

Cost-minimizing sample size allocation for comparing two proportions

Jiangtao Luo^{1*}, Chaoyuan Mary Liu², Ismail El Mouddeh³, Mohan D. Pant⁴, Yu Wang⁵, Zhong Wang^{5,6*} and Hongchao Zhang⁷

¹ Independent Researcher in Biostatistics, 9905 Kingshyre Way, Tampa, FL 33647, USA

² Mathematics and Statistics Department, Eastern Kentucky University, Richmond, KY 40475, USA

³ Macon & Joan Brock Virginia Health Sciences, Old Dominion University, Norfolk, VA 23501, USA

⁴ Department of Epidemiology, Biostatistics & Environmental Health, Joint School of Public Health, Old Dominion University, Norfolk 23529, VA, USA

⁵ Beijing Key Laboratory of Topological Statistics and Application for Complex Systems, Beijing Institute of Mathematical Sciences and Applications, Beijing 100084, China

⁶ School of Software Technology, Dalian University of Technology, Dalian 116024, China

⁷ Department of Mathematics, Louisiana State University, Baton Rouge, LA 77803-4918, USA

* Correspondence: johntowlord@gmail.com (Luo J); wangzhong@bimsa.cn (Wang Z)

Abstract

We develop cost-minimizing sample size allocation methods for comparing two proportions in medical studies while controlling the type I error rate and maintaining prespecified power. Closed-form formulas are derived for two-sided tests of equality, non-inferiority/superiority tests, and equivalence tests. We provide an illustrative example for each testing scenario to demonstrate the cost efficiency of the proposed allocation. The results may help investigators design cost-efficient studies when unit costs differ between treatment arms and financial resources are limited. We further show that the corresponding normal-approximation power constraint can be interpreted as fixing the asymptotic variance of the estimated difference between the two proportions, or, equivalently, the sum of the corresponding inverse Fisher information contributions.

Citation: Luo J, Liu CM, El Mouddeh I, Pant MD, Wang Y, et al. 2026. Cost-minimizing sample size allocation for comparing two proportions. *Statistics Innovation* 3: e009 <https://doi.org/10.48130/stati-0026-0009>

Introduction

Financial constraints are a central consideration in clinical trial design. When per-subject costs differ between treatment arms, equal allocation may achieve the desired power but may not minimize the total study cost. This motivates sample size allocation methods that minimize the variable cost while maintaining a prespecified type I error rate and power.

Optimal allocation problems have been studied in related settings. Brittain and Schlesselman discussed optimal allocation for comparing two proportions^[1], and Allison et al. considered statistically powerful study designs under financial constraints^[2]. Classical optimal allocation ideas also appear in survey sampling^[3]. More recent studies have examined cost-constrained or power-maximizing allocation problems for proportions, trimmed means, and other comparative designs^[4–6]. However, closed-form cost-minimizing allocation formulas for several common two-proportion testing scenarios remain valuable for practical trial planning.

This paper derives closed-form continuous sample-size allocations for comparing two independent proportions under unequal unit costs. We consider two-sided tests of equality, non-inferiority/superiority tests, and equivalence tests. For each scenario, we minimize total variable cost subject to the corresponding normal-approximation power constraint. We also show that the normal-approximation power constraint fixes the asymptotic variance of the estimated difference between the two proportions. This provides an inverse Fisher information interpretation of the proposed allocation problem.

We first introduce the notation and assumptions. We assume that the two samples are independent and drawn from two populations.

Specifically, let m subjects be assigned to the intervention group and n subjects to the control group with the corresponding binary response indicators X_1, \dots, X_m and Y_1, \dots, Y_n , respectively. We further assume that

$$X_1, \dots, X_m \sim \text{Bernoulli}(\pi_1), Y_1, \dots, Y_n \sim \text{Bernoulli}(\pi_2). \quad (1)$$

Then, we have

$$\sum_{i=1}^m X_i \sim \text{Binomial}(m, \pi_1), \sum_{j=1}^n Y_j \sim \text{Binomial}(n, \pi_2) \quad (2)$$

under Eq. (1). Our estimates for π_1 and π_2 are, respectively,

$$\hat{\pi}_1 = \frac{1}{m} \sum_{i=1}^m X_i \text{ and } \hat{\pi}_2 = \frac{1}{n} \sum_{j=1}^n Y_j. \quad (3)$$

The cost of a clinical trial consists of two parts. The first part is usually fixed and includes expenses for physicians, nurses, researchers, and other staff members involved in conducting and analyzing the trial. The second part depends on the number of study subjects in the trial. For simplicity, we assume that c_1 and c_2 denote the unit costs in the intervention and control groups, respectively. The variable cost is therefore

$$C = mc_1 + nc_2. \quad (4)$$

Our goal is to minimize the variable cost in Eq. (4) subject to the power constraint for each hypothesis-testing scenario. Because fixed costs do not affect the allocation problem, we focus on the variable cost in Eq. (4) and refer to it as the total variable cost hereafter. We also assume that our sample sizes are large enough to allow us to use the normal approximation for sample size calculation.

Main results

Cost minimization for the two-sided test of equality

The following hypotheses are used to test whether the expected response rates in the intervention (π_1) group and the control (π_2) group are statistically different:

$$H_0: \pi_1 = \pi_2 \text{ versus } H_1: \pi_1 \neq \pi_2. \quad (5)$$

The scientific question is whether the response probabilities differ between the two groups. This is commonly referred to as a two-sided test of equality for two proportions^[7].

Theorem 1. The sample sizes for hypotheses (Eq. [5]) to achieve the global minimum cost with power $1 - \beta$ at a significance level α are

$$m = \frac{[\pi_1(1 - \pi_1)]^{\frac{1}{2}} \left([c_1\pi_1(1 - \pi_1)]^{\frac{1}{2}} + [c_2\pi_2(1 - \pi_2)]^{\frac{1}{2}} \right)}{c_1^{\frac{1}{2}} \left(\frac{\pi_1 - \pi_2}{z_{1-\alpha/2} + z_{1-\beta}} \right)^2}$$

and

$$n = \frac{[\pi_2(1 - \pi_2)]^{\frac{1}{2}} \left([c_1\pi_1(1 - \pi_1)]^{\frac{1}{2}} + [c_2\pi_2(1 - \pi_2)]^{\frac{1}{2}} \right)}{c_2^{\frac{1}{2}} \left(\frac{\pi_1 - \pi_2}{z_{1-\alpha/2} + z_{1-\beta}} \right)^2}.$$

Proof. According to Chow et al.^[7], the constraint condition for intervention and control groups to achieve the power $1 - \beta$ at significant level α is

$$\frac{|\pi_1 - \pi_2|}{\sqrt{\frac{\pi_1(1 - \pi_1)}{m} + \frac{\pi_2(1 - \pi_2)}{n}}} - z_{1-\alpha/2} = z_{1-\beta}.$$

Equivalently,

$$\frac{\pi_1(1 - \pi_1)}{m} + \frac{\pi_2(1 - \pi_2)}{n} - \left(\frac{\pi_1 - \pi_2}{z_{1-\alpha/2} + z_{1-\beta}} \right)^2 = 0. \quad (6)$$

The corresponding Lagrangian for minimizing Eq. (4) subject to constraint Eq. (6) is, according to Bertsekas^[8]:

$$L(m, n, \lambda) = mc_1 + nc_2 + \lambda \left[\frac{\pi_1(1 - \pi_1)}{m} + \frac{\pi_2(1 - \pi_2)}{n} - \left(\frac{\pi_1 - \pi_2}{z_{1-\alpha/2} + z_{1-\beta}} \right)^2 \right]. \quad (7)$$

Setting

$$\nabla_{\{m,n\}} L(m, n, \lambda) = 0,$$

we obtain

$$m = \sqrt{\frac{\lambda\pi_1(1 - \pi_1)}{c_1}} \text{ and } n = \sqrt{\frac{\lambda\pi_2(1 - \pi_2)}{c_2}}. \quad (8)$$

Plugging Eq. (8) in Eq. (6), we obtain

$$\sqrt{\frac{c_1\pi_1(1 - \pi_1)}{\lambda}} + \sqrt{\frac{c_2\pi_2(1 - \pi_2)}{\lambda}} - \left(\frac{\pi_1 - \pi_2}{z_{1-\alpha/2} + z_{1-\beta}} \right)^2 = 0,$$

which implies that

$$\sqrt{\lambda} = \frac{\sqrt{c_1\pi_1(1 - \pi_1)} + \sqrt{c_2\pi_2(1 - \pi_2)}}{\left(\frac{\pi_1 - \pi_2}{z_{1-\alpha/2} + z_{1-\beta}} \right)^2}. \quad (9)$$

By plugging Eq. (9) into Eq. (8), we obtain the solutions in **Theorem 1.** Note that the Hessian matrix for the Lagrangian function Eq. (7) is

$$\nabla^2 L(m, n, \lambda) = \begin{bmatrix} 2\lambda \frac{\pi_1(1 - \pi_1)}{m^3} & 0 \\ 0 & 2\lambda \frac{\pi_2(1 - \pi_2)}{n^3} \end{bmatrix},$$

which is positive definite. The solution in **Theorem 1** is the only KKT point of the problem, which together with the positive definite Hessian, implies it is the global minimizer of the cost under the constraint in Eq. (6).

Example 1. Suppose the expected cure rates are 80% in the intervention group and 65% in the control group receiving standard treatment. Then, the sample sizes under equal allocation with 80% power at the 0.05 significance level are^[7]

$$m = n = \frac{(z_{1-\alpha/2} + z_{1-\beta})^2 [\pi_1(1 - \pi_1) + \pi_2(1 - \pi_2)]}{|\pi_1 - \pi_2|^2} = \frac{(1.96 + 0.84)^2 [0.8(1 - 0.8) + 0.65(1 - 0.65)]}{0.15^2} \approx 135.$$

If the per-subject costs are \$800 for the intervention group and \$200 for the control group, then the total cost will be

$$C = 135 \times 800 + 135 \times 200 = 135,000.$$

Using our results from **Theorem 1**, we have

$$m = \frac{[\pi_1(1 - \pi_1)]^{\frac{1}{2}} \left([c_1\pi_1(1 - \pi_1)]^{\frac{1}{2}} + [c_2\pi_2(1 - \pi_2)]^{\frac{1}{2}} \right)}{c_1^{\frac{1}{2}} \left(\frac{\pi_1 - \pi_2}{z_{1-\alpha/2} + z_{1-\beta}} \right)^2} = \frac{[0.8(1 - 0.8)]^{\frac{1}{2}} \left([800 \times 0.8(1 - 0.8)]^{\frac{1}{2}} + [200 \times 0.65(1 - 0.65)]^{\frac{1}{2}} \right)}{800^{\frac{1}{2}} \left(\frac{0.8 - 0.65}{1.96 + 0.84} \right)^2} \approx 89$$

and

$$n = \frac{[\pi_2(1 - \pi_2)]^{\frac{1}{2}} \left([c_1\pi_1(1 - \pi_1)]^{\frac{1}{2}} + [c_2\pi_2(1 - \pi_2)]^{\frac{1}{2}} \right)}{c_2^{\frac{1}{2}} \left(\frac{\pi_1 - \pi_2}{z_{1-\alpha/2} + z_{1-\beta}} \right)^2} = \frac{[0.65(1 - 0.65)]^{\frac{1}{2}} \left([800 \times 0.8(1 - 0.8)]^{\frac{1}{2}} + [200 \times 0.65(1 - 0.65)]^{\frac{1}{2}} \right)}{200^{\frac{1}{2}} \left(\frac{0.8 - 0.65}{1.96 + 0.84} \right)^2} \approx 212,$$

with a total cost of

$$C^* = 89 \times 800 + 212 \times 200 = 113,600.$$

Moreover,

$$\frac{C - C^*}{C} = \frac{135,000 - 113,600}{135,000} \approx 0.1585,$$

which indicates that the proposed allocation reduces the total cost by approximately 15.85% relative to equal allocation.

Cost minimization for non-inferiority and superiority tests

We use the following hypotheses to test non-inferiority or superiority^[7]:

$$H_0: \pi_1 - \pi_2 \leq \Delta \text{ versus } H_1: \pi_1 - \pi_2 > \Delta, \quad (10)$$

where Δ is defined as a signed margin, with positive values corresponding to a superiority margin and negative values corresponding to a non-inferiority margin. We assume $\pi_1 - \pi_2 - \Delta > 0$, so that the alternative hypothesis is separated from the null boundary by a positive effect size.

Theorem 2. The sample sizes for hypotheses Eq. (10) to achieve the global minimum cost with the power $1 - \beta$ at a significance level α are

$$m = \frac{[\pi_1(1 - \pi_1)]^{\frac{1}{2}} \left([c_1\pi_1(1 - \pi_1)]^{\frac{1}{2}} + [c_2\pi_2(1 - \pi_2)]^{\frac{1}{2}} \right)}{c_1^{\frac{1}{2}} \left(\frac{\pi_1 - \pi_2 - \Delta}{z_{1-\alpha} + z_{1-\beta}} \right)^2}$$

and

$$n = \frac{[\pi_2(1 - \pi_2)]^{\frac{1}{2}} \left([c_1\pi_1(1 - \pi_1)]^{\frac{1}{2}} + [c_2\pi_2(1 - \pi_2)]^{\frac{1}{2}} \right)}{c_2^{\frac{1}{2}} \left(\frac{\pi_1 - \pi_2 - \Delta}{z_{1-\alpha} + z_{1-\beta}} \right)^2}.$$

Proof. The constraint condition for the intervention and control groups to achieve power $1 - \beta$ at a significance level $\alpha^{[7]}$ is given as:

$$\frac{\pi_1 - \pi_2 - \Delta}{\sqrt{\frac{\pi_1(1 - \pi_1)}{m} + \frac{\pi_2(1 - \pi_2)}{n}}} - z_{1-\alpha} = z_{1-\beta},$$

which is equivalent to

$$\frac{\pi_1(1 - \pi_1)}{m} + \frac{\pi_2(1 - \pi_2)}{n} - \left(\frac{\pi_1 - \pi_2 - \Delta}{z_{1-\alpha} + z_{1-\beta}} \right)^2 = 0. \quad (11)$$

Therefore, the Lagrangian function for minimizing Eq. (4) subject to constraint Eq. (11) is^[8]

$$L(m, n, \lambda) = mc_1 + nc_2 + \lambda \left[\frac{\pi_1(1 - \pi_1)}{m} + \frac{\pi_2(1 - \pi_2)}{n} - \left(\frac{\pi_1 - \pi_2 - \Delta}{z_{1-\alpha} + z_{1-\beta}} \right)^2 \right]. \quad (12)$$

Setting

$$\nabla_{(m,n)} L(m, n, \lambda) = 0,$$

we obtain

$$m = \sqrt{\frac{\lambda\pi_1(1 - \pi_1)}{c_1}} \text{ and } n = \sqrt{\frac{\lambda\pi_2(1 - \pi_2)}{c_2}}, \quad (13)$$

which is the same as Eq. (8). Plugging Eq. (13) into Eq. (11), we obtain

$$\sqrt{\frac{c_1\pi_1(1 - \pi_1)}{\lambda}} + \sqrt{\frac{c_2\pi_2(1 - \pi_2)}{\lambda}} - \left(\frac{\pi_1 - \pi_2 - \Delta}{z_{1-\alpha} + z_{1-\beta}} \right)^2 = 0,$$

which implies that

$$\sqrt{\lambda} = \frac{\sqrt{c_1\pi_1(1 - \pi_1)} + \sqrt{c_2\pi_2(1 - \pi_2)}}{\left(\frac{\pi_1 - \pi_2 - \Delta}{z_{1-\alpha} + z_{1-\beta}} \right)^2}. \quad (14)$$

By substituting Eq. (14) into Eq. (13), we obtain the solutions in **Theorem 2.** Note that the Hessian matrix for the Lagrangian function Eq. (12) is exactly the same as that in **Theorem 1**, and the solution in **Theorem 2** is the only KKT point, which again shows that the solution is the global minimizer of the cost Eq. (4) under the constraint Eq. (11).

Example 2. Non-inferiority. Under the signed-margin convention, suppose that the non-inferiority margin is $\Delta = -10\%$ and the expected cure rates for the disease in the intervention and control groups are 80% and 75%, respectively. Then, the sample sizes for equal allocation with 80% power at the 0.05 significance level are, according to Chow et al.^[7],

$$m = n = \frac{(z_{1-\alpha} + z_{1-\beta})^2 [\pi_1(1 - \pi_1) + \pi_2(1 - \pi_2)]}{(\pi_1 - \pi_2 - \Delta)^2} = \frac{(1.64 + 0.84)^2 [0.8(1 - 0.8) + 0.75(1 - 0.75)]}{(0.8 - 0.75 - (-0.1))^2} \approx 95.$$

If the per-subject costs in the intervention and control groups are \$100 and \$800, respectively, then the total cost is

$$C = 95 \times 100 + 95 \times 800 = 85,500.$$

Applying **Theorem 2** yields

$$m = \frac{[0.8(1 - 0.8)]^{\frac{1}{2}} \left([100 \times 0.8(1 - 0.8)]^{\frac{1}{2}} + [800 \times 0.75(1 - 0.75)]^{\frac{1}{2}} \right)}{100^{\frac{1}{2}} \left(\frac{0.8 - 0.75 + 0.1}{1.64 + 0.84} \right)^2} \approx 178,$$

$$n = \frac{[0.75(1 - 0.75)]^{\frac{1}{2}} \left([100 \times 0.8(1 - 0.8)]^{\frac{1}{2}} + [800 \times 0.75(1 - 0.75)]^{\frac{1}{2}} \right)}{800^{\frac{1}{2}} \left(\frac{0.8 - 0.75 + 0.1}{1.64 + 0.84} \right)^2} \approx 68$$

and the corresponding cost is

$$C^* = 178 \times 100 + 68 \times 800 = 72,200.$$

The relative cost reduction is therefore

$$\frac{C - C^*}{C} = \frac{85,500 - 72,200}{85,500} \approx 0.1556.$$

Therefore, our method reduces the total cost by 15.56% relative to equal allocation.

Example 3. Superiority. Assume that the superiority margin is 5% and the cure rates are 80% and 65% for the intervention and control groups, respectively. The per-subject costs are \$800 and \$200 for the intervention and control groups, respectively. Suppose the objective is to test the superiority of the intervention over the control treatment with 80% power at a one-sided significance level of 0.05. Then, the sample sizes with equal allocation, according to Chow et al.^[7], are

$$m = n = \frac{(z_{1-\alpha} + z_{1-\beta})^2 [\pi_1(1 - \pi_1) + \pi_2(1 - \pi_2)]}{(\pi_1 - \pi_2 - \Delta)^2} = \frac{(1.64 + 0.84)^2 [0.8(1 - 0.8) + 0.65(1 - 0.65)]}{(0.8 - 0.65 - 0.05)^2} \approx 238$$

and the total cost is

$$C = 238 \times 800 + 238 \times 200 = 238,000.$$

Incorporating the unequal unit costs, **Theorem 2** gives

$$m = \frac{[0.8(1 - 0.8)]^{\frac{1}{2}} \left([800 \times 0.8(1 - 0.8)]^{\frac{1}{2}} + [200 \times 0.65(1 - 0.65)]^{\frac{1}{2}} \right)}{800^{\frac{1}{2}} \left(\frac{0.8 - 0.65 - 0.05}{1.64 + 0.84} \right)^2} \approx 157,$$

and

$$n = \frac{[0.65(1 - 0.65)]^{\frac{1}{2}} \left([800 \times 0.8(1 - 0.8)]^{\frac{1}{2}} + [200 \times 0.65(1 - 0.65)]^{\frac{1}{2}} \right)}{200^{\frac{1}{2}} \left(\frac{0.8 - 0.65 - 0.05}{1.64 + 0.84} \right)^2} \approx 375,$$

resulting in a total cost of

$$C^* = 157 \times 800 + 375 \times 200 = 200,600.$$

Therefore,

$$\frac{C - C^*}{C} = \frac{238,000 - 200,600}{238,000} \approx 0.1571.$$

Hence, our method reduces the total cost by 15.71% relative to equal allocation.

Cost minimization for equivalence tests

The equivalence hypotheses are formulated as Chow et al.^[7]:

$$H_0 : |\pi_1 - \pi_2| \geq \Delta \text{ versus } H_1 : |\pi_1 - \pi_2| < \Delta, \quad (15)$$

where Δ is the equivalence margin, which is usually specified by investigators rather than statisticians.

Theorem 3. The sample sizes for hypotheses (15) to achieve the global minimum cost with the $1 - \beta$ at a significance level α are

$$m = \frac{[\pi_1(1 - \pi_1)]^{\frac{1}{2}} \left([c_1\pi_1(1 - \pi_1)]^{\frac{1}{2}} + [c_2\pi_2(1 - \pi_2)]^{\frac{1}{2}} \right)}{c_1^{\frac{1}{2}} \left(\frac{\Delta - |\pi_1 - \pi_2|}{z_{1-\alpha} + z_{1-\beta/2}} \right)^2}$$

and

$$n = \frac{[\pi_2(1 - \pi_2)]^{\frac{1}{2}} \left([c_1\pi_1(1 - \pi_1)]^{\frac{1}{2}} + [c_2\pi_2(1 - \pi_2)]^{\frac{1}{2}} \right)}{c_2^{\frac{1}{2}} \left(\frac{\Delta - |\pi_1 - \pi_2|}{z_{1-\alpha} + z_{1-\beta/2}} \right)^2}$$

Proof. The proof is similar to that of **Theorem 2** and is therefore omitted.

Example 4. Equivalence. Assume that the cure rates are 75% ($\pi_1 = 0.75$) for the intervention and 80% for the control group ($\pi_2 = 0.80$), and the equivalence margin or limit is $\Delta = 20\%$. The intervention is easier to administer and less expensive than the control group. The per-subject costs for the intervention and control groups are \$100 and \$900, respectively. First, if we use equal allocation without considering costs and calculate the sample size for 80% power at the 0.05 significance level, we have^[7]

$$m = n = \frac{(z_{1-\alpha} + z_{1-\beta/2})^2 (\pi_1(1 - \pi_1) + \pi_2(1 - \pi_2))}{(\Delta - |\pi_1 - \pi_2|)^2} \\ = \frac{(1.64 + 1.28)^2 (0.75 \times 0.25 + 0.8 \times 0.2)}{(0.2 - |0.75 - 0.8|)^2} \approx 132.$$

The total cost is

$$C = 132 \times 100 + 132 \times 900 = 132,000.$$

Incorporating the unequal unit costs and applying **Theorem 3**, we obtain

$$m = \frac{[0.75 \times 0.25]^{\frac{1}{2}} \left([100 \times 0.75 \times 0.25]^{\frac{1}{2}} + [900 \times 0.8 \times 0.2]^{\frac{1}{2}} \right)}{100^{\frac{1}{2}} \left(\frac{0.2 - |0.75 - 0.8|}{1.64 + 1.28} \right)^2} \approx 268$$

and

$$n = \frac{[0.8 \times 0.2]^{\frac{1}{2}} \left([100 \times 0.75 \times 0.25]^{\frac{1}{2}} + [900 \times 0.8 \times 0.2]^{\frac{1}{2}} \right)}{900^{\frac{1}{2}} \left(\frac{0.2 - |0.75 - 0.8|}{1.64 + 1.28} \right)^2} \approx 83.$$

Therefore,

$$C^* = 268 \times 100 + 83 \times 900 = 101,500$$

and

$$\frac{C - C^*}{C} = \frac{132,000 - 101,500}{132,000} \approx 0.2311,$$

which implies that our method reduces the total cost by 23.11% relative to equal allocation.

Fisher information

The power constraints considered above can be interpreted in terms of inverse Fisher information. For the two independent Bernoulli samples, the Fisher information for π_1 and π_2 is^[9]

$$I_1(\pi_1) = \frac{m}{\pi_1(1 - \pi_1)}, \quad I_2(\pi_2) = \frac{n}{\pi_2(1 - \pi_2)}.$$

Therefore,

$$I_1^{-1}(\pi_1) = \frac{\pi_1(1 - \pi_1)}{m}, \quad I_2^{-1}(\pi_2) = \frac{\pi_2(1 - \pi_2)}{n}.$$

Thus, the normal-approximation power constraint fixes $I_1^{-1} + I_2^{-1}$. The allocation problem can therefore be viewed as minimizing cost subject to a fixed inverse-information requirement.

Theorem 4.1. The sum of the inverses Fisher information terms for hypotheses Eq. (5) with power $1 - \beta$ at a significance level α is

$$\left(\frac{\pi_1 - \pi_2}{z_{1-\alpha/2} + z_{1-\beta}} \right)^2.$$

Proof. Since Fisher information $I_1(\pi_1) = \frac{m}{\pi_1(1 - \pi_1)}$ and $I_2(\pi_2) = \frac{n}{\pi_2(1 - \pi_2)}$, and using Eq. (6), we have

$$I_1^{-1} + I_2^{-1} = \frac{\pi_1(1 - \pi_1)}{m} + \frac{\pi_2(1 - \pi_2)}{n} = \left(\frac{\pi_1 - \pi_2}{z_{1-\alpha/2} + z_{1-\beta}} \right)^2.$$

Theorem 4.2. The sum of the inverse Fisher information terms for hypotheses Eq. (10) with power $1 - \beta$ at a significance level α is

$$\left(\frac{\pi_1 - \pi_2 - \Delta}{z_{1-\alpha} + z_{1-\beta}} \right)^2.$$

Proof. Using Eq. (11), we have

$$I_1^{-1} + I_2^{-1} = \frac{\pi_1(1 - \pi_1)}{m} + \frac{\pi_2(1 - \pi_2)}{n} = \left(\frac{\pi_1 - \pi_2 - \Delta}{z_{1-\alpha} + z_{1-\beta}} \right)^2.$$

Theorem 4.3. The sum of the inverses of Fisher information terms for hypotheses (15) with power $1 - \beta$ at a significance level α is:

$$\left(\frac{\Delta - |\pi_1 - \pi_2|}{z_{1-\alpha} + z_{1-\beta/2}} \right)^2.$$

Proof. Similarly, we have

$$I_1^{-1} + I_2^{-1} = \frac{\pi_1(1 - \pi_1)}{m} + \frac{\pi_2(1 - \pi_2)}{n} = \left(\frac{\Delta - |\pi_1 - \pi_2|}{z_{1-\alpha} + z_{1-\beta/2}} \right)^2.$$

The sum of the inverse Fisher information terms for the two-sided equality, non-inferiority/superiority, and equivalence tests is fixed under the corresponding normal-approximation power constraint.

Simulation

We conducted two simulation studies using SAS[®]. The first simulation corresponds to Example 1. We used rand('BINOMIAL', p, n) to generate 1,000 binomial random samples under $n_1 = n_2 = 135$, $\pi_1 = 0.8$, and $\pi_2 = 0.65$, and also generated 1,000 replicates using rand('BINOMIAL', 0.80, 89) and rand('BINOMIAL', 0.65, 212). The power difference between the two comparisons was 0.8%. Therefore, the simulation results were consistent with the theoretical calculation.

The second simulation corresponds to Example 4. We generated 1,000 replicates under the equal allocation $(m, n) = (132, 132)$ and another 1,000 replicates under the optimized allocation $(m, n) = (268, 83)$. For each replicate, we conducted the prespecified equivalence test and estimated empirical power as the proportion of rejections. The empirical power difference was 3.2%; this discrepancy should be further evaluated with a larger number of simulations and by checking the effect of integer rounding.

These preliminary simulations suggest that the optimized allocations can achieve empirical power close to that of equal allocation while reducing cost. However, more extensive simulations with a larger number of Monte Carlo replicates are needed to quantify the effect of integer rounding and normal-approximation error.

Discussion

One limitation is that all results are derived under large-sample assumptions, so they may not apply to small samples. The corresponding optimization problems are more difficult for small samples because sample sizes must take integer values. The proposed large-sample formulas may be less suitable for rare-disease trials, where recruitment constraints and integer sample-size effects often play a more prominent role. These formulas may be useful at the planning stage of comparative studies in which the unit costs differ substantially between treatment arms and recruitment is not the primary limiting factor.

Conclusions

Financial constraints play an important role in clinical trial design. We derived closed-form formulas for continuous sample-size allocations for comparing two independent binary outcomes under fixed power and a fixed significance level. Because unit costs may differ between study arms, equal allocation does not necessarily minimize total variable cost. The formulas provide useful continuous allocations at the design stage. In practical applications, the rounded integer allocations should be checked against the target power. Although this paper focuses on comparing two proportions, similar cost-minimization ideas may be extended to other endpoints, such as means, relative risks, survival outcomes, and correlations.

Author contributions

The authors confirm contribution to the paper as follows: study conception and design: Luo J (www.researchgate.net/profile/Jiangtao-Luo-2), Liu CM, Pant MD, El Moudden I, Wang Z; theorem proof and algorithm: Luo J, Liu CM, Wang Y, Wang Z, Zhang H; draft manuscript preparation: Luo J, Liu CM, Wang Z, Zhang H. All authors reviewed the results and approved the final version of the manuscript.

Data availability

No real datasets were analyzed. Simulation code and generated data are available from the corresponding author, Jiangtao Luo, upon reasonable request.

Acknowledgments

Thanks are given to Dr. Rongling Wu for his suggestions.

Conflict of interest

The authors declare that they have no conflict of interest.

Dates

Received 18 August 2025; Revised 10 January 2026; Accepted 9 February 2026; Published online 23 June 2026

References

- [1] Brittain E, Schlesselman JJ. 1982. Optimal allocation for the comparison of proportions. *Biometrics* 38(4):1003
- [2] Allison DB, Allison RL, Faith MS, Paultre F, Pi-Sunyer FX. 1997. Power and money: designing statistically powerful studies while minimizing financial costs. *Psychological Methods* 2(1):20–33
- [3] Cochran WG. 1977. *Sampling Techniques*. John Wiley & Sons
- [4] Guo JH, Chen HJ, Luh WM. 2011. Sample size planning with the cost constraint for testing superiority and equivalence of two independent groups. *British Journal of Mathematical and Statistical Psychology* 64(3):439–461
- [5] Guo JH, Luh WM. 2009. Optimum sample size allocation to minimize cost or maximize power for the two-sample trimmed mean test. *British Journal of Mathematical and Statistical Psychology* 62(2):283–298
- [6] Luo J, Wang Y, Meza J. 2015. Optimal sample size allocation under financial constraint. *Austin Biometrics and Biostatistics* 2(3):1024
- [7] Chow SC, Shao J, Wang H, Lokhnygina Y. 2017. *Sample Size Calculations in Clinical Research*. 3rd Edition. Boca Raton: Chapman and Hall/CRC. doi: 10.1201/9781315183084
- [8] Bertsekas DP. 1997. Nonlinear programming. *Journal of the Operational Research Society* 48(3):334
- [9] Lehmann EL, Casella G. 1998. *Theory of Point Estimation*. New York, NY: Springer New York. doi: 10.1007/b98854



Copyright: © 2026 by the author(s). Published by Maximum Academic Press, Fayetteville, GA. This article is an open access article distributed under Creative Commons Attribution License (CC BY 4.0), visit <https://creativecommons.org/licenses/by/4.0/>.