z_m^1 - d, \quad j \in \mathcal{A}_m,$$

are eligible for randomized inclusion. For each such population, generate a random number $u \sim U(0, 1)$. If $u < x$, the population is included; otherwise, it is excluded.

Thus, adaptive elimination differs from the stopping rule. Elimination reduces the active set during sampling, whereas stopping terminates the whole experiment. It also differs from randomization, which acts only at the final selection step and does not alter the sampling process.

Figure 1 summarizes the order in which these rules are applied at each sampling stage.

Before turning to a worked example, we give an informal description of the idea behind adaptive elimination. Consider a drug screening setting in which several candidate treatments are tested at the same time, with the goal of keeping the promising ones for further study. As responses accumulate, a candidate may fall so far behind the current leader that, even if every one of its remaining patients responds, its best possible final count can at most reach the retention boundary but not clearly exceed it. Continuing to enroll patients on such a candidate adds little information and wastes experimental resources. Adaptive elimination formalizes this intuition: it permanently removes a candidate once its best possible future outcome can no longer move it past the current selection boundary, while sampling continues among the candidates that are still competitive. Because a candidate sitting exactly on the boundary is selected only with the randomization probability x , the rule

treats such borderline candidates as noncompetitive and removes them as well. This favors sampling efficiency at the cost of the small additional error analyzed in Section 3. The same operational principle appears in multi-arm clinical trials, where accrual to a research arm is stopped early once interim data indicate that the arm is unlikely to be worthwhile [6].

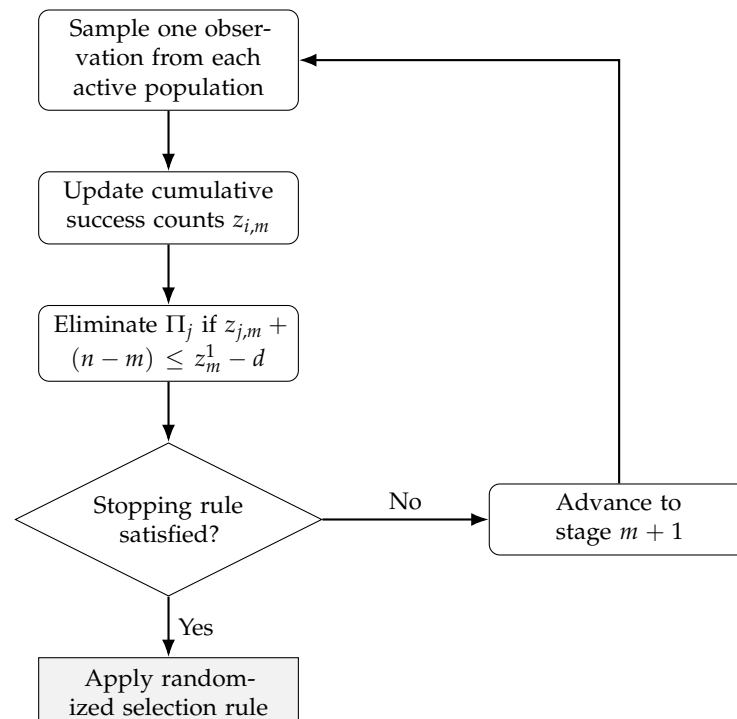


Figure 1. Flow of the adaptive elimination version of R_C at each sampling stage.

2.3. Example

We consider a scenario with $k = 5$ binomial populations, each allowed up to $n = 8$ observations. The success probabilities are $p = (0.9, 0.7, 0.5, 0.3, 0.1)$, and the selection threshold is set as $d = 2$. We assume no randomization in this illustrative path, so $x = 1$.

At each stage, one observation is drawn from each active population. Table 2 summarizes one possible sample path.

Table 2. Illustrative execution of the adaptive curtailed procedure.

Stage	Success (Pop 1–5)	Cumulative Success	Eliminated	Stop?
1	(1, 0, 0, 0, 0)	(1, 0, 0, 0, 0)	None	No
2	(1, 1, 0, 0, 0)	(2, 1, 0, 0, 0)	None	No
3	(1, 1, 1, 0, 0)	(3, 2, 1, 0, 0)	None	No
4	(1, 1, 1, 1, 1)	(4, 3, 2, 1, 1)	None	No
5	(1, 1, 0, 0, 0)	(5, 4, 2, 1, 1)	None	No
6	(1, 1, 1, 0, 0)	(6, 5, 3, 1, 1)	Π_4, Π_5	No
7	(1, 1, 0, –, –)	(7, 6, 3, –, –)	–	Yes (Π_1)

At stage 6, the current largest success count is $z_6^1 = 6$, so the selection boundary is $z_6^1 - d = 6 - 2 = 4$. For populations Π_4 and Π_5 , the maximum achievable final success count is $z_{j,6} + (8 - 6) = 1 + 2 = 3$ for $j = 4, 5$. Since $3 \leq 4$, both populations are eliminated, and sampling continues only among Π_1, Π_2 , and Π_3 , as the stopping rule has not yet been triggered.

At stage 7, the active cumulative success counts are $(7, 6, 3)$ for Π_1, Π_2 , and Π_3 . Population Π_1 now satisfies the stopping rule, since $z_{1,7} = 7 \geq z_{j,7} + 1$ for $j = 2, 3$. Sampling therefore stops after a total of 33 observations. For the final selection step, the boundary is $z_7^1 - d = 7 - 2 = 5$. Since $x = 1$, populations on the boundary are included with certainty, so the final selected subset is

$$S = \{\Pi_1, \Pi_2\}.$$

Without adaptive elimination, Π_4 and Π_5 would have continued to be sampled until the stopping time or the maximum stage. The example thus illustrates the operational role of adaptive elimination, which is that it reduces the active set before global stopping becomes available, while leaving the subset selection objective unchanged.

The example also shows why eliminating a population does not interfere with the stopping rule. The stopping rule is triggered only when the leader's lead over every other active population reaches $n - m$, that is, when no competitor can still overtake it. Populations Π_4 and Π_5 were eliminated at stage 6 because their best possible final counts (3) fell short of the retention boundary (4). They were already far behind the leader Π_1 and could neither become the leader nor stay close enough to it to keep the procedure running. Removing them leaves the decision unchanged at stage 7 where Π_1 still triggers stopping through $z_{1,7} \geq z_{j,7} + 1$ against the remaining competitors Π_2 and Π_3 . More generally, a population can be eliminated only when it already trails the leader by more than d , so it is never a candidate leader and never one of the close competitors that prevent stopping. Adaptive elimination therefore does not change the stopping time, and the only way it can turn a correct decision into an incorrect one is by removing the best population itself, which is the event analyzed in the next section.

3. Theoretical Analysis of Erroneous Elimination

This section studies the effect of adaptive elimination on the accuracy of the curtailed procedure, with or without randomization. The objective is not to analyze the full probability of wrong selection. Instead, we isolate a more specific error introduced solely by the elimination rule. All figures and tables in this section report author-generated simulation and analytical results.

The motivation comes from the modular use of adaptive elimination. In practice, an experimenter may first calibrate the design parameters (n, d, x) so that the curtailed procedure, with or without randomization, achieves the desired probability of correct selection. Adaptive elimination may then be added as a refinement intended to reduce sampling effort by removing clearly inferior populations early. From this perspective, the relevant theoretical question is how much additional risk is introduced when the original design parameters are reused without further recalibration.

Accordingly, the error of interest in this section is defined counterfactually. We consider the event that the best population is eliminated by the adaptive elimination rule even though it would still have been selected if sampling had continued under the corresponding curtailed procedure without elimination. Let this event be denoted by $\mathcal{E}$. This quantity differs from the usual probability of wrong selection. In particular, we do not count all wrong final outcomes. Instead, we focus only on sample paths on which adaptive elimination changes the final decision from correct to incorrect.

This distinction is important for interpreting the role of adaptive elimination. Eliminating the best population necessarily leads to an incorrect final outcome, but not every incorrect final outcome is caused by elimination. Some paths would already lead to an incorrect decision under the original curtailed procedure, even if no elimination rule were present. The present analysis therefore separates the marginal error introduced by adaptive elimination from the inherent error of the curtailed procedure itself.

This theoretical error criterion should also be distinguished from the simulations reported later in the paper. The analysis in the present section fixes (n, d, x) and studies the additional error event $\mathcal{E}$ under a counterfactual comparison with the non-elimination curtailed procedure. The later simulation study treats the adaptive procedure as a full procedure and reports its empirical P_{CS} , $E(N)$, and $E(S)$. Thus, the analytical results below are not used as formal guarantees for the recalibrated simulation study. Instead, they are used to quantify the additional risk attributable to the elimination mechanism itself.

3.1. Monte Carlo Evaluation of the Elimination-Induced Error Event

We emphasize that the Monte Carlo evaluation used in this section serves only to verify the analytical error bounds derived below. It is distinct from the simulation study of the full procedure reported in Section 4.

A simulated path is counted toward $\mathcal{E}$ only if the best population would be eliminated under the adaptive rule but would still be selected under the corresponding non-elimination curtailed procedure. Thus, this Monte Carlo calculation estimates the same counterfactual additional error event analyzed theoretically in this section.

3.2. Case Analysis: $k = 3$ and Elimination at Stage $n - 1$

We begin with the simplest nontrivial case, where $k = 3$ populations are compared. Population Π_1 is assumed to be the best population with success probability $p_1 = p + \delta$, while populations Π_2 and Π_3 share a common success probability p , where $\delta > 0$. Let n denote the maximum number of observations available from each population. The selection threshold d is calibrated under the original curtailed procedure without elimination so as to achieve the target probability of correct selection P^* .

We first consider the special case in which elimination occurs at the final possible stage, $m = n - 1$. Without loss of generality, suppose that population Π_2 attains the largest success count among the competitors at stage $n - 1$. The final-stage erroneous elimination event is defined as

$$\mathcal{E}_{n-1} := \left\{ \begin{array}{l} \Pi_1 \text{ is eliminated at stage } n - 1, \\ \Pi_1 \text{ would still be selected under the corresponding curtailed procedure without elimination} \end{array} \right\}.$$

This is a counterfactual event that it includes only those paths on which adaptive elimination removes the true best population, while the corresponding non-elimination procedure would still retain it.

For $\mathcal{E}_{n-1}$ to occur in the boundary configuration analyzed below, the following conditions must hold simultaneously:

1. Elimination of Π_1 at stage $n - 1$:

$$Z_{1,n-1} + 1 \leq Z_{2,n-1} - d.$$

2. No early stopping at stage $n - 1$: Population Π_2 does not yet dominate population Π_3 under the curtailed stopping rule. Since only one observation remains, this condition is

$$Z_{2,n-1} < Z_{3,n-1} + 1,$$

which, together with the assumption that Π_2 is a current leader among the competitors, gives

$$Z_{2,n-1} = Z_{3,n-1}.$$

3. **Counterfactual retention at stage n :** under the corresponding procedure without elimination, Π_1 would remain within the final selection boundary:

$$Z_{1,n} \geq Z_{\max,n} - d, \quad Z_{\max,n} := \max_{i=2,3} Z_{i,n}.$$

4. **Randomized boundary inclusion:** if Π_1 lies exactly on the final randomized boundary, then it is retained with probability x .

Let $(Z_{1,n-1}, Z_{2,n-1}, Z_{3,n-1}) = (\alpha, \beta, \gamma)$ denote the success counts observed up to stage $n - 1$, and let (X_1, X_2, X_3) denote the Bernoulli outcomes at the final stage. To obtain a tractable expression, we restrict attention to the boundary subevent in which Π_1 is eliminated exactly at the randomized selection boundary. This subevent is determined by four relations:

- $\beta = \alpha + d + 1$, from the elimination condition $\alpha + 1 = \beta - d$ taken at equality;
- $\beta = \gamma$, from the no stopping condition;
- $X_1 = 1$, so that Π_1 returns to the selection boundary at the final stage;
- $X_2 = X_3 = 0$, the most conservative continuation for the competitors.

Denote this boundary subevent by $\mathcal{E}_{n-1}^{(0)}$. Since the final-stage Bernoulli outcomes are independent of the counts observed up to stage $n - 1$, the probability of this subevent can be written as

$$P(\mathcal{E}_{n-1}^{(0)}) = x(p + \delta)(1 - p)^2 \sum_{\alpha=0}^{n-d-2} P(Z_{1,n-1} = \alpha) P(Z_{2,n-1} = \alpha + d + 1) \times P(Z_{3,n-1} = \alpha + d + 1). \quad (1)$$

The exact curtailed probabilities in (1) are difficult to compute directly because some sample paths may have stopped before stage $n - 1$. Replacing these curtailed probabilities by their corresponding full sampling binomial probabilities gives the computable bound

$$P(\mathcal{E}_{n-1}^{(0)}) \leq x(p + \delta)(1 - p)^2 \sum_{\alpha=0}^{n-d-2} \binom{n-1}{\alpha} (p + \delta)^\alpha (1 - p - \delta)^{n-1-\alpha} \times \left(\frac{n-1}{\alpha + d + 1} \right)^2 p^{2(\alpha+d+1)} (1 - p)^{2(n-\alpha-d-2)}. \quad (2)$$

The inequality in (2) follows because the full sampling binomial calculation includes paths that would already have terminated under the curtailed stopping rule before stage $n - 1$. Thus, within the final-stage boundary event considered here, the full sampling replacement provides a conservative upper bound. This calculation is used to quantify a tractable component of the elimination-induced error and to guide the numerical evaluation below. Meanwhile, broader erroneous elimination paths are assessed separately through the counterfactual Monte Carlo procedure described in Algorithm 1.

Algorithm 1 Monte Carlo evaluation of the elimination-induced error event

```

1: Initialize error count  $C \leftarrow 0$ 
2: for  $r = 1, \dots, R$  do
3:   Generate one sequential Bernoulli sample path for all populations under the speci-
     fied configuration  $p$ 
4:   Set  $\text{triggered} \leftarrow \text{False}$ 
5:   Set  $\text{selected\_without\_elimination} \leftarrow \text{False}$ 
6:   for  $m = 1, \dots, n$  do
7:     Update the cumulative success counts  $z_{i,m}$  for all populations
8:     if the adaptive elimination condition is met for the best population and
        $\text{triggered} = \text{False}$  then
9:       Record the trigger stage  $m$ 
10:      Set  $\text{triggered} \leftarrow \text{True}$ 
11:      if the curtailed stopping rule without elimination is satisfied then
12:        Apply the corresponding non-elimination selection rule
13:        if the best population would be selected then
14:          Set  $\text{selected\_without\_elimination} \leftarrow \text{True}$ 
15:      Exit the stage loop
16:   if  $\text{triggered} = \text{True}$  and  $\text{selected\_without\_elimination} = \text{True}$  then
17:      $C \leftarrow C + 1$ 
18: Estimate  $P(\mathcal{E})$  by  $C/R$ 

```

3.2.1. Numerical Evaluation of the Final-Stage Boundary Error

We next evaluate the final-stage boundary error component derived above for the case $k = 3$. The success probabilities are specified as

$$p_1 = p + \delta, \quad p_2 = p_3 = p,$$

where $\delta > 0$. Throughout this subsection, the randomization parameter is fixed at $x = 1$.

For each configuration, a reference value p^* is selected by numerically maximizing the computable bound in (2) over the admissible range of p . This choice provides a common reference point at which the Monte Carlo estimates and the analytical benchmark can be compared. Since the analytical expression in (2) is derived for the final-stage boundary subevent $\mathcal{E}_{n-1}^{(0)}$, the numerical results in this subsection should be interpreted as evidence for this tractable component of the elimination-induced error rather than as a complete evaluation of all possible erroneous elimination paths.

The column labeled *Analytical bound* reports the value of (2) evaluated at $p = p^*$. The column labeled *Simulated mean* reports the corresponding Monte Carlo estimate at the same value of p^* . To quantify Monte Carlo variability, we report a 95% confidence interval for the simulated mean. The confidence intervals are included to make clear that the comparison is between a deterministic analytical benchmark and a simulation estimate subject to Monte Carlo error.

Table 3 reports results for efficient values of d calibrated under the randomized curtailed procedure R_C . Table 4 reports corresponding results for more conservative choices of d based on Gupta and Sobel's fixed-size procedure R_{GS} . Related Bernoulli subset selection procedures, including the modified least-failures sampling approach, have been studied by Sanchez [10–12]. Multiple values of the superiority parameter δ are considered in both tables to examine how increasing separation between the best population and its competitors affects the magnitude of the final-stage boundary error.

Across all displayed configurations, the final-stage boundary error remains small in absolute magnitude. Under the efficient choices of d , the largest displayed analytical value is below 0.0012. When more conservative values of d are used, the corresponding values

are reduced further. In both tables, the error generally decreases as δ increases, reflecting the intuitive fact that clearer separation of the best population reduces the chance that it falls behind enough to be eliminated.

The analytical benchmark and the Monte Carlo estimates are close in magnitude across the displayed settings. As shown in Tables 3 and 4, the analytical value often lies inside or near the reported 95% confidence interval for the simulated mean. In a few cases, the simulated mean slightly exceeds the analytical value. These cases should not be interpreted as contradictions because the simulation estimate is subject to Monte Carlo variability and because the analytical expression is tied to the tractable final-stage boundary construction described above.

Table 3. Final stage boundary error under efficient d values (calibrated for R_C): analytical bound versus Monte Carlo estimates with 95% confidence intervals, for $k = 3$ and varying δ .

(n, d)	δ	p^*	Simulated Mean	CI (Low)	CI (High)	Analytical Bound
(10, 2)	0.05	0.396	0.001123	0.001012	0.001235	0.001104
(10, 2)	0.10	0.387	0.000757	0.000627	0.000886	0.000760
(10, 2)	0.15	0.376	0.000553	0.000462	0.000644	0.000489
(15, 2)	0.05	0.355	0.001147	0.001036	0.001257	0.001159
(15, 2)	0.10	0.357	0.000730	0.000639	0.000821	0.000751
(15, 2)	0.15	0.353	0.000463	0.000373	0.000553	0.000441
(20, 2)	0.05	0.331	0.001107	0.000991	0.001222	0.001078
(20, 2)	0.10	0.342	0.000657	0.000585	0.000728	0.000658
(20, 2)	0.15	0.344	0.000403	0.000333	0.000474	0.000353
(30, 2)	0.05	0.305	0.000877	0.000792	0.000961	0.000875
(30, 2)	0.10	0.330	0.000467	0.000383	0.000550	0.000475
(30, 2)	0.15	0.340	0.000227	0.000159	0.000294	0.000214
(50, 1)	0.05	0.235	0.000760	0.000653	0.000867	0.000820
(50, 1)	0.10	0.300	0.000313	0.000248	0.000378	0.000395
(50, 1)	0.15	0.329	0.000127	0.000089	0.000164	0.000142

Table 4. Final stage boundary error under conservative d values (calibrated for the Gupta and Sobel's fixed size procedure R_{GS}): analytical bound versus Monte Carlo estimates with 95% confidence intervals for $k = 3$ and varying δ .

(n, d)	δ	p^*	Simulated Mean	CI (Low)	CI (High)	Analytical Bound
(10, 4)	0.05	0.484	7.00×10^{-5}	4.01×10^{-5}	9.99×10^{-5}	7.02×10^{-5}
(10, 4)	0.10	0.464	2.67×10^{-5}	1.06×10^{-5}	4.28×10^{-5}	3.71×10^{-5}
(10, 4)	0.15	0.442	1.67×10^{-5}	3.10×10^{-6}	3.02×10^{-5}	1.82×10^{-5}
(15, 5)	0.05	0.466	5.67×10^{-5}	3.24×10^{-5}	8.10×10^{-5}	5.23×10^{-5}
(15, 5)	0.10	0.448	2.67×10^{-5}	8.03×10^{-6}	4.53×10^{-5}	2.30×10^{-5}
(15, 5)	0.15	0.428	1.00×10^{-5}	0.00	2.09×10^{-5}	9.05×10^{-6}
(20, 6)	0.05	0.462	2.33×10^{-5}	5.30×10^{-6}	4.14×10^{-5}	3.29×10^{-5}
(20, 6)	0.10	0.444	2.67×10^{-5}	1.06×10^{-5}	4.28×10^{-5}	1.20×10^{-5}
(20, 6)	0.15	0.424	6.67×10^{-6}	0.00	1.57×10^{-5}	3.81×10^{-6}
(30, 7)	0.05	0.448	3.67×10^{-5}	1.27×10^{-5}	6.06×10^{-5}	3.06×10^{-5}
(30, 7)	0.10	0.434	3.33×10^{-6}	0.00	9.87×10^{-6}	8.86×10^{-6}
(30, 7)	0.15	0.416	6.67×10^{-6}	0.00	1.57×10^{-5}	2.07×10^{-6}
(50, 10)	0.05	0.450	1.00×10^{-5}	0.00	2.09×10^{-5}	7.96×10^{-6}
(50, 10)	0.10	0.436	0.00	0.00	0.00	1.26×10^{-6}
(50, 10)	0.15	0.416	0.00	0.00	0.00	1.40×10^{-7}

Taken together, these results suggest that the final-stage boundary component of the elimination-induced error is small under the examined configurations. In the more conservative d settings, some simulated probabilities are effectively zero at the given Monte Carlo effort, while the analytical values remain on the order of 10^{-6} or smaller. This is consistent with a rare event setting (see Figure 2).

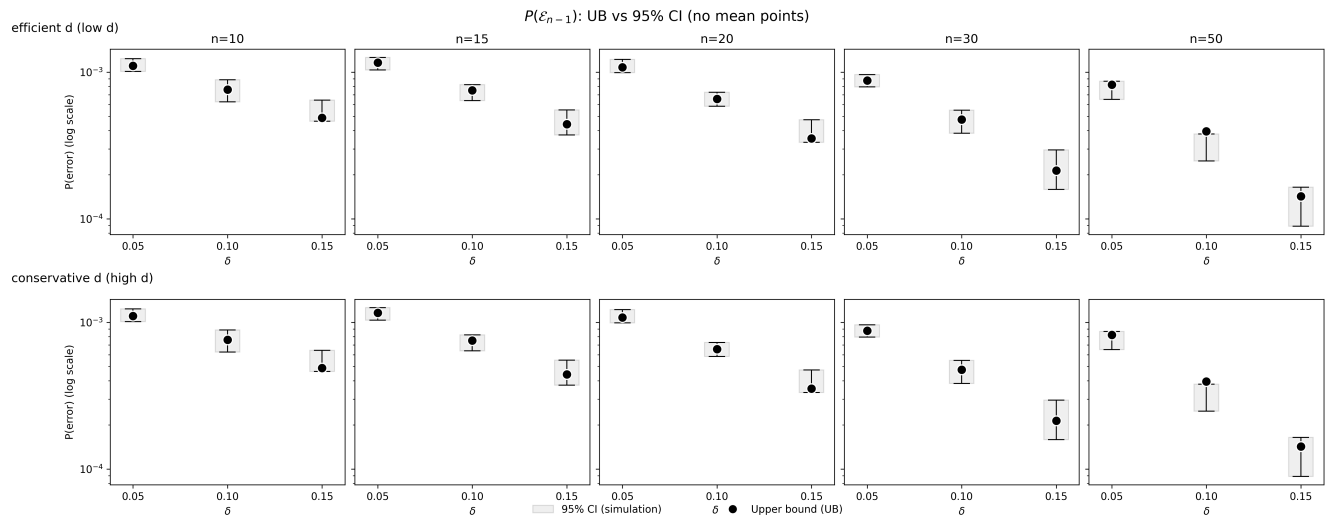


Figure 2. Final stage boundary error for $k = 3$: analytical bound versus 95% Monte Carlo confidence intervals, under efficient d (top row) and conservative d (bottom row).

At this stage, the purpose is to understand the additional error attributable to the elimination rule, not to evaluate the full operating characteristics of the adaptive procedure. The preceding results show that, in the focused case of final-stage boundary elimination, this additional error is small in magnitude. The next step is to develop a more convenient analytical approximation to the same final-stage boundary expression.

3.2.2. Normal Approximation to the Discrete Boundary Expression

In the previous discussion, the final-stage boundary component of the elimination-induced error was evaluated through a discrete binomial sum. Although this discrete expression is straightforward to compute numerically, repeated evaluation can still be inconvenient, especially when the success probability p must be optimized or when many parameter settings are examined. To complement the discrete calculation, we introduce a normal approximation to the binomial probabilities appearing in the discrete expression. The purpose of this approximation is to provide a simpler closed-form surrogate that captures the scale and qualitative behavior of the final-stage boundary component.

Recall that, under the configuration $p_1 = p + \delta$ and $p_2 = p_3 = p$, the final-stage boundary component is

$$P(\mathcal{E}_{n-1}^{(0)}) = x(p + \delta)(1 - p)^2 \sum_{\alpha=0}^{n-d-2} P(Z_{1,n-1} = \alpha) P(Z_{2,n-1} = \alpha + d + 1) \times P(Z_{3,n-1} = \alpha + d + 1). \quad (3)$$

Replacing the curtailed stage $n - 1$ probabilities in (3) by the corresponding full sampling binomial probabilities gives the discrete analytical bound

$$P(\mathcal{E}_{n-1}^{(0)}) \leq U_{n-1} := x(p + \delta)(1 - p)^2 \sum_{\alpha=0}^{n-d-2} \binom{n-1}{\alpha} (p + \delta)^\alpha (1 - p - \delta)^{n-1-\alpha} \\ \times \binom{n-1}{\alpha + d + 1}^2 p^{2(\alpha+d+1)} (1 - p)^{2(n-\alpha-d-2)}. \quad (4)$$

By the De Moivre–Laplace normal approximation, the binomial probabilities in (4) can be approximated by normal densities with means and variances

$$\mu_1 = (n - 1)(p + \delta), \quad \sigma_1^2 = (n - 1)(p + \delta)(1 - p - \delta)$$

for population Π_1 , and

$$\mu = (n - 1)p, \quad \sigma^2 = (n - 1)p(1 - p)$$

for populations Π_2 and Π_3 . Thus,

$$P(Z_{1,n-1} = \alpha) \approx \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{(\alpha - \mu_1)^2}{2\sigma_1^2}\right), \quad (5) \\ P(Z_{j,n-1} = \alpha + d + 1) \approx \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\alpha + d + 1 - \mu)^2}{2\sigma^2}\right), \quad j = 2, 3.$$

Using these approximations, we replace the discrete sum in U_{n-1} by a continuous integral over $-0.5 \leq \alpha \leq n - d - 2 + 0.5$, with continuity corrected limits $(\alpha_{\min}, \alpha_{\max}) = (-0.5, n - d - 2 + 0.5)$. Let $\tilde{U}_{n-1}$ denote the resulting approximation:

$$\tilde{U}_{n-1} = x(p + \delta)(1 - p)^2 \int_{\alpha_{\min}}^{\alpha_{\max}} \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{(\alpha - \mu_1)^2}{2\sigma_1^2}\right) \left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\alpha + d + 1 - \mu)^2}{2\sigma^2}\right) \right]^2 d\alpha \\ = x(p + \delta)(1 - p)^2 \frac{1}{(2\pi)^{\frac{3}{2}} \sigma_1 \sigma^2} \int_{\alpha_{\min}}^{\alpha_{\max}} \exp\left[-\frac{(\alpha - \mu_1)^2}{2\sigma_1^2} - \frac{(\alpha + d + 1 - \mu)^2}{\sigma^2}\right] d\alpha. \quad (6)$$

To simplify the exponential expression in (6), define

$$Q(\alpha) = \frac{(\alpha - \mu_1)^2}{2\sigma_1^2} + \frac{(\alpha + d + 1 - \mu)^2}{\sigma^2} \\ = A(\alpha - B)^2 + C, \quad (7)$$

where

$$A = \frac{1}{2\sigma_1^2} + \frac{1}{\sigma^2}, \quad (8)$$

$$B = \frac{\frac{\mu_1}{2\sigma_1^2} + \frac{\mu - d - 1}{\sigma^2}}{A} = \frac{\mu_1 \sigma^2 + 2(\mu - d - 1)\sigma_1^2}{\sigma^2 + 2\sigma_1^2}, \quad (9)$$

$$C = \frac{\mu_1^2}{2\sigma_1^2} + \frac{(\mu - d - 1)^2}{\sigma^2} - AB^2. \quad (10)$$

Substituting (7) into (6) yields

$$\tilde{U}_{n-1} = x(p + \delta)(1 - p)^2 \frac{1}{(2\pi)^{\frac{3}{2}} \sigma_1 \sigma^2} e^{-C} \int_{\alpha_{\min}}^{\alpha_{\max}} e^{-A(\alpha - B)^2} d\alpha. \quad (11)$$

Extending the finite integral in (11) to the whole real line gives a closed-form diagnostic, which we denote by $\tilde{U}_{n-1}^G$ (the superscript G refers to the Gaussian integral over $\mathbb{R}$):

$$\begin{aligned}\tilde{U}_{n-1}^G &:= x(p+\delta)(1-p)^2 \frac{1}{(2\pi)^{\frac{3}{2}}\sigma_1\sigma^2} e^{-C} \int_{-\infty}^{\infty} e^{-A(\alpha-B)^2} d\alpha \\ &= x(p+\delta)(1-p)^2 \frac{1}{(2\pi)^{\frac{3}{2}}\sigma_1\sigma^2} e^{-C} \sqrt{\frac{\pi}{A}} \\ &= \frac{x(p+\delta)(1-p)^2}{2\pi\sigma_1\sigma^2} \cdot \frac{e^{-C}}{\sqrt{2A}}.\end{aligned}\quad (12)$$

The quantity $\tilde{U}_{n-1}^G$ is a closed-form normal approximation to the discrete analytical bound U_{n-1} . It is not interpreted as a rigorous upper bound for $P(\mathcal{E}_{n-1}^{(0)})$ or for the full elimination-induced error probability. The discrete quantity U_{n-1} remains the conservative analytical bound for the final-stage boundary component and the normal expression is used only as a computationally convenient approximation to the scale and shape of U_{n-1} .

To assess the practical accuracy of the normal approximation, we compare $\tilde{U}_{n-1}^G$ with the discrete bound U_{n-1} across several combinations of (n, d, δ) , varying the success probability p over its admissible range while fixing $k = 3$ and $x = 1$. As shown in Figure 3, the normal approximation tracks the discrete bound closely across the displayed configurations. In each panel, the blue curve represents the discrete bound U_{n-1} and the orange curve represents its normal approximation $\tilde{U}_{n-1}^G$ (labeled “Gaussian bound” in the figure). The approximation may lie slightly above or below the discrete bound in small sample settings, but the two curves show similar peak locations and comparable magnitudes over the displayed range of p .

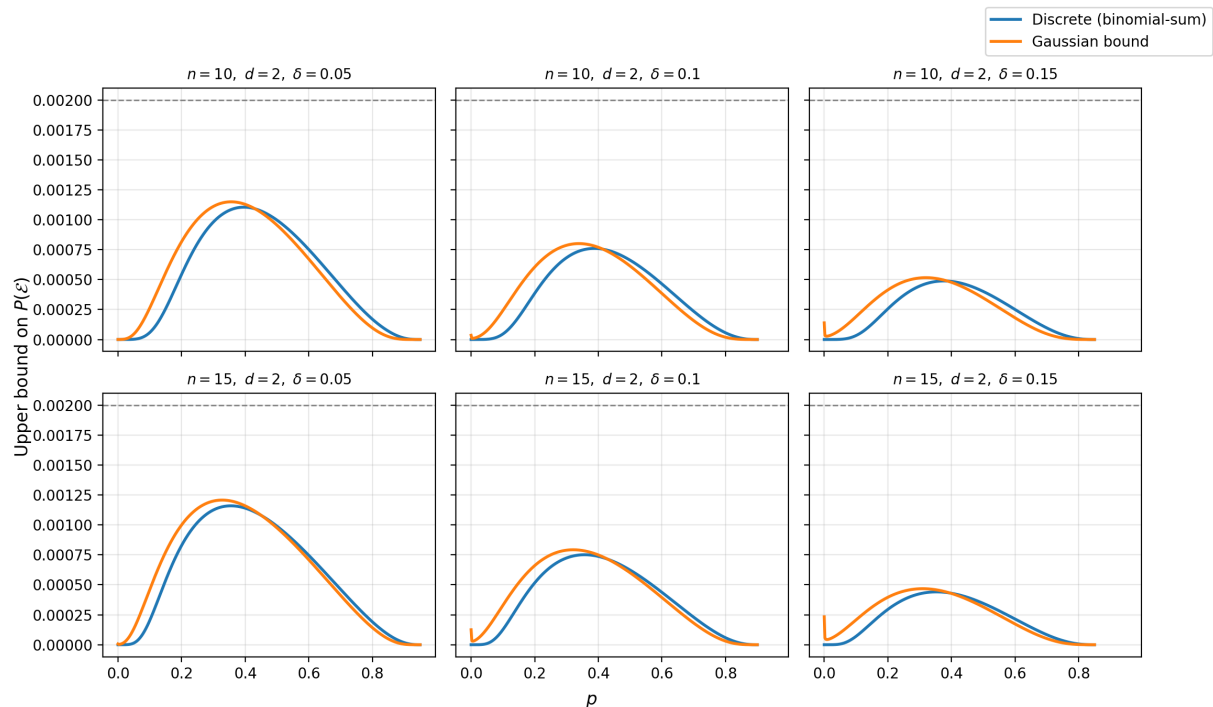


Figure 3. Discrete bound U_{n-1} and normal approximation $\tilde{U}_{n-1}^G$ for the final-stage boundary component.

These comparisons suggest that the normal approximation is a useful computational surrogate for the discrete final-stage boundary bound. It offers a convenient diagnostic for the scale of this boundary component, without requiring repeated evaluation of the full discrete sum. At the same time, the analysis in this subsection remains limited to the case

in which elimination occurs at the final stage $m = n - 1$. To understand earlier elimination events, we next extend the same line of reasoning to stages before termination.

3.3. Case Analysis: $k = 3$ and Elimination at an Arbitrary Stage m

We now extend the analysis to the case in which erroneous elimination may occur at an arbitrary stage $m \leq n - 1$, rather than only at the final possible stage $n - 1$. As before, consider three binomial populations (Π_1, Π_2, Π_3) with success probabilities

$$p_1 = p + \delta, \quad p_2 = p_3 = p,$$

where $\delta > 0$. Let $\mathcal{E}_m$ denote the counterfactual event that Π_1 is eliminated at stage m even though it would still be retained under the corresponding curtailed procedure without elimination.

The full event $\mathcal{E}_m$ contains many possible path configurations. To obtain a tractable expression, we analyze a structured boundary component of this event. Let $(Z_{1,m}, Z_{2,m}, Z_{3,m}) = (\alpha, \beta, \gamma)$ denote the cumulative success counts at stage m . Since populations Π_2 and Π_3 are exchangeable under $p_2 = p_3 = p$, we use the representative ordering $\beta \geq \gamma$. This convention is used only to describe the relevant boundary path structure. We do not introduce an additional multiplier for the two possible labelings of the inferior populations because the stopping and boundary conditions can create overlapping or tied configurations rather than two disjoint ordered path classes.

Under this representative ordering, the boundary component considered here is characterized as follows:

- Population Π_1 can return to the final randomized boundary under the non-elimination continuation only if it obtains the maximum possible number of successes in the remaining $n - m$ observations while the leading competitor Π_2 obtains no further success. This gives

$$\alpha + (n - m) = \beta - d, \quad \text{or equivalently} \quad \beta = \alpha + (n - m) + d. \quad (13)$$

- To avoid immediate stopping at stage m through the dominance of Π_2 over Π_3 , the third population must still be close enough that Π_2 has not become an insurmountable leader. Under the boundary representation above, this is written as

$$\gamma + (n - m) > \beta, \quad \text{so that} \quad \gamma = \alpha + d + j, \quad j = 1, \dots, n - m. \quad (14)$$

Let (X_1, X_2, X_3) denote the additional numbers of successes from stage $m + 1$ to stage n . Since future Bernoulli trials are independent of the observations up to stage m , conditional on the stage m configuration we have

$$\begin{aligned} X_1 &\sim \text{Bin}(n - m, p + \delta), \\ X_2 &\sim \text{Bin}(n - m, p), \\ X_3 &\sim \text{Bin}(n - m, p), \end{aligned}$$

with X_1, X_2, X_3 being mutually independent.

A sufficient continuation event that returns Π_1 to the final randomized boundary under the non-elimination continuation is

$$X_1 = n - m, \quad X_2 = 0, \quad X_3 \leq n - m - j, \quad (15)$$

where $j = 1, \dots, n - m$. This tail event is deliberately restrictive because it is chosen to preserve a tractable boundary expression rather than to enumerate all possible counterfactual selection paths.

Let $\mathcal{E}_m^{(0)}$ denote the boundary component determined by (13)–(15). Then

$$\mathbb{P}(\mathcal{E}_m^{(0)}) = x \sum_{\alpha=0}^{2m-n-d} \sum_{j=1}^{n-m} \underbrace{\mathbb{P}\left(\begin{matrix} Z_{1,m} = \alpha, Z_{2,m} = \alpha + (n-m) + d, \\ Z_{3,m} = \alpha + d + j \end{matrix}\right)}_{\text{stage-}m \text{ boundary configuration}} \underbrace{\times \mathbb{P}(X_1 = n-m) \mathbb{P}(X_2 = 0) \mathbb{P}(X_3 \leq n-m-j)}_{\text{independent continuation event}}. \quad (16)$$

As in the final-stage analysis, replacing the curtailed probabilities achieved at stage m by their full sampling binomial counterparts gives a computable analytical bound for this boundary component:

$$\begin{aligned} \mathbb{P}(\mathcal{E}_m^{(0)}) \leq U_m := & x \sum_{\alpha=0}^{2m-n-d} \sum_{j=1}^{n-m} \binom{m}{\alpha} (p + \delta)^\alpha (1 - p - \delta)^{m-\alpha} \\ & \times \binom{m}{\alpha + (n-m) + d} p^{\alpha + (n-m) + d} (1 - p)^{2m-n-d-\alpha} \\ & \times \binom{m}{\alpha + d + j} p^{\alpha + d + j} (1 - p)^{m-\alpha-d-j} \\ & \times (p + \delta)^{n-m} (1 - p)^{n-m} F_{\text{Bin}(n-m,p)}(n-m-j). \end{aligned} \quad (17)$$

Here $F_{\text{Bin}(N,p)}(\cdot)$ denotes the binomial distribution function with size parameter N and success probability p . When $m = n - 1$, we have $n - m = 1$, $j = 1$, and

$$F_{\text{Bin}(1,p)}(0) = 1 - p.$$

Thus, (17) reduces to the final-stage boundary expression derived in the previous subsection.

The summation is nonempty only when the counts achieved at stage m satisfy the feasibility constraints $\beta \leq m$ and $\gamma \leq m$. In particular, the condition

$$\beta = \alpha + (n - m) + d \leq m$$

implies the lower stage requirement

$$m \geq \left\lceil \frac{n + d}{2} \right\rceil.$$

Therefore, a stage summed analytical bound for the represented boundary component is

$$U^{(0)} = \sum_{m=\lceil (n+d)/2 \rceil}^{n-1} U_m. \quad (18)$$

The quantity $U^{(0)}$ should be interpreted as a structured analytical bound for the boundary component considered above, not as a complete closed-form expression for all possible erroneous elimination paths. A direct union bound over every labeling and every continuation pattern would be possible in principle, but it would also be substantially looser and less informative in this focused $k = 3$ setting because tied and overlapping path classes can occur. For this reason, we retain the representative boundary construction and use the

counterfactual Monte Carlo procedure to assess the actual scale of the elimination-induced error under calibrated design parameters.

Stage-Wise Boundary Bounds and Numerical Results

The preceding subsection gives a stage-wise analytical bound U_m for the represented boundary component $\mathcal{E}_m^{(0)}$. The purpose of the present numerical study is to examine the scale of this stage-wise component and to compare the analytical values with Monte Carlo estimates of the same counterfactual elimination event.

For each configuration, we conduct Monte Carlo simulations using the counterfactual procedure described in Algorithm 1. Each simulation path is generated from the Bernoulli populations under the specified configuration. A path is counted at stage m only if the best population triggers the adaptive elimination rule at that stage and would still be selected under the corresponding non-elimination curtailed procedure. Thus, the simulated stage-wise estimates are aligned with the elimination-induced error criterion, rather than with the full probability of wrong selection.

We consider

$$(n, d) \in \{(10, 2), (15, 2), (20, 2), (30, 2), (50, 1)\},$$

with $p = 0.5$, $x = 1$, and $\delta \in \{0.05, 0.10, 0.15\}$. For each configuration, 10,000 simulation replications are used within each independent run, and the runs are repeated to compute the empirical mean and a 95% Monte Carlo confidence interval.

Figure 4 compares the stage-wise analytical bound U_m with the simulated estimates across feasible sampling stages. In each panel, the red curve represents U_m , while the dashed blue curves and shaded band represent the 95% confidence interval for the simulated stage-wise error estimate.

Several patterns are visible in Figure 4. First, the stage-wise error is essentially zero during the early stages and becomes noticeable only when m approaches the terminal stage n . This pattern is consistent with the feasibility constraint $m \geq \lceil (n + d)/2 \rceil$ derived above, which determines the earliest stage at which the boundary construction can be nonempty.

Second, across the displayed parameter configurations, the analytical bound is close to or above the simulated estimates near the stages where the error is largest. The agreement is not interpreted as a proof that the analytical expression captures every possible erroneous elimination path. Instead, it shows that the represented boundary component has the same numerical scale as the counterfactual simulation estimates in these settings.

Third, the magnitude of the stage-wise error decreases as the signal strength δ increases. When the best population is better separated from the inferior populations, the probability that it falls far enough behind to trigger erroneous elimination becomes smaller.

Finally, even in the least favorable configurations considered here, the stage-wise elimination-induced error probabilities remain on the order of 10^{-3} or smaller. The stage-wise evidence therefore supports the conclusion that, for $k = 3$ and the parameter settings examined, adaptive elimination introduces only a small additional error. This conclusion is restricted to the calibrated settings and counterfactual error criterion considered here; the full operating characteristics of the adaptive procedure are evaluated separately in the simulation section below.

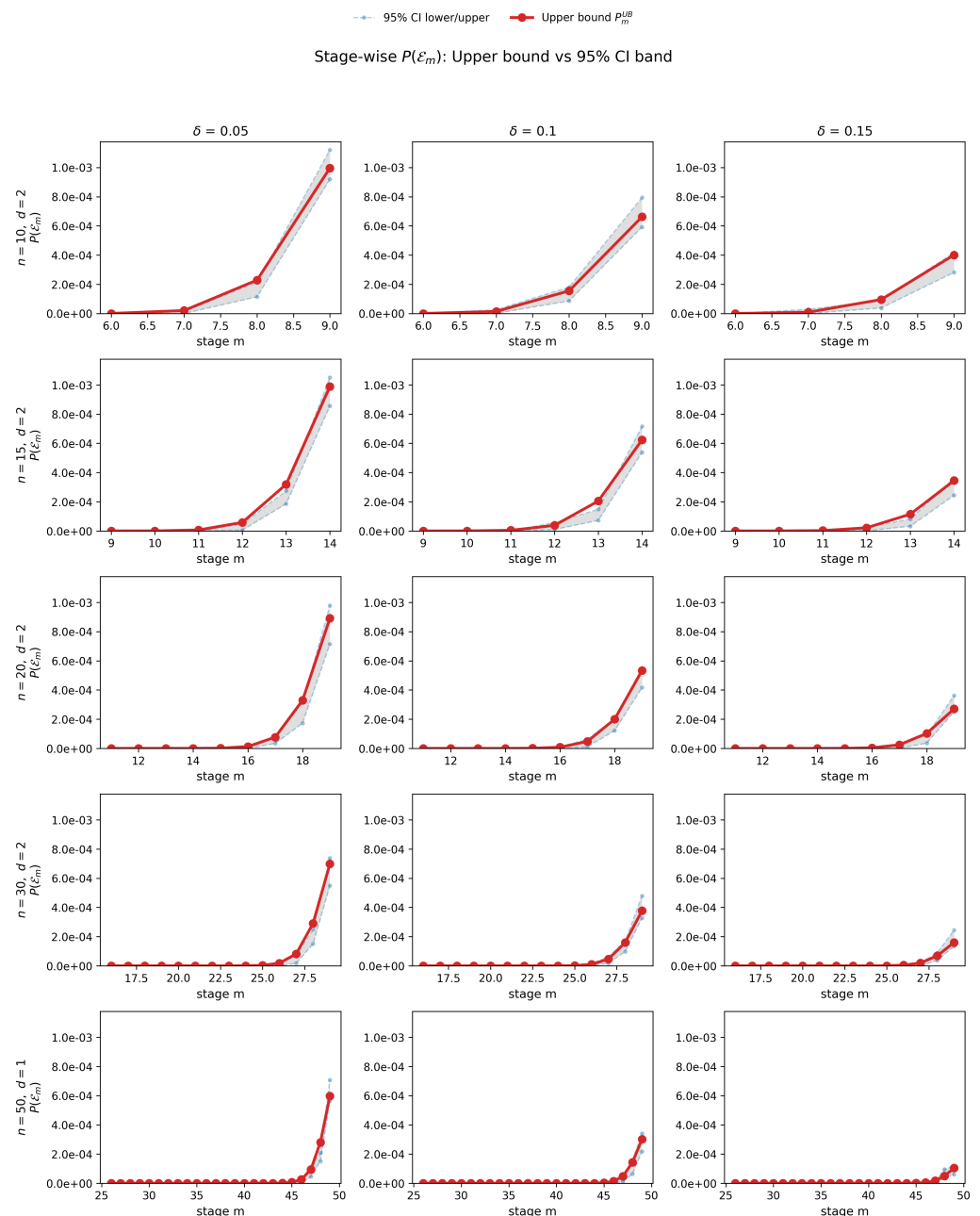


Figure 4. Stage-wise boundary bounds U_m and simulated 95% confidence intervals for the elimination-induced error estimates.

3.4. Implications for General k Under Adaptive Elimination

The analysis above gives a structured characterization of elimination-induced error for the case $k = 3$. We now discuss how the same reasoning may be extended when $k > 3$. The purpose of this subsection is to use the $k = 3$ derivation to identify the main path features that remain relevant when more populations are present, rather than to provide a complete proof for the general case. We then treat the resulting expression as a structural approximation whose adequacy is assessed numerically.

When the number of populations exceeds three, the erroneous elimination event becomes substantially more complicated. Several inferior populations may interact with the curtailed stopping rule at the same stage, and after the adaptive elimination condition is triggered, the remaining active populations may follow many different continuation paths before the procedure stops. A full enumeration of all possible leader assignments,

hold-back populations, ties, and continuation patterns would therefore require a much more detailed partition of the sample path space.

A direct combinatorial expansion over possible labels of the inferior populations is not used here. Such an expansion would be valid only if the underlying path classes were disjoint and were counted without overlap. In the present problem, this condition is difficult to verify because multiple inferior populations may be tied, may jointly prevent stopping, or may correspond to the same effective boundary configuration. A crude multiplicative correction could therefore be misleading since it may overcount overlapping path classes while still failing to represent the full curtailed path structure.

We therefore apply a structural approximation that extends the representative path reasoning used in the $k = 3$ analysis. As before, let population Π_1 denote the best population with success probability $p_1 = p + \delta$, and let populations $\Pi_2, \dots, \Pi_k$ denote inferior populations with common success probability p . We focus on paths where Π_1 satisfies the adaptive elimination condition at stage m , while the corresponding curtailed procedure without elimination would continue beyond that stage.

Let $Z_{i,m}$ denote the number of successes of population Π_i at stage m . We use a representative ordering in which Π_2 is the current leader at stage m . Under this ordering, the curtailed procedure can continue beyond stage m only if at least one other inferior population remains sufficiently close to Π_2 so that the stopping rule is not triggered. In this sense, the remaining inferior populations act collectively as hold-back populations. At the same time, the adaptive elimination condition for Π_1 depends only on the current leader through the boundary relation $Z_{1,m} + (n - m) = Z_{2,m} - d$.

Thus, after the representative leader Π_2 is fixed, the remaining inferior populations affect whether the non-elimination procedure can continue through the stopping rule, but they do not change the elimination boundary for Π_1 itself. This observation motivates a fixed size continuation construction that after stage m , the random curtailed continuation is replaced by $n - m$ additional observations. This construction does not provide a full theorem for general k . Its role is to give a tractable approximation to the boundary mechanism by removing the early stopping complication from the continuation phase.

Under this fixed size continuation construction, we focus on the following tail scenario:

- Population Π_1 obtains successes in all remaining $n - m$ observations.
- The representative leader Π_2 obtains no successes in the remaining $n - m$ observations.
- The remaining populations Π_j , for $j = 3, \dots, k$, do not exceed the final success count of Π_2 .

The resulting expression should therefore be interpreted as a structural approximation for the general k boundary mechanism, not as a rigorously established upper bound for all erroneous elimination paths. Its purpose is to provide an interpretable analytical guide for how the elimination-induced error may scale with k . The numerical simulations below are used to examine whether this approximation is consistent with the observed behavior of the adaptive procedure under the configurations considered.

3.4.1. A Structural Approximation for General k

To see how the structural argument extends beyond $k = 3$, it is helpful to first examine the case $k = 4$. Fix a stage m with $\left\lceil \frac{n+d}{2} \right\rceil \leq m \leq n - 1$, for a representative path configuration we suppose that population Π_2 is the current leader at stage m . Let $Z_{i,m}$ denote the number of successes observed from population Π_i up to stage m , and let X_i denote the number of successes obtained from the remaining $n - m$ observations under the fixed-size continuation construction. Thus,

$$Z_{i,m} \sim \text{Bin}(m, p_i), \quad X_i \sim \text{Bin}(n - m, p_i), \quad (19)$$

with $Z_{i,m}$ and X_i independent for each i .

As in the $k = 3$ boundary construction, suppose that the best population Π_1 is eliminated at stage m . Let $Z_{1,m} = \alpha$. The elimination boundary then requires the representative leader to satisfy

$$Z_{2,m} = \alpha + (n - m) + d. \quad (20)$$

For the corresponding non-elimination curtailed procedure to continue beyond stage m , at least one additional inferior population must remain close enough to the leader so that the stopping rule is not triggered. Let Π_3 denote one such population. We write

$$Z_{3,m} = \alpha + d + j, \quad j = 1, \dots, n - m, \quad (21)$$

so that

$$\alpha + d + 1 \leq Z_{3,m} \leq \alpha + d + (n - m). \quad (22)$$

Now consider the remaining inferior population Π_4 . In the structural approximation, we do not impose the full stage ordering restrictions that would be required by a complete path partition. Instead, we retain only the final tail restriction that the total success count of Π_4 should not exceed the final success count of the representative leader Π_2 . Conditional on $Z_{4,m} = z$, this requirement is

$$X_4 \leq \alpha + (n - m) + d - z. \quad (23)$$

Thus the contribution of Π_4 takes the form

$$\sum_z P(Z_{4,m} = z) P(X_4 \leq \alpha + (n - m) + d - z). \quad (24)$$

Under the fixed size continuation construction, $Z_{4,m}$ and X_4 are independent binomial counts from disjoint blocks of Bernoulli trials with common success probability p . Therefore,

$$Z_{4,m} + X_4 \sim \text{Bin}(n, p), \quad (25)$$

and the inner sum can be written exactly as

$$P(Z_{4,m} + X_4 \leq \alpha + (n - m) + d) = P(\text{Bin}(n, p) \leq \alpha + (n - m) + d). \quad (26)$$

For $k = 4$, the resulting structural expression is

$$\begin{aligned} A_4 = & x \sum_m \sum_{\alpha} P(Z_{1,m} = \alpha) P(X_1 = n - m) P(Z_{2,m} = \alpha + (n - m) + d) P(X_2 = 0) \\ & \times \sum_{j=1}^{n-m} P(Z_{3,m} = \alpha + d + j) P(X_3 \leq n - m - j) \\ & \times P(\text{Bin}(n, p) \leq \alpha + (n - m) + d), \end{aligned} \quad (27)$$

where the summation range of α is restricted by feasibility of the binomial counts. The notation A_4 is used to emphasize that this is a structural analytical expression, not a fully proven upper bound over all possible curtailed sample paths.

This representation suggests a natural extension to general k . Once Π_2 is fixed as the representative leader and Π_3 is used as one population that prevents stopping at stage m , every additional inferior population Π_r contributes through the requirement that its final success total does not exceed that of Π_2 , with $r = 4, \dots, k$. For each such population,

$$\sum_{z_r} P(Z_{r,m} = z_r) P(X_r \leq \alpha + (n - m) + d - z_r) = P(\text{Bin}(n, p) \leq \alpha + (n - m) + d). \quad (28)$$

Since these populations are independent and have the same inferior success probability p , their joint contribution is

$$[P(\text{Bin}(n, p) \leq \alpha + (n - m) + d)]^{k-3}. \quad (29)$$

Accordingly, for general $k \geq 4$, the structural expression becomes

$$\begin{aligned} A_k = & x \sum_m \sum_{\alpha} P(Z_{1,m} = \alpha) P(X_1 = n - m) P(Z_{2,m} = \alpha + (n - m) + d) P(X_2 = 0) \\ & \times \sum_{j=1}^{n-m} P(Z_{3,m} = \alpha + d + j) P(X_3 \leq n - m - j) [P(\text{Bin}(n, p) \leq \alpha + (n - m) + d)]^{k-3}. \end{aligned} \quad (30)$$

Using binomial pmf and cdf notation, this can be written more compactly as

$$\begin{aligned} A_k = & x \sum_m \sum_{\alpha} b_m(\alpha; p + \delta) (p + \delta)^{n-m} \cdot b_m(\alpha + (n - m) + d; p) (1 - p)^{n-m} \\ & \times \sum_{j=1}^{n-m} b_m(\alpha + d + j; p) B_{n-m}(n - m - j; p) [B_n(\alpha + (n - m) + d; p)]^{k-3}, \end{aligned} \quad (31)$$

where

$$b_n(x; p) = \binom{n}{x} p^x (1 - p)^{n-x}, \quad B_n(c; p) = P(\text{Bin}(n, p) \leq c). \quad (32)$$

Because the inferior populations $\Pi_2, \dots, \Pi_k$ are identically distributed with common success probability p , they are exchangeable under this construction. We use the labeling convention (Π_2, Π_3) only as a representative description of the path structure, with Π_2 denoting a current leader and Π_3 denoting one inferior population that keeps the curtailed procedure from stopping at stage m . The remaining inferior populations then contribute through the common final tail restriction in (29). This convention avoids a crude combinatorial multiplier at the approximation stage, since different labelings may overlap or reduce to the same effective tied configuration.

To obtain a more tractable approximation, we focus on the only nontrivial inner sum,

$$S_3(\alpha, m) = \sum_{j=1}^{n-m} P(Z_{3,m} = \alpha + d + j) P(X_3 \leq n - m - j), \quad (33)$$

where $Z_{3,m} \sim \text{Bin}(m, p)$ and $X_3 \sim \text{Bin}(n - m, p)$ are independent. Let $T_3 = Z_{3,m} + X_3$, then $T_3 \sim \text{Bin}(n, p)$, and the event $\{Z_{3,m} = \alpha + d + j, X_3 \leq n - m - j\}$ is equivalent to $\{Z_{3,m} \geq \alpha + d + 1, T_3 \leq \alpha + d + n - m\}$, after summing over j . Hence

$$S_3(\alpha, m) = P(Z_{3,m} \geq \alpha + d + 1, T_3 \leq \alpha + d + n - m). \quad (34)$$

We approximate the bivariate count vector $(Z_{3,m}, T_3)$ by a bivariate normal distribution. The mean vector is

$$\mu = \begin{pmatrix} mp \\ np \end{pmatrix}, \quad (35)$$

and the covariance matrix is

$$\Sigma = \begin{pmatrix} mp(1-p) & mp(1-p) \\ mp(1-p) & np(1-p) \end{pmatrix}. \quad (36)$$

The covariance identity is

$$\text{Cov}(Z_{3,m}, T_3) = \text{Cov}(Z_{3,m}, Z_{3,m}) + \text{Cov}(Z_{3,m}, X_3) = \text{Var}(Z_{3,m}), \quad (37)$$

because $Z_{3,m}$ and X_3 are independent. Therefore the corresponding correlation is

$$\rho = \frac{mp(1-p)}{\sqrt{mp(1-p)}\sqrt{np(1-p)}} = \sqrt{\frac{m}{n}}. \quad (38)$$

Using a continuity correction, the event in (34) is approximated by

$$\{Z_{3,m} \geq \alpha + d + \frac{1}{2}, T_3 \leq \alpha + d + n - m + \frac{1}{2}\}. \quad (39)$$

Define

$$a_{m,\alpha} = \frac{\alpha + d + \frac{1}{2} - mp}{\sqrt{mp(1-p)}}, \quad b_{m,\alpha} = \frac{\alpha + d + n - m + \frac{1}{2} - np}{\sqrt{np(1-p)}}. \quad (40)$$

Then

$$S_3(\alpha, m) \approx \Phi(b_{m,\alpha}) - \Phi_2(a_{m,\alpha}, b_{m,\alpha}; \rho), \quad (41)$$

where Φ is the standard normal cdf and $\Phi_2(\cdot, \cdot; \rho)$ is the cdf of the standard bivariate normal distribution with correlation $\rho = \sqrt{m/n}$.

Substituting this approximation into (31) gives

$$\begin{aligned} \tilde{A}_k = x \sum_{m=\lceil (n+d)/2 \rceil}^{n-1} \sum_{\alpha=0}^{2m-n-d} b_m(\alpha; p+\delta)(p+\delta)^{n-m} \cdot b_m(\alpha+n-m+d; p)(1-p)^{n-m} \\ \times \left[\Phi(b_{m,\alpha}) - \Phi_2(a_{m,\alpha}, b_{m,\alpha}; \sqrt{m/n}) \right] \cdot \left[B_n(\alpha+n-m+d; p) \right]^{k-3}. \end{aligned} \quad (42)$$

The expression $\tilde{A}_k$ is a structural approximation motivated by the $k = 3$ boundary analysis. It is intended to capture the dominant qualitative behavior of the general error mechanism, especially the effect of additional inferior populations through the factor $[B_n(\alpha+n-m+d; p)]^{k-3}$. It is not be interpreted as a rigorously proven upper bound for all erroneous elimination paths when $k \geq 4$.

The simulation results below are used to examine whether the additional error caused by adaptive elimination remains small under the calibrated design parameters (n, d, x) and whether its qualitative behavior is consistent with (42). This replaces the need to state a formal conjecture as a theorem level conclusion in the main text.

3.4.2. Simulation Evidence for General k

We now provide simulation evidence for several values of k and δ to assess whether the qualitative behavior suggested by the structural general approximation is reflected in the simulated elimination-induced error probabilities. Figure 5 presents simulation estimates of $P(\mathcal{E})$ for $k = 4, 5, 6$. Each column corresponds to a different number of populations, and each row corresponds to a different signal level $\delta \in \{0.05, 0.10, 0.15\}$. The red curve represents the simulated mean of $P(\mathcal{E})$, while the dashed blue curves represent the corresponding 95% confidence interval across repeated runs.

The simulation results in Figure 5 suggest several patterns. First, for each fixed (k, δ) , the elimination-induced error probability tends to decrease as n increases. Second, for each fixed k , the magnitude of $P(\mathcal{E})$ becomes smaller as δ increases, indicating that stronger separation of the best population reduces the risk of erroneous elimination. Third, for fixed (n, δ) , the error probability tends to increase moderately with k , reflecting the greater complexity introduced by additional competing populations.

Across all settings shown in the figure, the simulated error probabilities remain on the order of 10^{-3} or smaller. These numerical results suggest that the additional risk introduced by adaptive elimination remains small in absolute magnitude under the configurations considered. They also provide simulation-based support for the qualitative behavior

suggested by the structural general approximation, without treating that approximation as a theorem-level guarantee.

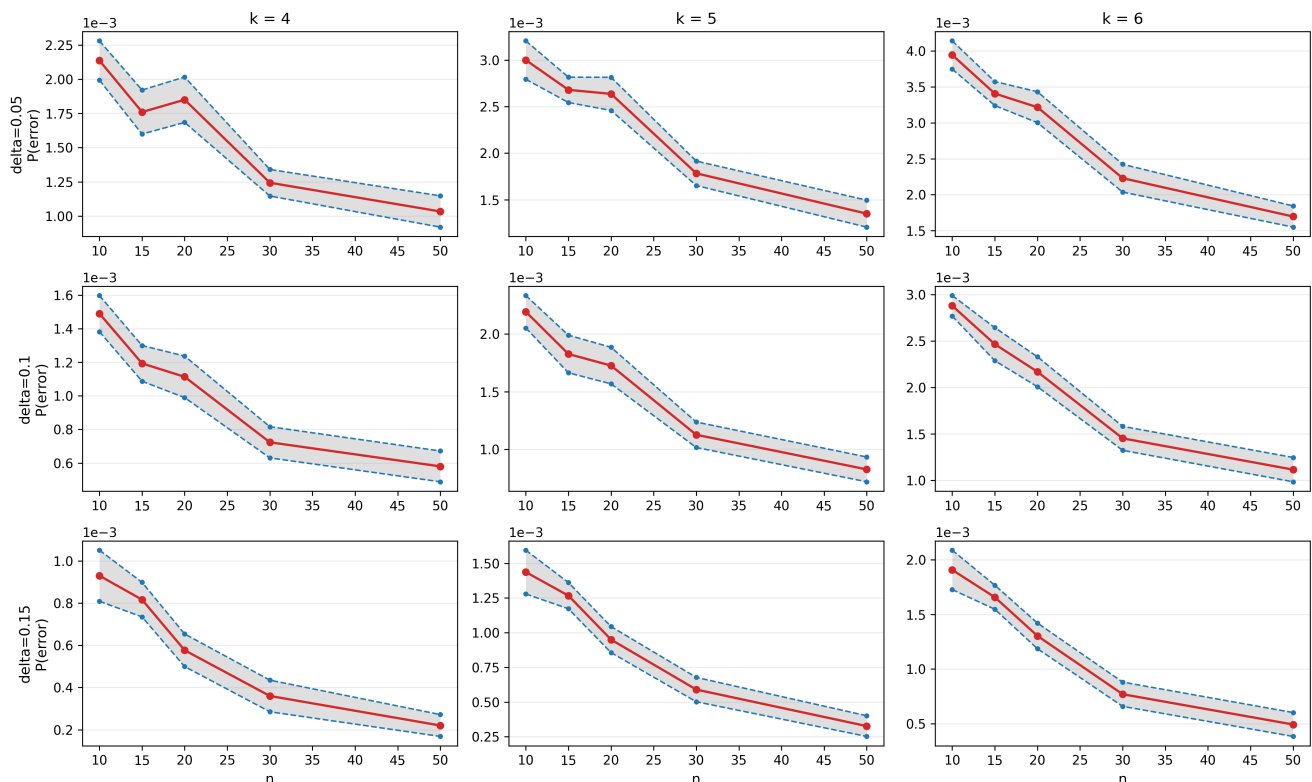


Figure 5. Simulation estimates of the elimination-induced error probability $P(\mathcal{E})$ for general k under adaptive elimination.

The simulations in this subsection are used to assess the scale of the elimination-induced error probability and the qualitative behavior suggested by the structural general approximation. The next section turns to the operating characteristics of the full adaptive procedure, comparing the adaptive elimination refinement with the original randomized curtailed procedure.

4. Simulation Study

4.1. Simulation Setup

The theoretical analysis in the previous section focuses on the elimination-induced error caused by adaptive elimination. In contrast, this section evaluates the operating performance of the full adaptive procedure. The purpose is to compare the original randomized curtailed procedure R_C with its adaptive elimination refinement, referred to as Randomized R_C with elimination.

This comparison is designed to isolate the incremental effect of adding adaptive elimination to the randomized curtailed framework. Therefore, we do not compare against external sequential procedures in this section. The baseline procedure and the adaptive elimination procedure share the same subset selection objective, the same target probability of correct selection, and the same final randomized selection rule. The only difference is that the adaptive version removes populations during sampling when they can no longer reach the current selection boundary.

All tables and figures in this section present author-generated simulation results, and are not reproduced from any external source.

We consider settings with a small or moderate number of populations, specifically $k = 3, 4, 5$. These settings are motivated by multi-arm screening experiments, where several candidate treatments or alternatives are compared under limited sampling resources. Adaptive experimental designs commonly use interim information to modify trial conduct, including stopping or dropping poorly performing arms, but such modifications must be evaluated through their operating characteristics under the planned design rules [2]. This motivates the simulation-based evaluation below. For each value of k , we examine three representative configurations of success probabilities:

1. **Equally spaced probabilities** (Table 5);
2. **One clear best population** (Table 6);
3. **One clear worst population** (Table 7).

These configuration families represent balanced separation, strong dominance of one population, and settings in which weak populations are present but the stopping rule may still be difficult to trigger. For each setting, the target probability of correct selection is fixed at $P^* = 0.95$. The planned maximum sample sizes are $n \in \{10, 15, 20, 30, 50\}$. For each combination of (k, n) and procedure, the selection threshold d and the randomization probability x are calibrated so that the estimated probability of correct selection, denoted by P_{CS} , is close to the target level P^* .

All simulation data are generated from independent Bernoulli observations. For population Π_i , each observation has success probability p_i . Under the original randomized curtailed procedure, one observation is sampled from every population that remains in the experiment at each stage until the stopping rule is triggered or the maximum stage n is reached. Under the adaptive elimination refinement, one observation is sampled from each active population at each stage, and the adaptive elimination rule is applied before checking the stopping rule.

The following performance measures are reported:

- P_{CS} : the estimated probability of correct selection;
- $E(N)$: the expected total number of observations used across all populations;
- $E(S)$: the expected number of populations retained in the final selected subset.

Each reported result is computed from 100 independent repetitions. In each repetition, 10,000 Monte Carlo sample paths are generated to estimate P_{CS} , $E(N)$, and $E(S)$. For each performance measure, the tables report the Monte Carlo mean together with its standard error across the 100 independent repetitions. The simulations were implemented in R using a fixed random seed for reproducibility. The simulation code and generated summary tables are available from the authors upon reasonable request.

4.2. Simulation Results Across Different Configurations

The simulation results are summarized in Tables 5–7. The tables compare Randomized R_C with Randomized R_C with adaptive elimination under the same configuration families and planned sample sizes. Values are reported as Monte Carlo means with standard errors in parentheses. The main goal is to assess whether adaptive elimination reduces $E(N)$ while keeping P_{CS} near the target level and preserving the expected selected subset size $E(S)$.

Equally Spaced Probabilities

Table 5. Equally spaced probabilities: Randomized R_C versus randomized R_C with adaptive elimination. Values are Monte Carlo means with standard errors in parentheses.

k	n	Method	(d, x)	P_{Cs}	$E(N)$	$E(S)$
3	10	R_C	(2, 0.9500)	0.958 (0.0002)	25.47 (0.003)	2.02 (0.001)
		$R_C + AE$	(2, 0.9600)	0.955 (0.0002)	25.28 (0.003)	1.98 (0.001)
3	15	R_C	(3, 0.1500)	0.957 (0.0002)	38.62 (0.005)	1.74 (0.001)
		$R_C + AE$	(3, 0.1000)	0.956 (0.0002)	38.39 (0.004)	1.72 (0.001)
3	20	R_C	(3, 0.1000)	0.955 (0.0002)	51.79 (0.006)	1.52 (0.001)
		$R_C + AE$	(3, 0.1200)	0.956 (0.0002)	51.29 (0.005)	1.52 (0.001)
3	30	R_C	(2, 0.8700)	0.957 (0.0002)	78.11 (0.007)	1.27 (0.000)
		$R_C + AE$	(2, 0.8700)	0.956 (0.0002)	76.36 (0.008)	1.26 (0.001)
3	50	R_C	(1, 0.6900)	0.956 (0.0002)	130.62 (0.011)	1.05 (0.000)
		$R_C + AE$	(1, 0.9900)	0.957 (0.0002)	126.30 (0.009)	1.05 (0.000)
4	10	R_C	(2, 0.8800)	0.957 (0.0002)	34.45 (0.004)	2.13 (0.001)
		$R_C + AE$	(2, 0.9200)	0.957 (0.0002)	33.54 (0.004)	2.10 (0.001)
4	15	R_C	(2, 0.9600)	0.957 (0.0002)	51.98 (0.005)	1.73 (0.001)
		$R_C + AE$	(2, 0.9800)	0.956 (0.0002)	49.96 (0.004)	1.71 (0.001)
4	20	R_C	(2, 0.9000)	0.954 (0.0002)	69.52 (0.007)	1.49 (0.001)
		$R_C + AE$	(2, 0.9800)	0.958 (0.0002)	66.26 (0.006)	1.50 (0.001)
4	30	R_C	(2, 0.6500)	0.956 (0.0002)	104.58 (0.009)	1.24 (0.000)
		$R_C + AE$	(2, 0.6400)	0.956 (0.0002)	98.67 (0.007)	1.24 (0.000)
4	50	R_C	(1, 0.4500)	0.957 (0.0002)	174.51 (0.014)	1.04 (0.000)
		$R_C + AE$	(1, 0.5700)	0.957 (0.0002)	161.76 (0.010)	1.04 (0.000)
5	10	R_C	(2, 0.4900)	0.957 (0.0002)	45.56 (0.004)	2.56 (0.001)
		$R_C + AE$	(2, 0.5200)	0.955 (0.0002)	44.15 (0.003)	2.49 (0.001)
5	15	R_C	(2, 0.5800)	0.956 (0.0002)	68.38 (0.005)	2.06 (0.001)
		$R_C + AE$	(2, 0.6300)	0.957 (0.0002)	65.35 (0.004)	2.04 (0.001)
5	20	R_C	(2, 0.6100)	0.957 (0.0002)	91.23 (0.006)	1.76 (0.001)
		$R_C + AE$	(2, 0.6100)	0.956 (0.0002)	86.40 (0.005)	1.73 (0.001)
5	30	R_C	(2, 0.4500)	0.956 (0.0002)	136.91 (0.008)	1.40 (0.001)
		$R_C + AE$	(2, 0.4400)	0.956 (0.0002)	128.25 (0.007)	1.39 (0.000)
5	50	R_C	(1, 0.7200)	0.955 (0.0002)	228.08 (0.013)	1.10 (0.000)
		$R_C + AE$	(1, 0.9000)	0.957 (0.0002)	209.33 (0.009)	1.11 (0.000)

One Clear Best Population

Table 6. One clear best population: Randomized R_C versus randomized R_C with adaptive elimination. Values are Monte Carlo means with standard errors in parentheses.

k	n	Method	(d, x)	P_{Cs}	$E(N)$	$E(S)$
3	10	R_C	(3, 0.0900)	0.956 (0.0002)	26.39 (0.003)	2.36 (0.001)
		$R_C + AE$	(3, 0.1200)	0.957 (0.0002)	26.35 (0.003)	2.37 (0.001)
3	15	R_C	(3, 0.4400)	0.957 (0.0002)	40.02 (0.004)	2.17 (0.001)
		$R_C + AE$	(3, 0.4200)	0.955 (0.0002)	39.88 (0.004)	2.14 (0.001)
3	20	R_C	(3, 0.5400)	0.957 (0.0002)	53.64 (0.005)	1.96 (0.001)
		$R_C + AE$	(3, 0.5600)	0.957 (0.0002)	53.39 (0.004)	1.95 (0.001)
3	30	R_C	(3, 0.4100)	0.957 (0.0002)	80.82 (0.006)	1.60 (0.001)
		$R_C + AE$	(3, 0.4200)	0.957 (0.0002)	80.29 (0.006)	1.59 (0.001)
3	50	R_C	(2, 0.5400)	0.957 (0.0002)	134.76 (0.009)	1.18 (0.000)
		$R_C + AE$	(2, 0.4600)	0.954 (0.0002)	133.16 (0.008)	1.17 (0.000)

Table 6. Cont.

k	n	Method	(d, x)	P_{CS}	$E(N)$	$E(S)$
4	10	R_C	(1, 0.9700)	0.954 (0.0002)	33.88 (0.004)	1.53 (0.001)
		$R_C + AE$	(2, 0.0100)	0.957 (0.0002)	33.35 (0.003)	1.55 (0.001)
4	15	R_C	(1, 0.5300)	0.955 (0.0002)	50.34 (0.005)	1.15 (0.000)
		$R_C + AE$	(1, 0.7500)	0.957 (0.0002)	48.43 (0.005)	1.16 (0.000)
4	20	R_C	(1, 0.0000)	0.966 (0.0002)	66.51 (0.006)	1.02 (0.000)
		$R_C + AE$	(1, 0.0000)	0.965 (0.0002)	64.04 (0.006)	1.02 (0.000)
4	30	R_C	(1, 0.0000)	0.990 (0.0001)	98.39 (0.008)	1.01 (0.000)
		$R_C + AE$	(1, 0.0000)	0.990 (0.0001)	94.96 (0.008)	1.01 (0.000)
4	50	R_C	(1, 0.0000)	0.999 (0.0000)	161.31 (0.013)	1.00 (0.000)
		$R_C + AE$	(1, 0.0000)	0.999 (0.0000)	156.38 (0.009)	1.00 (0.000)
5	10	R_C	(1, 0.0000)	0.996 (0.0001)	37.43 (0.004)	1.01 (0.000)
		$R_C + AE$	(1, 0.0000)	0.996 (0.0001)	35.15 (0.003)	1.01 (0.000)
5	15	R_C	(1, 0.0000)	0.999 (0.0000)	54.77 (0.005)	1.00 (0.000)
		$R_C + AE$	(1, 0.0000)	0.999 (0.0000)	51.23 (0.003)	1.00 (0.000)
5	20	R_C	(1, 0.0000)	1.000 (0.0000)	72.01 (0.006)	1.00 (0.000)
		$R_C + AE$	(1, 0.0000)	1.000 (0.0000)	67.30 (0.003)	1.00 (0.000)
5	30	R_C	(1, 0.0000)	1.000 (0.0000)	106.46 (0.007)	1.00 (0.000)
		$R_C + AE$	(1, 0.0000)	1.000 (0.0000)	99.36 (0.005)	1.00 (0.000)
5	50	R_C	(1, 0.0000)	1.000 (0.0000)	175.20 (0.010)	1.00 (0.000)
		$R_C + AE$	(1, 0.0000)	1.000 (0.0000)	163.38 (0.006)	1.00 (0.000)

One Clear Worst Population

Table 7. One clear worst population: Randomized R_C versus randomized R_C with adaptive elimination. Values are Monte Carlo means with standard errors in parentheses.

k	n	Method	(d, x)	P_{CS}	$E(N)$	$E(S)$
3	10	R_C	(3, 0.4500)	0.956 (0.0002)	26.57 (0.002)	2.49 (0.001)
		$R_C + AE$	(3, 0.4700)	0.955 (0.0002)	26.50 (0.003)	2.47 (0.001)
3	15	R_C	(3, 0.9200)	0.958 (0.0002)	40.41 (0.003)	2.35 (0.001)
		$R_C + AE$	(3, 0.9300)	0.956 (0.0002)	40.15 (0.003)	2.32 (0.001)
3	20	R_C	(4, 0.3700)	0.957 (0.0002)	54.35 (0.004)	2.24 (0.001)
		$R_C + AE$	(4, 0.3600)	0.956 (0.0002)	54.08 (0.003)	2.23 (0.001)
3	30	R_C	(4, 0.9100)	0.956 (0.0002)	82.44 (0.005)	2.04 (0.001)
		$R_C + AE$	(4, 0.9800)	0.958 (0.0002)	81.63 (0.005)	2.05 (0.001)
3	50	R_C	(5, 0.7300)	0.957 (0.0002)	139.11 (0.008)	1.82 (0.001)
		$R_C + AE$	(5, 0.7000)	0.956 (0.0002)	137.16 (0.007)	1.81 (0.001)
4	10	R_C	(2, 0.9500)	0.957 (0.0002)	36.39 (0.003)	2.72 (0.001)
		$R_C + AE$	(3, 0.0000)	0.962 (0.0002)	36.02 (0.003)	2.76 (0.001)
4	15	R_C	(3, 0.4200)	0.954 (0.0002)	55.03 (0.004)	2.50 (0.001)
		$R_C + AE$	(3, 0.4800)	0.956 (0.0002)	53.90 (0.004)	2.49 (0.001)
4	20	R_C	(3, 0.7500)	0.957 (0.0002)	73.82 (0.004)	2.34 (0.001)
		$R_C + AE$	(3, 0.7500)	0.956 (0.0002)	71.69 (0.004)	2.31 (0.001)
4	30	R_C	(4, 0.1100)	0.956 (0.0002)	111.57 (0.006)	2.08 (0.001)
		$R_C + AE$	(4, 0.1200)	0.956 (0.0002)	108.08 (0.005)	2.08 (0.001)
4	50	R_C	(4, 0.6700)	0.957 (0.0002)	187.48 (0.009)	1.81 (0.001)
		$R_C + AE$	(4, 0.6200)	0.956 (0.0002)	178.97 (0.007)	1.79 (0.001)

Table 7. Cont.

k	n	Method	(d, x)	P_{CS}	$E(N)$	$E(S)$
5	10	R_C	(2, 0.8000)	0.956 (0.0002)	46.75 (0.003)	3.23 (0.001)
		$R_C + AE$	(3, 0.0000)	0.975 (0.0001)	45.35 (0.003)	3.39 (0.001)
5	15	R_C	(3, 0.0500)	0.956 (0.0002)	70.39 (0.004)	2.95 (0.001)
		$R_C + AE$	(3, 0.1000)	0.958 (0.0002)	67.22 (0.003)	2.97 (0.001)
5	20	R_C	(3, 0.4100)	0.957 (0.0002)	94.17 (0.005)	2.81 (0.001)
		$R_C + AE$	(3, 0.4500)	0.957 (0.0002)	89.09 (0.003)	2.79 (0.001)
5	30	R_C	(3, 0.6600)	0.954 (0.0002)	141.83 (0.007)	2.44 (0.001)
		$R_C + AE$	(3, 0.7400)	0.956 (0.0002)	132.74 (0.005)	2.43 (0.001)
5	50	R_C	(3, 0.9100)	0.956 (0.0002)	237.33 (0.009)	1.96 (0.001)
		$R_C + AE$	(3, 0.9200)	0.955 (0.0002)	219.79 (0.007)	1.94 (0.001)

In some strongly separated configurations, the calibrated values are $d = 1$ and $x = 0$. As discussed in our earlier work [9], randomization is useful only when $P(CS \mid d - 1) < P^* < P(CS \mid d)$, in which case the inclusion probability x interpolates between two adjacent integer thresholds to meet the target. Here, however, $d = 1$ already yields $P(CS \mid d = 1) \geq P^*$ over the whole parameter range, so the target is met without any randomization and $x = 0$. In this regime $d = 1$ serves to keep the selected subset as small as possible, while randomization is simply not required. This is therefore a boundary case of the same selection rule, not a different selection objective.

Overall Summary and Trade-Off

Table 5 reports configurations in which the success probabilities are approximately evenly separated. For every displayed k and n , the adaptive elimination procedure uses a smaller expected total sample size than the original randomized curtailed procedure, and the gap widens as k or n grows; for $k = 5$ and $n = 50$, $E(N)$ decreases from 228.08 to 209.33. The estimated P_{CS} stays close to the target in most settings, with small deviations arising from the simulation-based calibration of (d, x) . The expected subset size $E(S)$ is similar between the two procedures, indicating that adaptive elimination mainly affects the sampling burden and leaves the final selection behavior largely unchanged.

Table 6 reports configurations in which one population has a noticeably larger success probability than the others. Here the curtailed stopping rule is already effective, since a dominant population usually separates early, yet adaptive elimination still reduces $E(N)$ in every displayed setting. The reduction is clearest for $k = 5$: when $n = 50$, $E(N)$ falls from 175.20 to 163.38. In these cases $E(S)$ stays essentially equal to one and P_{CS} is close to one, so elimination removes noncompetitive populations without altering the final decision in most simulated paths.

Table 7 reports configurations containing one clearly weak population. These are the most favorable settings for adaptive elimination because a poor population may become unable to reach the selection boundary well before the experiment stops. The procedure reduces $E(N)$ throughout, most strongly when $k = 5$ and n is large; for $k = 5$ and $n = 50$, $E(N)$ falls from 237.33 to 219.79. The subset size shifts slightly with the calibrated (d, x) but stays of the same general magnitude, and P_{CS} remains close to the target apart from a few configurations that deviate from 0.95 under the discrete, simulation-based calibration.

Across all three families, the consistent effect of adaptive elimination is a reduction in $E(N)$, growing with k and with the planned horizon as more populations and stages create more chances to remove weak arms before global stopping. The effects on P_{CS} and $E(S)$ are more variable: P_{CS} sits near the target $P^* = 0.95$ in most settings and modestly above or below it in others, while $E(S)$ shifts slightly where the calibrated d and x differ between the two procedures. This is the central trade-off—the saving in sampling effort is not

free, since (d, x) must be recalibrated and P_{CS} and $E(S)$ can then drift from their baseline values. The main and most reliable benefit is therefore the reduction in $E(N)$, with selection accuracy and final subset size broadly comparable to those of the original procedure.

5. Discussion

The results of this paper show that adaptive elimination can serve as a useful refinement of the curtailed procedure R_C for binomial randomized subset selection. Its main value lies in improving the efficiency of the sampling process when the original stopping rule is difficult to satisfy. This happens most naturally in configurations where several relatively strong populations remain close to one another, while one or more clearly inferior populations continue to consume observations. In such settings, the adaptive rule helps shrink the active set before terminal stopping becomes available, and thereby reduces unnecessary sampling without altering the basic structure of the curtailed framework. The proposed refinement therefore addresses scenarios where early stopping alone fails to fully use the information generated during sequential sampling.

A central issue is whether this additional flexibility comes at a meaningful cost in accuracy. For this reason, the present paper focused not on all possible incorrect final decisions, but on the extra risk introduced specifically by elimination. This distinction is important. The curtailed procedure already has its own calibrated probability of correct selection, determined through the original design parameters (n, d, x) . Adaptive elimination is introduced afterward as a refinement. From this perspective, the relevant question is whether the elimination step itself introduces substantial new error when the original calibrated parameters are retained. Both the analytical bounds and the simulation results indicate that this incremental risk is small under the settings examined, which supports the practical use of adaptive elimination when sampling efficiency is an important concern.

It is also useful to clarify why the elimination rule uses " $\leq$ " rather than the stricter " $<$ ". A population is eligible for elimination when its largest possible future success count cannot reach the current selection boundary. Under the randomized curtailed procedure, a population that lands exactly on the boundary is not selected with certainty, but only with the calibrated probability x . The choice of " $\leq$ " treats such a population as noncompetitive for sampling, which is what makes the additional error possible. The best population may be eliminated early after a slow start, even though, had sampling continued, it could in principle have succeeded in all remaining observations and then been retained at the final randomized step. The theoretical analysis in this paper is precisely an attempt to show that the probability of this event is small, so that the more convenient " $\leq$ " rule can be used without a meaningful loss in accuracy.

The main theoretical limitation of the present work lies in the general case. As the number of populations grows, exact analysis becomes increasingly difficult because of the dependencies between the current leader, the populations that prevent termination, and the downstream evolution of the process under the baseline rule. For this reason, the general k results in this paper are intentionally conservative. They are designed to show that the additional elimination risk remains controlled, not to provide the sharpest possible expression. In this sense, the present work should be viewed as establishing the feasibility and stability of adaptive elimination, rather than as closing all theoretical questions about its exact operating characteristics.

Another point worth emphasizing is that adaptive elimination is not uniformly superior in every experimental configuration. Its usefulness depends on the structure of the underlying success probabilities. When one population becomes dominant early, the original curtailed stopping rule may already be sufficient, leaving less room for elimination to add benefit. By contrast, when the competition among the leading populations remains

close but weak populations are clearly present, adaptive elimination can noticeably improve efficiency. This configuration sensitive behavior is consistent with the broader view of R_C , randomization, and elimination as a sequence of refinements rather than a sequence of universally dominating replacements.

Although this paper has been developed at the methodological level, the adaptive logic behind elimination suggests several natural directions for future application. In multi-arm and platform type clinical settings, adaptive designs are increasingly used to allow prospectively planned modifications based on interim evidence, including arm-level decisions and other changes intended to improve efficiency [2–5,13]. A related operational intuition also appears in modern online experimentation. In digital advertising and creative testing, adaptive experimentation systems may reduce traffic to poorly performing variants as information accumulates, rather than continuing to allocate exposure uniformly throughout a fixed-horizon test [14,15]. The statistical objectives in those settings differ from subset selection, and the relevant optimization criteria are not the same as controlling the probability of correct selection P_{CS} . Even so, they reflect a similar practical principle that once accumulating evidence indicates that some options are clearly inferior, continued resource allocation to those options becomes increasingly difficult to justify. It would therefore be of interest to study whether the elimination logic developed here can be adapted more explicitly to settings involving online binary response experimentation, creative screening, or other sequential decision systems with limited experimental budgets.

Several methodological extensions also remain open. One natural development is to strengthen the theory for general k , either through sharper approximations or through more computationally efficient bounding strategies. Another is to extend the framework beyond binomial outcomes. A multinomial version would not be a direct substitution because each population would then be represented by a vector of category counts rather than a single success count, so the selection boundary, the adaptive elimination rule, and the definition of correct selection would all need to be redefined, possibly through an ordered category score, a utility function, or a target response category. It would also be useful to implement the procedure with additional structural information, such as control group settings, heterogeneous populations, stratified designs, or prior information on success probabilities. These directions would move the current refinement from a modular add-on toward a more fully integrated adaptive subset selection framework.

Overall, the present paper supports adaptive elimination as a statistically stable and practically meaningful refinement of the original curtailed procedure R_C [9]. It is most effective not as a replacement for curtailment, but as a complementary mechanism that removes populations once they are no longer competitive for selection. Viewed in this way, adaptive elimination extends the original curtailed procedure in a natural direction where it improves the efficiency of sequential sampling when early stopping alone is not enough, while introducing only small additional elimination-induced risk under the settings studied.

Author Contributions: Conceptualization, Y.Z.; methodology, Y.Z.; validation, Y.Z.; formal analysis, Y.Z.; investigation, Y.Z.; resources, Y.Z.; writing—original draft, Y.Z.; writing—review and editing, Y.Z. and P.C.; visualization, Y.Z.; supervision, P.C.; project administration, Y.Z. and P.C. All authors have read and agreed to the published version of the manuscript.

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.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Gupta, S.S.; Sobel, M. Selecting a Subset Containing the Best of Several Binomial Populations. In *Contributions to Probability and Statistics*; Stanford University Press: Stanford, CA, USA, 1960; Chapter 20, pp. 224–248.
2. U.S. Food and Drug Administration. *Adaptive Designs for Clinical Trials of Drugs and Biologics: Guidance for Industry*; Guidance for Industry; U.S. Food and Drug Administration: Silver Spring, MD, USA, 2019.
3. Pallmann, P.; Bedding, A.W.; Choodari-Oskooei, B.; Dimairo, M.; Flight, L.; Hampson, L.V.; Holmes, J.; Mander, A.P.; Odondi, L.; Sydes, M.R.; et al. Adaptive designs in clinical trials: Why use them, and how to run and report them. *BMC Med.* **2018**, *16*, 29. [[CrossRef](#)] [[PubMed](#)]
4. Burnett, T.; König, F.; Jaki, T. Adding experimental treatment arms to Multi-Arm Multi-Stage platform trials in progress. *arXiv* **2020**, arXiv:2007.04951.
5. Greenstreet, P.; Jaki, T.; Bedding, A.; Harbron, C.; Mozgunov, P. A multi-arm multi-stage platform design that allows pre-planned addition of arms while still controlling the family-wise error. *arXiv* **2021**, arXiv:2112.06195.
6. Sydes, M.R.; Parmar, M.K.B.; Mason, M.D.; Clarke, N.W.; Amos, C.; Anderson, J.; de Bono, J.; Dearnaley, D.P.; Dwyer, J.; Green, C.; et al. Flexible Trial Design in Practice: Stopping Arms for Lack-of-Benefit and Adding Research Arms Mid-Trial in STAMPEDE: A Multi-Arm Multi-Stage Randomized Controlled Trial. *Trials* **2012**, *13*, 168. [[CrossRef](#)] [[PubMed](#)]
7. James, N.D.; de Bono, J.S.; Spears, M.R.; Clarke, N.W.; Mason, M.D.; Dearnaley, D.P.; Ritchie, A.W.; Amos, C.L.; Gilson, C.; Jones, R.J.; et al. Abiraterone for prostate cancer not previously treated with hormone therapy. *N. Engl. J. Med.* **2016**, *377*, 338–351.
8. Villar, S.S.; Bowden, J.; Wason, J. Multi-armed bandit models for the optimal design of clinical trials: Benefits and challenges. *Stat. Sci.* **2015**, *30*, 199–215. [[CrossRef](#)] [[PubMed](#)]
9. Zhang, Y.; Chen, P. Curtailed Procedures for Binomial Random-Sized Subset Selection. *J. Biopharm. Stat.* **2025**, advance online publication. [[CrossRef](#)] [[PubMed](#)]
10. Sanchez, S.M. Contributions to the Bernoulli Selection Problem. Ph.D. Thesis, Cornell University, Ithaca, NY, USA, 1986.
11. Sanchez, S.M. A Modified Least-Failure Sampling Procedure for Bernoulli Subset Selection. *Commun. Stat. Methods* **1987**, *16*, 3609–3629. [[CrossRef](#)]
12. Sanchez, S.M. Small-Sample Performance of a Modified Least-Failures Sampling Procedure for Bernoulli Subset Selection. *Commun. Stat. Comput.* **1987**, *16*, 1051–1065. [[CrossRef](#)]
13. U.S. Food and Drug Administration. *Master Protocols: Efficient Clinical Trial Design Strategies to Expedite Development of Oncology Drugs and Biologics: Guidance for Industry*; Guidance for Industry; U.S. Food and Drug Administration: Silver Spring, MD, USA, 2022.
14. Geng, T.; Li, X.; Nair, H.S.; Hao, J.; Xiang, B.; Fan, S. Comparison Lift: Bandit-based Experimentation System for Online Advertising. *Proc. AAAI Conf. Artif. Intell.* **2021**, *35*, 15117–15125. [[CrossRef](#)]
15. Wang, Z.; Carrion, C.; Lin, X.; Ji, F.; Bao, Y.; Yan, W. Adaptive Experimentation with Delayed Binary Feedback. In Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Lyon, France, 25–29 April 2022; pp. 4014–4022. [[CrossRef](#)]

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.