Subset Selection with Curtailment Among Treatments with Two Binary Endpoints in Comparison with a Control

Chishu Yin *, Elena Buzaianu, Pinyuen Chen, Lifang Hsu

Posted Date: 3 July 2025

doi: 10.20944/preprints202507.0255.v1

Keywords: curtailment; sample size; subset selection; two binary endpoints

 doi:10.20944/preprints202507.0255.v1

Article

Subset Selection with Curtailment Among Treatments with Two Binary Endpoints in Comparison with a Control

Chishu Yin [1,*], Elena M. Buzaianu [2], Pinyuen Chen [1] and Lifang Hsu [3]

[1] Department of Mathematics, Syracuse University, Syracuse, NY 13244, USA
[2] Department of Mathematics and Statistics, University of North Florida, Jacksonville, FL 32224, USA
[3] Department of Mathematics, Le Moyne College, Syracuse, NY 13214, USA
* Correspondence: cyin05@syr.edu

Abstract

This paper proposes a closed adaptive sequential procedure for selecting a random-sized subset of size $t(>0)$among $k(\geq t)$ experimental treatments so that the selected subset contains all treatments superior to the control treatment. All the experimental treatments and the control are assumed to produce two binary endpoints, and the procedure is based on those two binary endpoints. A treatment is considered superior if its both endpoints are larger than those of the control. While responses across treatments are assumed to be independent, dependence between endpoints within each treatment is allowed and modeled via an odds ratio. The proposed procedure comprises explicit sampling, stopping, and decision rules. We show that for any sample size n and any parameter configuration, the sequential procedure maintains the same probability of correct selection as the corresponding fixed-sample-size procedure. We use the bivariate binomial and multinomial distributions in the computation and derive design parameters under three scenarios: (i) independent endpoints, (ii) dependent endpoints with known association, and (iii) dependent endpoints with unknown association. We provide tables with the sample size savings achieved by the proposed procedure compared to its fixed-sample-size counterpart. Examples are given to illustrate the procedure.

Keywords: curtailment; sample size; subset selection; two binary endpoints

1. Background, Introduction, and Motivation

In Phase II clinical trials, new treatments are evaluated with respect to both efficacy and safety. We propose a *closed adaptive sequential procedure* for comparing k (>1) experimental treatments against a control treatment. Each treatment, including the control, is assessed using two binary endpoints: one for efficacy and one for safety. An experimental treatment is considered superior to the control if it demonstrates higher success probabilities in both endpoints. Let the control treatment be denoted by π_0, and the k experimental treatments by $\pi_1, \pi_2, \ldots, \pi_k$. The outcomes for each treatment consist of two binary endpoints that are modeled marginally as Bernoulli random variables with unknown success probabilities. Specifically, for the control, the success probabilities are denoted by $(p_{0,1}, p_{0,2})$, and for each experimental treatment π_i $(i = 1, \ldots, k)$, the corresponding success probabilities are $(p_{i,1}, p_{i,2})$. Comparisons are made between each experimental treatment and the control by evaluating whether $p_{i,1} > p_{0,1}$ and $p_{i,2} > p_{0,2}$. The proposed procedure incorporates *curtailment*, allowing for early termination of sampling when sufficient evidence has been gathered to make a decision, thus potentially stopping before reaching the maximum sample size N. The term *closed* refers to the presence of this upper limit N, while the procedure is *adaptive* in the sense that the decision to continue sampling depends dynamically on the outcomes observed thus far.

The problem of selecting the best among k $(k>1)$ treatments with Bernoulli outcomes, or comparing k Bernoulli experiments with each other and with a control or standard, has a long-standing history

 doi:10.20944/preprints202507.0255.v1

in selection theory, particularly with applications in medical trials and the pharmaceutical industry. Sobel and Huyett (1957) proposed a fixed-sample-size procedure for identifying the best Bernoulli population based on the Indifference Zone approach. Later, Gupta and Sobel (1958) introduced a Subset Selection method to select a group that includes the best Bernoulli population. Dunnett (1984) and Thall, Simon, and Ellenberg (1988, 1989) also employed the Indifference Zone framework to select the best among k Bernoulli experiments and compare the selected treatment with a control or standard. Notably, all of these works considered only a single Bernoulli endpoint and were formulated under fixed-sample-size, non-curtailed sampling schemes.

In contrast, this paper investigates a curtailed selection procedure involving treatments with two Bernoulli endpoints. In a curtailed procedure, the experimenter (1) sets an upper limit N on the number of observations per treatment, and (2) continues sampling sequentially from each treatment until either there is sufficient evidence that a treatment is no longer a contender, or the maximum sample size N is reached. The notion of a "contending treatment" is formally defined in Procedure **R** (Section 3). This early stopping mechanism allows for a potential reduction in the total number of observations required.

Curtailment has previously been applied in clinical trials with Bernoulli outcomes, primarily in the context of hypothesis testing (e.g., Carsten and Chen, 2016) and selection procedures (e.g., Bechhofer and Kulkarni, 1982; Jennison, 1983; Buzaianu and Chen, 2008). However, these works are limited to the case of a single Bernoulli endpoint. The procedure proposed in this paper extends the concept of curtailment to treatment comparisons involving two Bernoulli endpoints. For related work, see Jennison and Turnbull (1993) for normally distributed outcomes, and Bryant and Day (1995), Conway and Petroni (1995), and Chen and Chi (2012) for designs with binary outcomes.

More recently, Buzaianu et al (2022) discussed a curtailed procedure for subset selection involving two binary Bernoulli endpoints. However, their approach compares each experimental treatment to a well-established *standard treatment*. This design is most appropriate when a widely accepted reference treatment exists. In contrast, our procedure compares new treatments against a *control treatment*, which may or may not be a recognized standard.

There are many situations in which our approach is more applicable. For example, in the absence of a universally accepted standard treatment—such as when placebo is the only baseline option—it becomes necessary to evaluate new treatments in relation to a control. Similarly, even when a standard treatment exists, it may have been validated only in limited populations (e.g., specific age groups, races, or genders). In such cases, it is important to assess whether the standard treatment continues to perform well in broader or different patient populations. Our design, which explicitly includes the control treatment in the experiment, enables such comparisons and provides a flexible and inclusive framework for decision-making.

This paper addresses the two-endpoint problem using a subset selection approach. We introduce a curtailed, closed sequential procedure in which the total number of observations drawn from each of the $k+1$ contending populations (including the control) is a bounded random variable. We assume that the time between treatment administration and observation of the response is short relative to the overall duration of the experiment. The proposed curtailed procedure uses the fixed-sample-size method as a reference, which will be described in the following sections. We demonstrate that this closed sequential procedure maintains the same probability of correct selection as the fixed-sample-size procedure, while reducing the number of observations drawn from inferior treatments. Section 2 outlines the assumptions, objectives, and probability requirements for two-endpoint clinical trials. Section 3 presents the fixed-sample-size procedure that serves as the benchmark for evaluating the performance of the proposed curtailed procedure.

In Section 4, we propose a sequential selection procedure with curtailment to achieve our objective. We show that the probability of correct selection under the proposed procedure is equal to that of the corresponding fixed-sample-size procedure, uniformly over the parameter space. In Section 5, we evaluate the performance of the proposed curtailed procedure in comparison to its non-curtailed

 doi:10.20944/preprints202507.0255.v1

counterpart, with respect to expected sample size. Section 6 provides two numerical examples to illustrate the application of the proposed method. Finally, concluding remarks are presented in Section 7.

2. Assumptions, Goal, and Probability Requirements

Suppose that n independent subjects are assigned to a treatment, and that two binary endpoints—typically representing therapeutic efficacy ("response") and safety ("nontoxicity")—are observed for each subject. Following the notation of Conway and Petroni (1995), let X_{ij} denote the number of subjects classified as outcome i on the first endpoint and outcome j on the second endpoint, where $i = 1, 2$ and $j = 1, 2$, with 1 representing "success" and 2 representing "failure." The resulting data can be summarized in a 2×2 contingency table (see Table 1).

We assume that the random vector $X = (X_{11}, X_{12}, X_{21}, X_{22})$ follows a multinomial distribution with probabilities $P = (p_{11}, p_{12}, p_{21}, p_{22})$, where:

- p_{11} is the probability of success on both endpoints,
- p_{12} is the probability of success on endpoint 1 and failure on endpoint 2,
- p_{21} is the probability of failure on endpoint 1 and success on endpoint 2,
- p_{22} is the probability of failure on both endpoints.

Let $X_1 = X_{11} + X_{12}$ and $X_2 = X_{11} + X_{21}$ represent the number of successes on endpoints 1 and 2, respectively. The marginal probabilities of success are given by $p_1 = p_{11} + p_{12}$ and $p_2 = p_{11} + p_{21}$, respectively. Consequently, $X_1 \sim \text{Binomial}(n, p_1)$ and $X_2 \sim \text{Binomial}(n, p_2)$. We denote the binomial probability mass function with parameters n and p by $b(n, p, \cdot)$.

The joint distribution of (X_1, X_2) depends not only on p_1 and p_2, but also on the association between the two endpoints. To quantify this association, we use the odds ratio $\phi = \frac{p_{11}p_{22}}{p_{21}p_{12}}$, which is a natural and widely used measure in 2×2 tables. Notably, ϕ is independent of the marginal probabilities p_1 and p_2. When $\phi = 1$, the two endpoints are independent; $\phi > 1$ indicates a positive association, and $\phi < 1$ indicates a negative association.

Table 1. Classification Table.

	Second Endpoint		
First Endpoint	1	2	
1	X_{11}	X_{12}	X_1
2	X_{21}	X_{22}	$X_{\bar{1}}$
	X_2	$X_{\bar{2}}$	n

In this paper, we compare the two binary endpoints of each experimental treatment to those of a control treatment, whose success probabilities on the efficacy and safety endpoints are denoted by $p_{0,1}$ and $p_{0,2}$, respectively. Let π_i, for $i = 1, \ldots, k$, represent the k experimental treatments under investigation, and let π_0 denote the control treatment. Each treatment is associated with two binary outcomes. To distinguish between the two endpoints within each treatment, we use a second subscript j in the notation, where $j = 1$ corresponds to the efficacy endpoint and $j = 2$ to the safety endpoint. Thus, the success probabilities for treatment π_i are denoted by $p_{i,1}$ and $p_{i,2}$ for the efficacy and safety endpoints, respectively, for $i = 0, 1, \ldots, k$. We assume that the $k + 1$ treatments are mutually independent, meaning that responses across different treatments are independent. However, responses within a single treatment may exhibit association between the two endpoints. To classify treatments based on their performance, we partition the parameter space $\{(p_{1,1}, p_{1,2}), \ldots, (p_{k,1}, p_{k,2}) \mid 0 < p_{i,j} < 1 \text{ for } i = 1, \ldots, k,\ j = 1, 2\}$ using four

 doi:10.20944/preprints202507.0255.v1

prespecified constants: $\delta^*_{0,1}, \delta^*_{0,2}, \delta^*_{1,1}$, and $\delta^*_{1,2}$. These constants satisfy the conditions $-\infty < \delta^*_{0,1} < \delta^*_{1,1}$ with $\delta^*_{1,1} > 0$, and $-\infty < \delta^*_{0,2} < \delta^*_{1,2}$ with $\delta^*_{1,2} > 0$. In this framework, a treatment π_i is considered *ineffective* if $p_{i,1} \leq p_{0,1} + \delta^*_{0,1}$ or $p_{i,2} \leq p_{0,2} + \delta^*_{0,2}$, and considered *effective* if $p_{i,1} \geq p_{0,1} + \delta^*_{1,1}$ and $p_{i,2} \geq p_{0,2} + \delta^*_{1,2}$, where $p_{0,1}$ and $p_{0,2}$ are success probabilities of the control treatment, and we assume that these two probabilities are known prior to conducting the selection procedure. Our objective is to classify the k experimental treatments into two groups: those that are *effective* and those that are *ineffective*. We now describe the formal selection goal.

Our Goal: Select a subset consisting of those treatments π_i for which $p_{i,1} > p_{0,1}$ and $p_{i,2} > p_{0,2}$; that is, include all experimental treatments that demonstrate superiority over the control treatment with respect to both efficacy and safety. If no such treatment exists—i.e., if no π_i satisfies both $p_{i,1} > p_{0,1}$ and $p_{i,2} > p_{0,2}$—then none of the k experimental treatments should be selected.

Our probability requirements: Let P^*_0 and P^*_1 be pre-specified probability constants satisfying $2^{-k} < P^*_0 < 1$ and $(1 - 2^{-k})/k < P^*_1 < 1$. The probability requirements for the selection procedure are defined as follows:

$$P\big(\text{All } \pi_i \text{ with } p_{i,1} \geq p_{0,1} + \delta^*_{1,1} \text{ and } p_{i,2} \geq p_{0,2} + \delta^*_{1,2} \\ \text{are included in the selected subset, for } i = 1, \ldots, k\big) \geq P^*_1, \tag{1}$$

and

$$P\big(\text{No } \pi_i \text{ is selected whenever } p_{i,1} \leq p_{0,1} + \delta^*_{0,1} \text{ or } p_{i,2} \leq p_{0,2} + \delta^*_{0,2} \\ \text{for all } i = 1, \ldots, k\big) \geq P^*_0. \tag{2}$$

Let CS_1 denote the event that the selected subset correctly includes all effective treatments, provided such treatments exist. Specifically, CS_1 occurs when every treatment π_i satisfying $p_{i,1} \geq p_{0,1} + \delta^*_{1,1}$ and $p_{i,2} \geq p_{0,2} + \delta^*_{1,2}$ is included in the selected subset. Similarly, let CS_0 denote the event that no treatment is selected when none are truly effective. That is, CS_0 occurs if $p_{i,1} \leq p_{0,1} + \delta^*_{0,1}$ or $p_{i,2} \leq p_{0,2} + \delta^*_{0,2}$ holds for all $i = 1, \ldots, k$. Under this framework, the selection procedure is required to satisfy the following probability criteria:

$$P(CS_1) \geq P^*_1,$$
$$P(CS_0) \geq P^*_0,$$

where P^*_1 and P^*_0 are prespecified thresholds that represent the minimum acceptable probabilities for correctly identifying effective treatments and correctly excluding ineffective treatments, respectively.

Remark 2.1: When effective experimental treatments exist, a correct selection is made if the selected subset includes all such effective treatments. The rationale for selecting a subset—rather than identifying a single best treatment—is that no natural ordering can be established among the pairs of success probabilities $(p_{1,1}, p_{1,2}), \ldots, (p_{k,1}, p_{k,2})$ unless one endpoint is explicitly prioritized over the other. Since this paper does not assume any preference between the two endpoints, we adopt a subset selection approach.

3. Fixed Sample Size Procedure

In this section, we first present the fixed-sample-size selection procedure, which serves as a reference for the curtailed procedure introduced in Section 4. This fixed-sample-size procedure was derived by Buzaianu et al. (2025). We also include results related to the derivation of the design parameters that

 doi:10.20944/preprints202507.0255.v1

ensure the fixed-sample-size procedure satisfies the probability requirements stated in Conditions 2.1 and 2.2.
For prespecified design parameters n, c_1, c_2 , the selection procedure is defined as follows:

Procedure H:
Take n observations from each of the k Bernoulli experimental treatments and the control treatment. Let $X_{i,1}$ and $X_{i,2}$ be the numbers of successes from the first and second endpoints of treatment $i, i = 0, 1, 2, \ldots, k$. For positive integers c_1 and c_2, Procedure **H** is defined as follows:
(1) Include in the subset all the treatments π_i with $X_{i,1} - X_{0,1} \geq c_1$ and $X_{i,2} - X_{0,2} \geq c_2$;
(2) If there is no treatment π_i so that $X_{i,1} - X_{0,1} \geq c_1$ and $X_{i,2} - X_{0,2} \geq c_2$ do not select any experimental treatment.

Typically, ranking and selection problems are solved by obtaining an analytical expression for the probability of a correct selection $P(CS)$ and then finding the least favorable configuration(LFC), that parameter configuration where the $P(CS)$ is minimized. Then design parameters are obtained by setting $P(CS|LFC)$ to be at least some pre-specified value P^*. In this subset selection problem, it was not possible to derive an expression for the $P(CS)$. Instead, a lower bound for $P(CS)$ was derived, along with the parameter configuration that minimizes this bound. Then if the minimum value of this lower bound is higher than P^*, the $P(CS)$ will be at least P^* for any parameter configuration.

We denote by CFG_1 the configuration where $p_{i,1} = p_{0,1} + \delta^*_{1,1}, p_{i,2} = p_{0,2} + \delta^*_{1,2}, i = 1, .., k$ and by CFG_0 the parameter configuration where $p_{i,1} = p_{0,1} + \delta^*_{0,1}, p_{i,2} = p_{0,2} + \delta^*_{0,2}, i = 1, .., k$. Then CFG_1 and CFG_0 are the configurations under which the lower bounds $P_L(CS_1)$ and $P_L(CS_0)$ of the probabilities of correct selections $P(CS_1)$ and $P(CS_0)$, respectively, were computed. $P_L(CS_1)$ also depends on the odds ratios between the two endpoints of each of the k treatments, while $P_L(CS_0)$ does not. We assume that there is the same association between the two endpoints of each of the k treatments. Three cases were considered: independent endpoints, dependent endpoints with known association and endpoints with unknown association. When the association is not known, it was shown that the minimum of $P_L(CS_1)$ is attained when the odds ratio is zero. However, numerical computations showed that the sample size varies very little with the odds ratio. Below we state the theorems on the lower bounds $P_L(CS_0)$ and $P_L(CS_1)$ of the probabilities of correct selections $P(CS_1)$ and $P(CS_0)$, respectively, whose profs were given in Buzaianu et al (2025).

Case 1: $\phi_i = 1$, ***i = 0,1,.., k.***. We first consider the case of two independent endpoints. That is, we assume $\phi_i = 1$, *i = 0,1, 2, ..., k.* In this case, $X_{i,1}$ and $X_{i,2}$ are independent random variables following binomial distributions, with parameters $(n, p_{i,1})$ and $(n, p_{i,2})$, respectively, *i = 0,1, 2, ..., k.*

Theorem 1. *For fixed* $k, p_{0,1}, p_{0,2}, \delta^*_{0,1}, \delta^*_{0,2}, \delta^*_{1,1}, \delta^*_{1,2}$, *the probability requirements are satisfied by choosing values of* n, c_1, c_2 *that simultaneously satisfy*

$$(\sum_{x_{0,1}=0}^{n}[\sum_{x_{i,1}=c_1+x_{0,1}}^{n} b(n, p_{0,1} + \delta^*_{1,1}, x_{i,1})]^k b(n, p_{0,1}, x_{0,1})) \times$$
$$(\sum_{x_{0,2}=0}^{n}[\sum_{x_{i,2}=c_2+x_{0,2}}^{n} b(n, p_{0,2} + \delta^*_{1,2}, x_{i,2})]^k b(n, p_{0,2}, x_{0,2})) \geq P_1^*$$

and

$$\sum_{x_{0,1}=0}^{n}\sum_{x_{0,2}}^{n}(1 - \max[P(X_{i,1} \geq c_1 + x_{0,1}|p_{i,1} = p_{0,1} + \delta^*_{0,1}), P(X_{i,2} \geq c_2 + x_{0,2}|p_{i,2} = p_{0,2} + \delta^*_{0,2})])^k \times$$
$$b(n, p_{0,1}, x_{0,1})b(n, p_{0,2}, x_{0,2}) \geq P_0^*.$$

 doi:10.20944/preprints202507.0255.v1

Case 2: $\phi_i \neq 1$ **specified** ***i =0, 1,.., k.*** We now consider the case when the endpoints of each of treatment are dependent with known association.

Theorem 2. *For fixed values of* $k, p_{0,1}, p_{0,2}, \delta^*_{0,1}, \delta^*_{0,2}, \delta^*_{1,1}, \delta^*_{1,2}, \phi_i, i = 0, 1, .., k$, *the probability requirements are satisfied by choosing values of* n, c_1, c_2 *that simultaneously satisfy*

$\sum_{x_{0,1}=0}^{n} \sum_{x_{0,2}=0}^{n} [\Pi_{i=1}^{k} (\sum_{x_{i,1}=c_1+x_{0,1}}^{n} \sum_{x_{i,2}=c_2+x_{0,2}}^{n} P(X_{i,1} = x_{i,1}, X_{i,2} = x_{i,2} | p_{i,1} = p_{0,1} + \delta^*_{1,1}, p_{i,2} = p_{0,2} + \delta^*_{1,2}, \phi_i))] \times P(X_{0,1} = x_{0,1}, X_{0,2} = X_{0,2} | p_{0,1}, p_{0,2}, \phi_0) \geq P^*_1$

and

$\sum_{x_{0,1}=0}^{n} \sum_{x_{0,2}}^{n} (1 - \max[P(X_{i,1} \geq c_1 + x_{0,1} | p_{i,1} = p_{0,1} + \delta^*_{0,1}), P(X_{i,2} \geq c_2 + x_{0,2} | p_{i,2} = p_{0,2} + \delta^*_{0,2})])^k \times P(X_{0,1} = x_{0,1}, X_{0,2} = X_{0,2} | p_{0,1}, p_{0,2}, \phi_0) \geq P^*_0$.
where $p_{i,11} = \frac{a_i - \sqrt{a_i^2 + b_i}}{2(\phi_i - 1)}$, $a_i = 1 + (\phi_i - 1)(p_{0,1} + \delta^*_{1,1} + p_{0,2} + \delta^*_{1,2})$, $b_i = -4\phi_i(\phi_i - 1)(p_{0,1} + \delta^*_{1,1})(p_{0,2} + \delta^*_{1,2})$, $p_{i,12} = p_{0,1} + \delta^*_{1,1} - p_{i,11}$.

Case 3: ϕ_i **unspecified** ***i = 1, 2, ..., k,*** ϕ_0 **specified.** We now consider the case when the endpoints of each of tested treatment have unknown association.

Theorem 3. *For fixed* $k, p_{0,1}, p_{0,2}, \delta^*_{0,1}, \delta^*_{0,2}, \delta^*_{1,1}, \delta^*_{1,2}$, *the probability requirements are satisfied by choosing values of* n, c_1, c_2 *that simultaneously satisfy*

$\sum_{x_{0,1}=0}^{n} \sum_{x_{0,2}=0}^{n} [\sum_{x_{i,1}=c_1+x_{0,1}}^{n} \sum_{x_{i,2}=c_2+x_{0,2}}^{n} b(n, p_{0,2} + \delta^*_{1,2}, x_{i,2}) f(x_{i,1}, x_{i,2})]^k \times P(X_{0,1} = x_{0,1}, X_{0,2} = x_{0,2} | p_{0,1}, p_{0,2}, \phi_0) \geq P^*_1$

and

$(\sum_{x_{0,1}=0}^{n} \sum_{x_{0,2}}^{n} (1 - \max[P(X_{i,1} \geq c_1 + x_{0,1} | p_{i,1} = p_{0,1} + \delta^*_{0,1}), P(X_{i,2} \geq c_2 + x_{0,2} | p_{i,2} = p_{0,2} + \delta^*_{0,2})])^k \times P(X_{0,1} = x_{0,1}, X_{0,2} = X_{0,2} | p_{0,1}, p_{0,2}, \phi_0) \geq P^*_0$

where

$$f(x_{i,1}, x_{i,2}) = \begin{cases} b\left(n - x_{i,2}, \dfrac{p_{0,1} + \delta^*_{1,1}}{1 - (p_{0,2} + \delta^*_{1,2})}, x_{i,1}\right) & \text{if } p_{0,1} + \delta^*_{1,1} < 1 - (p_{0,2} + \delta^*_{1,2}) \\ b\left(x_{i,2}, \dfrac{p_{0,1} + \delta^*_{1,1} + p_{0,2} + \delta^*_{1,2} - 1}{p_{0,2} + \delta^*_{1,2}}, x_{i,1} + x_{i,2} - n\right) & \text{if } p_{0,1} + \delta^*_{1,1} > 1 - (p_{0,2} + \delta^*_{1,2}) \\ b\left(n, p_{0,2} + \delta^*_{1,2}, x_{i,2}\right) \mathbb{1}_{x_{i,1} + x_{i,2} = n} & \text{if } p_{0,1} + \delta^*_{1,1} = 1 - (p_{0,2} + \delta^*_{1,2}) \end{cases} \tag{3}$$

,

$$\mathbb{1}_{x_{i,1} + x_{i,2} = n} = \begin{cases} 1 & x_{i,1} + x_{i,2} = n \\ 0 & x_{i,1} + x_{i,2} \neq n \end{cases}$$

 doi:10.20944/preprints202507.0255.v1

and

$$\phi_i = 0, i = 1, 2, \ldots, k$$

Remark 1: The lower bound on $P(CS_0)$ depends only on the odds ratio for the control treatment.

Remark 2: Buzaianu et al. (2025) demonstrated that $P(CS_1)$ increases with the odds ratios of the experimental treatments π_i, for $i = 1, 2, \ldots, k$. Therefore, the minimum value of $P(CS_1)$ is achieved when the odds ratios of all experimental treatments are zero. To obtain a lower bound for $P(CS_1)$, we evaluated it under the assumption that all treatments tested have odds ratios equal to zero. As a result, the scenario with unspecified odds ratios is effectively handled by considering the scenario where the odds ratios of the tested treatments are zeros.

Remark 3: Our results are derived under the assumption that the association between the two endpoints is of the same type for each of the k treatments. For example, either the two endpoints are independent for all k treatments, or they are dependent with a known form of association for all k treatments. However, based on the structure of our derivations, scenarios in which treatments exhibit different types of associations between the two endpoints—such as independence for some treatments and unknown dependence for others—can be readily accommodated.

4. Proposed Curtailment Procedure

We propose a curtailed procedure to achieve the objective outlined in Section 2. The proposed procedure, denoted by **R**, is a sequential method that employs curtailment to reduce the sample size for treatments that are either clearly inferior or sufficiently effective. Let n denote the maximum number of observations per treatment that the experimenter is permitted to collect.

Curtailment Procedure R:

A contending treatment is a treatment that has not been eliminated from the experiment. Procedure **R** begins with all *k+1* populations being the contending populations. We will use the vector-at-a-time sampling rule. By "Step M", where $1 \leq M \leq n$, we mean that a total of M vectors have been sampled thus far. Let $Y_{(i,1),M}$ and $Y_{(i,2),M}$ respectively denote the numbers of successes from the two endpoints of π_i through Step M.

Sampling Rule. We use a vector-at-a-time sampling rule with the following restrictions:

(a) At most n observations can be taken from each of the $(k+1)$ populations. Observations are taken from each contending treatment one at a time until either the total number of observations from that treatment reaches n, or the treatment is eliminated according to conditions (b) or (c) below.

(b) At any step M, if the number of successes for the two endpoints, $Y_{(i,1),M}$ and $Y_{(i,2),M}$, of treatment π_i satisfy

$$Y_{(i,1),M} + n - M < c_1 + Y_{(0,1),M} \quad \text{or} \quad Y_{(i,2),M} + n - M < c_2 + Y_{(0,2),M},$$

then eliminate treatment π_i and stop sampling from it.

(c) At any step M, if the number of successes for the two endpoints, $Y_{(i,1),M}$ and $Y_{(i,2),M}$, of treatment π_i satisfy

$$Y_{(i,1),M} \geq c_1 + Y_{(0,1),M} + n - M \quad \text{and} \quad Y_{(i,2),M} \geq c_2 + Y_{(0,2),M} + n - M,$$

 doi:10.20944/preprints202507.0255.v1

then stop sampling from treatment π_i.

Stopping Rule:

Stop the experiment at the first step M when any of the following three conditions is satisfied:

(i) There exists a partition A, B of the set $\{1, 2, \ldots, k\}$ such that:

$$Y_{(i,1),M} \geq c_1 + Y_{(0,1),M} + n - M \quad \text{and} \quad Y_{(i,2),M} \geq c_2 + Y_{(0,2),M} + n - M, \quad \text{for all } i \in A,$$
$$Y_{(j,1),M} + n - M < c_1 + Y_{(0,1),M} \quad \text{or} \quad Y_{(j,2),M} + n - M < c_2 + Y_{(0,2),M}, \quad \text{for all } j \in B.$$

(ii) For all $i \in \{1, 2, \ldots, k\}$,

$$Y_{(i,1),M} + n - M < c_1 + Y_{(0,1),M} \quad \text{or} \quad Y_{(i,2),M} + n - M < c_2 + Y_{(0,2),M}.$$

(iii) $M = n$.

Decision Rule:

(a) If the sampling stops according to (i) of the above ***Stopping Rule***, we include in the selected subset all π_i in A.

(b) If the sampling stops according to (ii) of the above ***Stopping Rule***, we declare that no experimental treatment is significantly better than the control treatment π_0.

(c) If the sampling stops according to (iii) of the above ***Stopping Rule***, we include in the selected subset all the treatments π_i with $Y_{(i,1),n} \geq c_1 + Y_{(0,1),n}$ and $Y_{(i,2),n} \geq c_2 + Y_{(0,2),n}$. If the selected subset is an empty set, we declare that no experimental treatment is significantly better than the control treatment π_0.

Theorem 4. *For given k and n, both **H** and **R** select the same subset of k experimental treatments if both use the same c_1 and c_2. The result is uniform in $\{(p_{0,1}, p_{0,2}), (p_{1,1}, p_{1,2}), \ldots, (p_{k,1}, p_{k,2})\}$.*

Proof. Decision Rule (c) of Procedure **R**, which is applied if and only if sampling stops according to Stopping Rule (iii), is identical to the decision rule of Procedure **H** when $m = n$. Thus, the same subset will be selected by **R** and **H** for any sampling outcome in which a total of n observations is taken from each of the k treatments. Therefore, we only need to consider the case in which the decision under **R** is made according to Decision Rule (a) or (b).

Decision Rules (a) and (b) are invoked if and only if sampling stops according to Stopping Rule (i) or (ii), respectively. Note that $M < n$ whenever the sampling stops due to Stopping Rule (i) or (ii). When this occurs, we have $n - M > 0$.

If sampling stops according to Stopping Rule (i), then under Procedure **R**, Decision Rule (a) selects the subset consisting of treatments π_i such that

$$Y_{(i,1),M} \geq c_1 + Y_{(0,1),M} + n - M \quad \text{and} \quad Y_{(i,2),M} \geq c_2 + Y_{(0,2),M} + n - M.$$

Now suppose the experiment were to continue as it would under Procedure **H**. Let $Y_{(i,1),n}$ and $Y_{(i,2),n}$ denote the total number of successes for treatment π_i at endpoints 1 and 2, respectively, after n observations. Then, by Rule (2) of Procedure **H**, treatment π_i would be selected if

$$Y_{(i,1),n} - c_1 \geq Y_{(0,1),n} \quad \text{and} \quad Y_{(i,2),n} - c_2 \geq Y_{(0,2),n}.$$

 doi:10.20944/preprints202507.0255.v1

Observe that:

$$Y_{(i,1),n} - c_1 \geq Y_{(i,1),M} - c_1 \geq Y_{(0,1),M} + n - M \geq Y_{(0,1),n},$$
$$Y_{(i,2),n} - c_2 \geq Y_{(i,2),M} - c_2 \geq Y_{(0,2),M} + n - M \geq Y_{(0,2),n}.$$

Hence, the same subset of treatments would be selected by Procedure **H**.

Similarly, if sampling stops according to Stopping Rule (ii), then Decision Rule (b) of Procedure **R** selects no experimental treatment. This is exactly the same decision that would be made by Rule (2) of Procedure **H**.

This completes the proof of the theorem. □

5. Tables

In this section, we evaluate the performance of the curtailment procedure in terms of sample size savings relative to the corresponding non-curtailment procedure **H**. We assume the same association structure between the two endpoints for each of the $(k+1)$ treatments. However, based on our results, parameter derivations for the curtailment procedure can also be extended to scenarios where different treatments exhibit different associations between the two endpoints.

We consider the following cases: $k = 2, 3$, $p_{0,1} = 0.40, 0.50, 0.60$, $p_{0,2} = 0.60$, $\delta^*_{0,1} = 0.01$, $\delta^*_{0,2} = 0.01$, $\delta^*_{1,1} = 0.30$, $\delta^*_{1,2} = 0.25$, $P^*_0 = 0.90$, $P^*_1 = 0.80, 0.85, 0.90$, and $\phi = 0, 0.01, 0.1, 1, 2, 4, 8, 100$, which follow the settings used by Buzaianu et al. (2025) to create Tables 2–7 for the fixed sample size procedure **H**. For each case, they first determined the minimum number of observations per treatment, n, and the associated (c_1, c_2) values that satisfy the required probability constraints. If multiple (c_1, c_2) combinations met these constraints, the design yielding the highest probability of selecting a single effective treatment was chosen. This n is then used as the maximum number of observations per treatment under the curtailment procedure **R**. According to Theorem 4.1, with this choice of n and (c_1, c_2), the curtailment procedure satisfies the same probability requirements as procedure **H**.

We denote by N the total number of observations required by the fixed sample size procedure to satisfy the probability requirements, where $N = (k+1) \times n$. This N also serves as the upper bound on the total number of observations under the curtailment procedure. Let $E_1(N|R)$ and $E_0(N|R)$ denote the expected sample sizes for the curtailment procedure **R** under configurations CFG_1 and CFG_0 defined in Section 3. These configurations were used to compute the lower bounds $P_L(CS_1)$ and $P_L(CS_0)$ for selecting a correct subset under the alternative and null hypotheses, respectively, and were used to derive the design parameters for the fixed sample size procedure **H**.

We define the average expected sample size under the curtailment procedure as $E(N \mid R) = \frac{1}{2}(E_1(N \mid R) + E_0(N \mid R))$, following the approach of Rhall, Simon, and Ellenberg, to account for performance under both configurations. The quantities $E_1(N \mid R)$ and $E_0(N \mid R)$ are estimated via simulation (10,000 repetitions), implemented in R. To generate bivariate binary data with marginal probabilities p_1, p_2 and odds ratio $\phi \neq 1$, we compute:

$$p_{11} = \frac{a - \sqrt{a^2 + b}}{2(\phi - 1)}, \quad a = 1 + (\phi - 1)(p_1 + p_2), \quad b = -4\phi(\phi - 1)p_1 p_2,$$
$$p_{12} = p_1 - p_{11}, \quad p_{21} = p_2 - p_{11}, \quad p_{22} = 1 - (p_{11} + p_{12} + p_{21}),$$

and then simulate binary outcomes from a 2×2 table with cell probabilities p_{ij}.

Tables 2–7 report, for each specification, the total sample size N required by the fixed sample size procedure, the expected sample sizes $E_1(N|R)$ and $E_0(N|R)$ for the curtailment procedure, and the percentage of observations saved using curtailment: $RS(\%) = \frac{N - E(N|R)}{N} \times 100$. It is evident that the

 doi:10.20944/preprints202507.0255.v1

curtailment procedure **R** requires substantially fewer observations than procedure **H** to satisfy the same performance criteria.

Tables 2–7 also demonstrate that the odds ratio ϕ has little impact on the expected sample size when ϕ values are relatively close, under both CFG_1 and CFG_0, for the curtailment procedure. Greater variability is observed under CFG_0. This pattern is consistent with the findings of Chen and Chi (2012), who considered only moderate odds ratios ($\phi = 2, 4, 8$) in the context of hypothesis testing, and observed minimal sample size variation under curtailment, with larger variability under the null hypothesis. In our procedure, however, when ϕ varies substantially—for example, from 1 to 100—we observe a marked decrease in expected sample size under both CFG_1 and CFG_0. Chen and Chi (2012) did not report average expected sample sizes, but instead presented results under both the null and alternative hypotheses and calculated percentage savings. Their findings indicated modest savings under the alternative and substantial savings under the null, which align with our observations.

Table 2. Design parameters when $k = 2, P_1^* = .80$.

$p_{0,1}$	ϕ	n	c_1	c_2	N	$E_0(N\|R)$	$E_1(N\|R)$	$E(N\|R)$	$RS\%$
0.4	0	81	14	13	243	195.96	232.41	214.18	11.86
	0.01	81	14	13	243	196.61	232.10	214.36	11.79
	0.1	78	14	12	234	190.27	222.84	206.55	11.73
	1	77	14	12	231	189.83	219.01	204.42	11.51
	2	75	13	12	225	186.30	212.83	199.56	11.3
	4	75	13	12	225	187.11	212.41	199.76	11.22
	8	71	13	11	213	177.07	200.73	188.90	11.32
	100	69	12	11	207	174.27	194.16	184.22	11.01
0.5	0	80	14	13	240	193.65	229.38	211.52	11.87
	0.01	80	14	13	240	193.97	229.26	211.61	11.83
	0.1	77	14	12	231	187.35	220.16	203.76	11.79
	1	76	14	12	228	186.77	216.31	201.54	11.61
	2	74	13	12	222	183.31	210.23	196.77	11.37
	4	74	13	12	222	184.22	209.72	196.97	11.27
	8	70	13	11	210	174.25	198	186.12	11.37
	100	67	12	11	201	169.16	188.56	178.86	11.01
0.6	0	75	14	12	225	180.99	215.14	198.06	11.97
	0.01	75	14	12	225	181.04	215.12	198.08	11.97
	0.1	73	13	12	219	177.48	208.88	193.18	11.79
	1	73	13	12	219	179.61	207.89	193.75	11.53
	2	71	14	11	213	173.74	202.21	187.97	11.75
	4	69	13	11	207	170.59	195.87	183.23	11.48
	8	67	12	11	201	167.16	189.63	178.40	11.25
	100	62	12	10	186	155.61	174.58	165.10	11.24

Note: $k = 2, p_{0,2} = .60, \delta_{0,1}^* = .01, \delta_{0,2}^* = .01, \delta_{1,1}^* = .30, \delta_{1,2}^* = .25$ $P_0^* = .90, P_1^* = .80$

 doi:10.20944/preprints202507.0255.v1

Table 3. Design parameters when $k = 2, P_1^* = .85$.

$p_{0,1}$	ϕ	n	c_1	c_2	N	$E_0(N\|R)$	$E_1(N\|R)$	$E(N\|R)$	$RS(\%)$
0.4	0	87	15	13	261	212.04	248.41	230.23	11.79
	0.01	87	15	13	261	212.71	248.08	230.40	11.73
	0.10	85	14	13	255	209.71	241.68	225.70	11.49
	1	85	14	13	255	212.24	240.44	226.34	11.24
	2	81	14	12	243	202.46	228.73	215.60	11.28
	4	81	14	12	243	203.30	228.21	215.75	11.21
	8	81	14	12	243	204.02	227.76	215.89	11.16
	100	78	13	12	234	198.24	218.67	208.46	10.92
0.5	0	86	15	13	258	209.73	245.41	227.57	11.79
	0.01	86	15	13	258	209.97	245.28	227.63	11.77
	0.10	84	14	13	252	206.76	238.98	222.87	11.56
	1	84	14	13	252	209.18	237.80	223.49	11.31
	2	80	14	12	240	199.46	226.11	212.79	11.34
	4	80	14	12	240	200.36	225.57	212.97	11.26
	8	78	13	12	234	196.99	219.37	208.18	11.03
	100	73	13	11	219	185.24	204.25	194.75	11.07
0.6	0	81	15	12	243	196.90	231.23	214.06	11.91
	0.01	81	15	12	243	197.00	231.18	214.09	11.90
	0.10	81	15	12	243	197.70	230.81	214.25	11.83
	1	78	14	12	234	192.70	221.34	207.02	11.53
	2	78	14	12	234	193.54	220.89	207.22	11.45
	4	77	13	12	231	193.19	217.16	205.18	11.18
	8	74	14	11	222	184.30	208.86	196.58	11.45
	100	70	12	11	210	178.72	195.91	187.32	10.80

Note: $k = 2, p_{0,2} = .60, \delta_{0,1}^* = .01, \delta_{0,2}^* = .01, \delta_{1,1}^* = .30, \delta_{1,2}^* = .25$ $P_0^* = .90, P_1^* = .85$

 doi:10.20944/preprints202507.0255.v1

Table 4. Design parameters when $k = 2, P_1^* = .90$.

$p_{0,1}$	ϕ	n	c_1	c_2	N	$E_0(N\|R)$	$E_1(N\|R)$	$E(N\|R)$	$RS(\%)$
0.4	0	96	15	14	288	237.24	272.11	254.67	11.57
	0.01	96	15	14	288	237.93	271.76	254.85	11.51
	0.10	96	15	14	288	239.33	271.17	255.25	11.37
	1	92	15	13	276	231.39	258.66	245.02	11.22
	2	92	15	13	276	232.31	258.15	245.23	11.15
	4	90	14	13	270	228.93	251.92	240.42	10.95
	8	90	14	13	270	229.75	251.42	240.58	10.90
	100	85	14	12	255	217.44	236.74	227.09	10.94
0.5	0	95	15	14	285	234.97	269.09	252.03	11.57
	0.01	95	15	14	285	235.24	268.96	252.10	11.54
	0.10	93	16	13	279	230.01	263.14	246.57	11.62
	1	91	15	13	273	228.35	256.03	242.19	11.29
	2	91	15	13	273	229.28	255.49	242.39	11.21
	4	89	14	13	267	225.99	249.27	237.63	11.00
	8	89	14	13	267	226.93	248.71	237.82	10.93
	100	82	13	12	246	211.21	227.98	219.59	10.73
0.6	0	89	15	13	267	219.34	252.27	235.80	11.68
	0.01	89	15	13	267	219.46	252.20	235.83	11.67
	0.10	89	15	13	267	220.23	251.86	236.04	11.59
	1	88	14	13	264	221.17	247.47	234.32	11.24
	2	87	16	12	261	216.51	245.11	230.81	11.57
	4	85	15	12	255	213.46	238.66	226.06	11.35
	8	83	14	12	249	210.29	232.34	221.31	11.12
	100	78	14	11	234	198.57	217.13	207.85	11.17

Note: $k = 2, p_{0,2} = .60, \delta_{0,1}^* = .01, \delta_{0,2}^* = .01, \delta_{1,1}^* = .30, \delta_{1,2}^* = .25$ $P_0^* = .90, P_1^* = .90$

 doi:10.20944/preprints202507.0255.v1

Table 5. Design parameters when $k = 3, P_1^* = .80$.

$p_{0,1}$	ϕ	n	c_1	c_2	N	$E_0(N\|R)$	$E_1(N\|R)$	$E(N\|R)$	$RS(\%)$
0.4	0.00	95	16	15	380	308.89	362.28	335.59	11.69
	0.01	94	17	14	376	305.01	357.95	331.48	11.84
	0.10	93	17	14	372	303.28	354.08	328.68	11.65
	1.00	91	16	14	364	300.94	344.40	322.67	11.35
	2.00	89	15	14	356	296.73	335.82	316.27	11.16
	4.00	88	15	14	352	293.36	331.98	312.67	11.17
	8.00	86	14	14	344	288.88	323.93	306.41	10.93
	100.00	82	14	13	328	277.11	307.69	292.40	10.85
0.5	0.00	92	17	14	368	297.22	351.12	324.17	11.91
	0.01	92	17	14	368	297.68	351.02	324.35	11.86
	0.10	92	17	14	368	299.18	350.45	324.82	11.73
	1.00	90	16	14	360	296.63	340.58	318.61	11.50
	2.00	88	15	14	352	292.18	332.41	312.30	11.28
	4.00	87	15	14	348	289.80	328.60	309.20	11.15
	8.00	85	16	13	340	282.10	320.37	301.24	11.40
	100.00	80	14	13	320	270.15	300.19	285.17	10.88
0.6	0.00	88	16	14	352	284.93	335.77	310.35	11.83
	0.01	88	16	14	352	285.05	335.56	310.30	11.85
	0.10	86	15	14	344	280.29	327.42	303.86	11.67
	1.00	86	17	13	344	279.82	326.74	303.28	11.84
	2.00	85	16	13	340	279.78	321.82	300.80	11.53
	4.00	82	15	13	328	271.86	309.96	290.91	11.31
	8.00	80	14	13	320	267.28	301.55	284.42	11.12
	100.00	74	13	12	296	250.71	277.28	263.99	10.81

Note: $k = 3, p_{0,2} = .60, \delta_{0,1}^* = .01, \delta_{0,2}^* = .01, \delta_{1,1}^* = .30, \delta_{1,2}^* = .25$ $P_0^* = .90, P_1^* = .80$

 doi:10.20944/preprints202507.0255.v1

Table 6. Design parameters when $k = 3, P_1^* = .85$.

$p_{0,1}$	ϕ	n	c_1	c_2	N	$E_0(N\|R)$	$E_1(N\|R)$	$E(N\|R)$	$RS(\%)$
0.4	0.00	101	17	15	404	330.13	383.44	356.78	11.69
	0.01	101	17	15	404	330.85	382.99	356.92	11.65
	0.10	101	17	15	404	333.10	382.08	357.59	11.49
	1.00	98	16	15	392	326.96	369.43	348.19	11.18
	2.00	97	17	14	388	323.57	364.62	344.09	11.32
	4.00	95	16	14	380	319.33	356.19	337.76	11.12
	8.00	94	16	14	376	316.11	352.35	334.23	11.11
	100.00	91	15	14	364	309.03	340.14	324.59	10.83
0.5	0.00	99	17	15	396	323.11	376.07	349.59	11.72
	0.01	99	17	15	396	323.68	375.74	349.71	11.69
	0.10	99	17	15	396	325.23	375.14	350.18	11.57
	1.00	97	16	15	388	322.80	365.79	344.29	11.26
	2.00	96	17	14	384	319.40	361.15	340.28	11.39
	4.00	93	16	14	372	311.32	349.43	330.38	11.19
	8.00	93	16	14	372	312.71	348.76	330.74	11.09
	100.00	86	15	13	344	291.61	320.66	306.13	11.01
0.6	0.00	94	17	14	376	306.26	357.10	331.68	11.79
	0.01	94	17	14	376	306.41	356.78	331.60	11.81
	0.10	94	17	14	376	307.53	356.55	332.04	11.69
	1.00	92	16	14	368	305.01	347.17	326.09	11.39
	2.00	91	16	14	364	302.20	343.25	322.72	11.34
	4.00	90	15	14	360	301.71	338.37	320.04	11.10
	8.00	87	16	13	348	290.26	326.94	308.60	11.32
	100.00	84	15	13	336	284.41	313.77	299.09	10.98

Note: $k = 3, p_{0,2} = .60, \delta_{0,1}^* = .01, \delta_{0,2}^* = .01, \delta_{1,1}^* = .30, \delta_{1,2}^* = .25$ $P_0^* = .90, P_1^* = .85$

 doi:10.20944/preprints202507.0255.v1

Table 7. Design parameters when $k = 3, P_1^* = .90$.

$p_{0,1}$	ϕ	n	c_1	c_2	N	$E_0(N\|R)$	$E_1(N\|R)$	$E(N\|R)$	$RS(\%)$
0.4	0.00	111	18	16	444	365.66	419.29	392.48	11.60
	0.01	111	18	16	444	366.71	418.74	392.73	11.55
	0.10	110	17	16	440	366.62	413.66	390.14	11.33
	1.00	108	18	15	432	361.99	404.50	383.24	11.29
	2.00	106	17	15	424	358.00	395.96	376.98	11.09
	4.00	106	17	15	424	359.44	395.26	377.35	11.00
	8.00	104	16	15	416	354.66	387.29	370.97	10.82
	100.00	101	17	14	404	343.97	375.04	359.50	11.01
0.5	0.00	109	19	15	436	357.96	411.50	384.73	11.76
	0.01	109	19	15	436	358.33	411.49	384.91	11.72
	0.10	109	19	15	436	360.19	410.60	385.40	11.61
	1.00	107	18	15	428	358.00	401.18	379.59	11.31
	2.00	105	17	15	420	354.04	392.32	373.18	11.15
	4.00	104	17	15	416	351.23	388.19	369.71	11.13
	8.00	103	16	15	412	350.99	383.49	367.24	10.86
	100.00	97	16	14	388	332.10	359.83	345.97	10.83
0.6	0.00	104	18	15	416	341.64	392.59	367.11	11.75
	0.01	103	17	15	412	340.19	388.10	364.14	11.62
	0.10	103	17	15	412	341.38	387.88	364.63	11.50
	1.00	102	19	14	408	337.08	383.75	360.41	11.66
	2.00	100	18	14	400	332.66	375.32	353.99	11.50
	4.00	98	17	14	392	329.37	366.73	348.05	11.21
	8.00	97	16	14	388	328.98	361.58	345.28	11.01
	100.00	91	16	13	364	309.59	337.64	323.62	11.09

Note: $k = 3, p_{0,2} = .60, \delta_{0,1}^* = .01, \delta_{0,2}^* = .01, \delta_{1,1}^* = .30, \delta_{1,2}^* = .25$ $P_0^* = .90, P_1^* = .90$

6. Examples

6.0.1. Immunotherapy in Elderly Patients with Non-Small Cell Lung Cancer

This example considers an experimental trial involving two immunotherapy-based treatments for elderly patients ($\geq$ 75 years old) diagnosed with advanced non-small cell lung cancer (NSCLC). The trial compares two immunotherapy strategies—*PD1-A* (anti-PD-1 monotherapy) and *PD1-B* (anti-PD-1 combined with low-dose chemotherapy)—against the standard chemotherapy regimen consisting of carboplatin and pemetrexed, which serves as the control treatment.

While carboplatin plus pemetrexed is considered the standard of care in general NSCLC populations, this regimen has not been adequately studied in patients aged 75 and above. As a result, its efficacy and safety profile in this elderly subgroup remain uncertain. Historically, in younger NSCLC populations, this standard chemotherapy yields approximately 40% objective response rate (ORR), and around 40% of patients experience grade 3 or higher treatment-related adverse events. These outcomes establish the benchmark efficacy and safety rates for the control treatment as $p_{0,1} = 0.40$ and $p_{0,2} = 0.60$, respectively. Prior analyses in younger patients suggest an odds ratio of approximately 2 between efficacy and safety, indicating that patients who do not experience toxicity are more likely to respond to treatment.

The goal of the trial is to evaluate whether either *PD1-A* or *PD1-B* is superior to the control treatment in terms of both efficacy and safety. Specifically, the experimenter seeks an increase in the response rate of at least 0.30 and a reduction in high-grade toxicity of at least 0.25, corresponding to $\delta_{1,1}^* = 0.30$ and $\delta_{1,2}^* = 0.25$. If both experimental treatments fail to demonstrate improvements over the control, the standard chemotherapy will be selected, with thresholds $\delta_{0,1}^* = \delta_{0,2}^* = 0.01$.

 doi:10.20944/preprints202507.0255.v1

When $P_1^* = 0.85$ and $P_0^* = 0.90$, Table 3 shows that the fixed sample size procedure requires $n = 81$ observations per treatment, with corresponding critical values $c_1 = 14$ and $c_2 = 12$. Therefore, the total number of observations required for the fixed sample size procedure to satisfy the probability constraints is $3 \times 81 = 243$. In contrast, the curtailment procedure, while also using at most $n = 81$ observations per treatment and the same critical values $c_1 = 14$, $c_2 = 12$, is expected to achieve the same probability guarantees with fewer observations on average. According to Table 3, the expected relative sample size saving from using the curtailment procedure is approximately 11.2767%.

6.0.2. Chemotherapy of Acute Leukemia

This example involves an experimental trial comparing two different combinations of Gemcitabine and Cyclophosphamide—denoted as GemCy1 and GemCy2—each with varying dosage proportions, against the standard Ara-C regimen for treating patients with good-prognosis acute myelogenous leukemia (AML) or myelodysplastic syndrome. Historically, the standard treatment Ara-C yields approximately 60% of patients achieving complete remission (CR), while around 40% either die or experience severe myelosuppression within the first five weeks. These historical outcomes establish the efficacy and safety rates for the control treatment as $p_{0,1} = 0.6$ and $p_{0,2} = 0.6$, respectively. Additionally, the odds ratio between efficacy and safety is estimated to be 4, indicating that patients who do not experience toxicity are more likely to achieve complete remission.

The goal of the trial is to determine whether either GemCy1 or GemCy2 surpasses the control treatment in both efficacy and safety by at least 0.3 and 0.25, respectively, corresponding to threshold values $\delta_{1,1}^* = 0.3$ and $\delta_{1,2}^* = 0.25$. If both experimental treatments fail to outperform Ara-C in terms of both endpoints, the control treatment will be selected, with equivalence thresholds set at $\delta_{0,1}^* = \delta_{0,2}^* = 0.01$.

When $P_1^* = 0.80$ and $P_0^* = 0.90$, Table 2 shows that the fixed sample size procedure requires $n = 69$ observations per treatment, with corresponding critical values $c_1 = 13$ and $c_2 = 11$. Therefore, the total number of observations required for the fixed sample size procedure to satisfy the probability constraints is $3 \times 69 = 207$. In contrast, the curtailment procedure, while also using at most $n = 69$ observations per treatment and the same critical values $c_1 = 13$, $c_2 = 11$, is expected to achieve the same probability guarantees with fewer observations on average. According to Table 2, the expected relative sample size saving from using the curtailment procedure is approximately 11.48%.

7. Conclusions

This paper considers a curtailment procedure for selecting a random-size subset that contains the best treatment whenever it is significantly better than the control treatment. The comparison is made according to the two endpoints associated with the Bernoulli outcomes from each of the k experimental treatments and the control treatment. The proposed procedure is based on the fixed sample size procedure defined by Buzaianu et al (2025). The proposed procedure satisfies the same probability requirements in reaching the selection goal as does the original fixed sample size procedure, but requires fewer observations from the experimental treatments. Based on our simulations, using the curtailment procedure over the original fixed sample size procedure, would produce between a relative total sample size saving of $10\% - 12\%$. The sampling rule with curtailment is a highly desirable feature, not only because it reduces the overall sample size, but because it reduces the sample sizes from potentially undesirable populations. However, in order for such a curtailment procedure to be used, it is desirable that the time between application of the treatment and the observation of the response is small compared to the duration of the experiment.

Also, our simulations showed that the odds ratios have minimal impact on sample size; being non-sensitive to the odds ratio makes the procedure robust with regard to departures from independence.

 doi:10.20944/preprints202507.0255.v1

We only considered cases where there is the same type of association between the two endpoints of a treatment, for all treatments. However, this can be easily relaxed to accommodate situations where there are treatments that do not display the same type of association between their two endpoints.

Author Contributions: Conceptualization, Pinyuen Chen and Lifang Hsu; Methodology, Elena Buzaianu and Chishu Yin; Software, Chishu Yin; Validation, Elena Buzaianu, Pinyuen Chen, and Lifang Hsu; Formal analysis, Chishu Yin; Investigation, Elena Buzaianu and Chishu Yin; Resources, Chishu Yin; Data curation, Elena Buzaianu and Chishu Yin; Writing—original draft preparation, Chishu Yin; Writing—review and editing, Elena Buzaianu, Pinyuen Chen, and Lifang Hsu; Visualization, Chishu Yin; Supervision, Pinyuen Chen; Project administration, Lifang Hsu; Funding acquisition, Pinyuen Chen. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: The authors thank the anonymous reviewers for their valuable comments and suggestions.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Bechhofer, R.E.; Kulkarni, R.V. Closed Adaptive Sequential Procedures for Selecting the Best of $k \geq 2$ Bernoulli Populations. In *Proceedings of the Third Purdue Symposium in Statistical Decision Theory and Related Topics*; Gupta, S.S.; Berger, J., Eds.; Volume I, pp. 61–108, 1982.
2. Bryant, J.; Day, R. Incorporating Toxicity Considerations into the Design of Two-Stage Phase II Clinical Trials. *Biometrics* **1995**, *51*, 1372–1383.
3. Buzaianu, E.M.; Chen, P. Curtailment Procedure for Selecting Among Bernoulli Populations. *Commun. Stat. Theory Methods* **2008**, *37*(7), 1085–1102.
4. Buzaianu, E.M.; Chen, P.; Hsu, L. A Curtailed Procedure for Selecting Among Treatments With Two Bernoulli Endpoints. *Sankhya B* **2022**, *84*(1), 320–339. https://doi.org/10.1007/s13571-021-00261-2
5. Buzaianu, E.M.; Chen, P.; Hsu, L. Selecting among treatments with two Bernoulli endpoints. *Commun. Stat. Theory Methods* **2024**, *53*(6), 1964–1984. https://doi.org/10.1080/03610926.2022.2117557
6. Yin, C.; Buzaianu, E.M.; Chen, P.; Hsu, L. A Design for Selecting Among Treatments with Two Binary Endpoints. *Unpublished manuscript*, 2025.
7. Carsten, C.; Chen, P. Curtailed Two-Stage Matched Pairs Design in Double-Arm Phase II Clinical Trials. *J. Biopharm. Stat.* **2016**, *26*(5), 816–822.
8. Conway, M.R.; Petroni, G.R. Bivariate Sequential Designs for Phase II Trials. *Biometrics* **1995**, *51*, 656–664.
9. Chen, C.M.; Chi, Y. Curtailed two-stage designs with two dependent binary endpoints. *Pharm. Stat.* **2012**, *11*, 57–62.
10. Dunnett, C.M. Selection of the Best Treatment in Comparison to a Control with an Application to a Medical Trial. In *Design of Experiments, Ranking, and Selection*; Santner, T.J.; Tamhane, A.C., Eds.; Marcel Dekker: New York, 1984; pp. 47–66.
11. Gupta, S.S.; Sobel, M. On Selecting a Subset which Contains all Populations Better than a Standard. *Ann. Math. Stat.* **1958**, *29*, 235–244.
12. Jennison, C. Equal Probability of Correct Selection for Bernoulli Selection Procedures. *Commun. Stat. Theory Methods* **1983**, *12*(24), 2887–2896.
13. Jennison, C.; Turnbull, B.W. Group sequential tests for bivariate response: interim analysis of clinical trials with both efficacy and safety endpoints. *Biometrics* **1993**, *49*, 741–752.

 doi:10.20944/preprints202507.0255.v1

14. Sobel, M.; Hyuett, M.J. Selecting the Best One of Several Binomial Populations. *Bell Syst. Tech. J.* **1957**, *36*, 537–576.
15. Royzman, I. Hurdles for Neulasta Biosimilars. Available online: https://www.biologicsblog.com/hurdles-for-neulasta-biosimilars/ (accessed on June 27, 2025).
16. Thall, P.F.; Simon, R.; Ellenberg, S.S. Two-stage designs for comparative clinical trials. *Biometrika* **1988**, *75*, 303–310.
17. Wikipedia. Pegfilgrastim. Available online: https://en.wikipedia.org/wiki/Pegfilgrastim (accessed on June 27, 2025).