AIMS Mathematics, 11(7): 20224–20253
DOI:10.3934/math.2026821
Received: 27 April 2026
Revised: 08 June 2026
Accepted: 02 July 2026
Published: 09 July 2026

https://www.aimspress.com/journal/Math

*Research article*

# Dunnett-type saddlepoint simultaneous confidence intervals for many-to-one comparisons of proportion differences with correlated paired binary data

**Juanjuan Zhang**[1], **Yujuan Luo**[2] **and Weixian Wang**[2,3,*]

[1] School of Digital Economy and Trade, Guangzhou Huashang College, Guangzhou 511300, China
[2] School of Mathematics and Statistics, Guangxi Normal University, Guilin 541006, China
[3] Center for Applied Mathematics of Guangxi, Guangxi Normal University, Guilin 541006, China

* **Correspondence:** Email: wangweixian@email.cufe.edu.cn.

**Abstract:** Correlated paired binary data often arise in biomedical studies involving paired organs, where each subject may contribute zero, one, or two positive responses. This paper considers simultaneous confidence interval construction for many-to-one comparisons of proportion differences between several comparison groups and a common control. Existing methods, including method of variance estimates recovery (MOVER), Wald-type, profile likelihood, score, and Normal-Dunnett procedures, mainly rely on first-order normal or chi-square approximations, whose finite-sample accuracy may be affected by sparse counts, unbalanced designs, strong within-subject dependence, or an increasing number of groups. We propose a Dunnett-type saddlepoint simultaneous confidence interval method under the constant-$R$ model. By using the three-category structure of paired binary responses, the cumulant generating function of the treatment-control contrast vector is derived explicitly and used for saddlepoint calibration of the joint distribution. Simulation studies show that the proposed method improves simultaneous coverage accuracy in sparse and strongly dependent settings, while remaining comparable to existing methods in moderate settings. A retinitis pigmentosa dataset is analyzed to illustrate the proposed procedure.

**Keywords:** correlated paired binary data; saddlepoint approximation; simultaneous confidence interval; many-to-one comparison; proportion difference; constant-$R$ model
**Mathematics Subject Classification:** 62F12, 62G10

## 1. Introduction

Correlated paired binary data frequently arise in ophthalmologic, otolaryngologic, dental, orthopaedic, and other biomedical studies, where binary responses are collected from paired organs or paired body parts of the same subject. Typical examples include responses from two eyes, two ears,

two teeth, or two limbs. In such studies, each subject may contribute no response, a unilateral response, or bilateral responses, and the two outcomes from the same subject are usually correlated. Ignoring this intra-subject dependence may lead to biased variance estimation, inaccurate test statistics, and misleading confidence intervals. Therefore, appropriate statistical methods are required to account for the correlation structure in paired binary responses. Rosner [1] proposed the constant-$R$ model for ophthalmologic data, in which the conditional response probability of one eye given a response in the other eye is proportional to the marginal response probability. Dallal [2] studied paired Bernoulli trials from a different modeling perspective, while Donner [3] developed an adjusted chi-square approach for correlated bilateral binary data by incorporating intraclass correlation.

Based on these dependence models, considerable attention has been devoted to statistical inference for correlated paired or bilateral binary data. Tang et al. [4] investigated statistical inference procedures for correlated ophthalmologic data and discussed the importance of accounting for the within-subject association. For otolaryngologic studies, Tang et al. [5] considered tests for the equality of proportions under correlated binary responses. Ma et al. [6] further studied homogeneity tests for correlated binary data and examined their finite-sample performance. In addition to hypothesis testing, confidence interval construction has also been extensively investigated. Tang et al. [7] developed asymptotic confidence intervals for the difference between two proportions in medical studies with bilateral data, whereas Peng et al. [8] considered confidence interval estimation for proportion ratios based on correlated paired data. These studies provide important tools for single comparison problems involving correlated paired binary outcomes.

In modern multi-arm clinical and biomedical studies, however, the inferential target is often not a single comparison, but a set of many-to-one comparisons between several treatment or disease groups and a common control. Dunnett's multiple comparison procedure [9] provides a classical framework for comparing several treatments with a control while controlling the family-wise error rate. For dichotomous responses, Piegorsch [10] extended multiple comparison ideas to binary data analysis. More recently, simultaneous inference for correlated paired or bilateral binary data has received increasing attention. Yang et al. [11] constructed simultaneous confidence intervals for many-to-one comparisons of proportion differences based on correlated paired data under the constant-$R$ model. Zhang and Ma [12] studied homogeneity testing for differences between two proportions in stratified bilateral and unilateral data. Liang et al. [13] further developed homogeneity tests and interval estimation methods for risk differences in stratified bilateral and unilateral correlated data. For other effect measures, Zhang and Ma [14] considered simultaneous confidence intervals for many-to-one comparisons of proportion ratios, Hua and Ma [15] investigated common odds ratio tests and interval estimation, and Wang and Ma [16] studied interval estimation of relative risks for combined unilateral and bilateral correlated data.

Despite these developments, most existing simultaneous inference procedures for many-to-one comparisons in correlated paired binary data are still built upon first-order normal, multivariate normal, or chi-square approximations. For example, Normal-Dunnett-type (ND-type) procedures approximate the joint distribution of treatment-control contrast estimators by a multivariate normal distribution and determine the simultaneous critical value from this first-order approximation. Likelihood ratio, profile likelihood, and score-based procedures can improve upon simple Wald-type intervals by using the likelihood structure more effectively; however, their calibration is still typically based on asymptotic quadratic expansions of the log-likelihood or chi-square approximations to the

corresponding test statistics. Although these approximations are asymptotically valid and computationally convenient, their accuracy may be unsatisfactory in finite samples, especially when the sample size is small, the group sizes are highly unbalanced, the cell counts are sparse, the response probabilities are close to the boundary of the parameter space, or the within-subject dependence is strong. These situations are commonly encountered in biomedical studies involving bilateral organs, where the number of subjects in each treatment group may be limited and the dependence between paired responses cannot be ignored.

Saddlepoint approximation provides a powerful higher-order alternative to conventional asymptotic approximations by directly exploiting the cumulant generating function of the statistic of interest. The classical foundations of saddlepoint approximation were established by Daniels [17] and Lugannani and Rice [18], and systematic reviews and theoretical treatments were given by Reid [19], Jensen [20], and Kolassa [21]. In recent years, saddlepoint methods have received renewed attention in modern statistical applications where tail probability accuracy is essential. For example, Zhou et al. [22] used saddlepoint approximation to calibrate score tests in large-scale genetic association studies with severe case-control imbalance and sample relatedness. Bi et al. [23] developed SPACox for genome-wide time-to-event analysis and showed that saddlepoint calibration improves the accuracy of rare-variant association tests. Johnsen et al. [24] studied saddlepoint approximations to score test statistics in logistic regression for binary genome-wide association studies and showed that the normal approximation can be inaccurate when responses are imbalanced or minor allele counts are small. Newer [25] applied a double saddlepoint approximation to weighted log-rank tests for censored clustered data and demonstrated improved accuracy over asymptotic normal approximation. More recently, Ma et al. [26] incorporated saddlepoint approximation into SPAmix, a scalable framework for large-scale genetic association studies involving complex traits and admixed populations.

These recent developments show that saddlepoint approximation is particularly useful in problems involving sparse counts, imbalanced outcomes, rare events, clustered dependence, or extreme tail probabilities. However, most existing saddlepoint applications focus on single test statistics, score tests, likelihood quantities, or genome-wide association testing. Relatively little attention has been paid to simultaneous confidence interval construction for many-to-one comparisons in correlated paired binary data. This gap is important because the problem considered here involves two sources of dependence: The within-subject correlation between paired binary responses and the between-comparison dependence induced by sharing a common control group. Accurate simultaneous inference therefore requires both multiplicity adjustment and finite-sample tail probability calibration.

Motivated by these considerations, this paper develops a Dunnett-type saddlepoint approximation (DSPA) procedure for simultaneous confidence interval construction for many-to-one comparisons of proportion differences in correlated paired binary data. The key methodological novelty of the proposed DSPA method is that it does not approximate the treatment-control contrasts only through a first-order multivariate normal distribution, nor does it rely solely on chi-square calibration of likelihood or score statistics. Instead, the proposed method starts from the exact cumulant generating function of the relevant risk-difference contrast statistics under the correlated paired binary data model. The associated saddlepoint equations are then solved to obtain saddlepoint-calibrated tail probabilities, which are further incorporated into the Dunnett-type simultaneous inference framework. In this way, the proposed method combines the dependence-adjusted multiplicity structure of

Dunnett's procedure with a higher-order approximation to the finite-sample distribution of the contrast statistics.

Compared with the existing ND approach, the proposed DSPA method uses the exact cumulant generating function and saddlepoint calibration to approximate joint tail probabilities more accurately. Compared with likelihood-, profile-likelihood-, and score-based approaches, it provides a direct higher-order calibration for the distribution of the contrast statistics rather than relying only on asymptotic chi-square approximations. Therefore, the proposed method is expected to improve finite-sample simultaneous inference while accounting for both within-subject correlation and between-comparison dependence. Simulation studies and a real data analysis are conducted to evaluate the empirical performance of the proposed procedure in comparison with existing methods.

## 2. Correlated paired binary data and model formulation

### *2.1. Data structure and notation*

Correlated paired binary data commonly arise in biomedical studies where responses are collected from paired organs or paired body parts of the same subject. Examples include ophthalmologic studies involving two eyes, otolaryngologic studies involving two ears, dental studies involving paired sites, and orthopaedic studies involving paired joints. In such settings, the two responses from the same subject are usually correlated, and this within-subject dependence should be explicitly incorporated into statistical inference.

Suppose that there are $g$ independent groups. Without loss of generality, group 1 is taken as the control group, and groups $2, \ldots, g$ are treatment or comparison groups. For the $i$th group, let $m_i$ denote the number of subjects. For subject $j$ in group $i$, define

$$Z_{ijk} = \begin{cases} 1, & \text{if the } k\text{th paired organ has a positive response,} \\ 0, & \text{otherwise,} \end{cases} \qquad i = 1, \ldots, g, \;\; j = 1, \ldots, m_i, \;\; k = 1, 2.$$

The marginal response probability in group $i$ is denoted by

$$\pi_i = P(Z_{ijk} = 1), \quad i = 1, \ldots, g.$$

For each subject, define the total number of positive responses from the paired organs as

$$L_{ij} = Z_{ij1} + Z_{ij2}.$$

Then,

$$L_{ij} \in \{0, 1, 2\}.$$

Specifically, $L_{ij} = 0$ means that neither paired organ responds, $L_{ij} = 1$ means that exactly one paired organ responds, and $L_{ij} = 2$ means that both paired organs respond.

Let

$$m_{li} = \sum_{j=1}^{m_i} I(L_{ij} = l), \quad l = 0, 1, 2, \quad i = 1, \ldots, g,$$

where $m_{li}$ is the number of subjects in group $i$ with $l$ positive responses. Therefore, the observed data can be summarized by Table 1.

**Table 1.** Observed response counts across $g$ groups.

| Number of responses | 1 | 2 | $\cdots$ | g | Total |
|---|---|---|---|---|---|
| 0 | $m_{01}$ | $m_{02}$ | $\cdots$ | $m_{0g}$ | $S_0$ |
| 1 | $m_{11}$ | $m_{12}$ | $\cdots$ | $m_{1g}$ | $S_1$ |
| 2 | $m_{21}$ | $m_{22}$ | $\cdots$ | $m_{2g}$ | $S_2$ |
| Total | $m_1$ | $m_2$ | $\cdots$ | $m_g$ | $N$ |

where

$$S_l = \sum_{i=1}^{g} m_{li}, \quad l = 0, 1, 2,$$

and

$$N = \sum_{i=1}^{g} m_i.$$

Under the assumption that subjects are independent within and between groups, the count vector for group $i$ follows a multinomial distribution:

$$(m_{0i}, m_{1i}, m_{2i})^\top \sim \text{Multinomial}\,(m_i; p_{0i}, p_{1i}, p_{2i}),$$

where

$$p_{li} = P(L_{ij} = l), \quad l = 0, 1, 2.$$

The primary inferential goal is to compare each treatment group with the common control group in terms of the marginal response probabilities. For $i = 2, \ldots, g$, define the many-to-one proportion difference

$$\Delta_i = \pi_i - \pi_1.$$

The objective is to construct simultaneous confidence intervals for

$$\Delta_2, \Delta_3, \ldots, \Delta_g.$$

### *2.2. The constant-R model*

To model the dependence between the two paired binary responses from the same subject, we consider Rosner's constant-$R$ model. Under this model,

$$P(Z_{ijk} = 1) = \pi_i$$

and

$$P(Z_{ijk} = 1 \mid Z_{ij,3-k} = 1) = R\pi_i,$$

where $R > 0$ is a dependence parameter shared by all groups.

The parameter $R$ describes the association between the two paired responses. When $R = 1$, the two paired responses are independent. When $R > 1$, the paired responses are positively associated, subject to the admissible parameter space. When $R < 1$, the paired responses tend to be negatively associated.

Under the constant-$R$ model, the cell probabilities for group $i$ are

$$p_{0i} = P(L_{ij} = 0) = R\pi_i^2 - 2\pi_i + 1,$$

$$p_{1i} = P(L_{ij} = 1) = 2\pi_i(1 - R\pi_i)$$

and

$$p_{2i} = P(L_{ij} = 2) = R\pi_i^2.$$

It is straightforward to verify that

$$p_{0i} + p_{1i} + p_{2i} = 1.$$

The admissible parameter space is determined by

$$0 < \pi_i < 1, \quad p_{0i} \geq 0, \quad p_{1i} \geq 0, \quad p_{2i} \geq 0, \quad i = 1, \ldots, g.$$

In particular, $p_{1i} \geq 0$ implies

$$R\pi_i \leq 1.$$

The within-subject correlation between the two paired binary responses in group $i$ is

$$\rho_i = \mathrm{corr}(Z_{ij1}, Z_{ij2}) = \frac{\pi_i}{1 - \pi_i}(R - 1), \quad i = 1, \ldots, g.$$

Thus, the constant-$R$ model links the marginal response probability $\pi_i$ with the dependence structure of the paired responses.

### *2.3. Likelihood function and reparameterization*

Let

$$\boldsymbol{\theta} = (\pi_1, \ldots, \pi_g, R)^\top.$$

Since the $g$ groups are independent, the log-likelihood function, up to an additive constant, is

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{g} \ell_i(\pi_i, R),$$

where

$$\ell_i(\pi_i, R) = m_{0i} \log(R\pi_i^2 - 2\pi_i + 1) + m_{1i} \log\{2\pi_i(1 - R\pi_i)\} + m_{2i} \log(R\pi_i^2).$$

Equivalently,

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{g} \left[ m_{0i} \log p_{0i} + m_{1i} \log p_{1i} + m_{2i} \log p_{2i} \right].$$

For many-to-one comparisons, the parameter of interest is

$$\Delta_i = \pi_i - \pi_1, \quad i = 2, \ldots, g.$$

For a fixed comparison between group $i$ and the control group, we may reparameterize

$$\pi_i = \pi_1 + \Delta_i.$$

For example, for $i = 2$,

$$\pi_2 = \pi_1 + \Delta_2.$$

Then, the log-likelihood can be written as

$$\ell_2(\Delta_2; \pi_1, \pi_3, \ldots, \pi_g, R) = \ell_1(\pi_1, R) + \ell_2(\pi_1 + \Delta_2, R) + \sum_{r=3}^{g} \ell_r(\pi_r, R),$$

where

$$\begin{aligned}\ell_2(\pi_1 + \Delta_2, R) =& m_{02} \log\{R(\pi_1 + \Delta_2)^2 - 2(\pi_1 + \Delta_2) + 1\} \\ &+ m_{12} \log\{2(\pi_1 + \Delta_2)[1 - R(\pi_1 + \Delta_2)]\} + m_{22} \log\{R(\pi_1 + \Delta_2)^2\}.\end{aligned}$$

In this representation, $\Delta_2$ is the parameter of interest, while

$$(\pi_1, \pi_3, \ldots, \pi_g, R)$$

are nuisance parameters. The same reparameterization can be used for any comparison $i = 2, \ldots, g$.

### *2.4. Multiplicity adjustment for many-to-one comparisons*

The purpose of simultaneous confidence interval construction is to find intervals

$$CI_2, \ldots, CI_g$$

such that the family-wise coverage probability satisfies

$$P(\Delta_i \in CI_i,\ i = 2, \ldots, g) \geq 1 - \alpha.$$

If multiplicity is ignored and each interval is constructed separately at level $1 - \alpha$, the overall coverage probability may fall below the nominal level. Therefore, multiplicity adjustment is essential for many-to-one comparisons.

A simple approach is Bonferroni adjustment. In this case, each individual confidence interval is constructed at the level

$$1 - \frac{\alpha}{g - 1}.$$

For two-sided intervals, the corresponding critical value is

$$c_B = z_{1-\alpha/[2(g-1)]}.$$

Although the Bonferroni method is easy to implement and guarantees family-wise coverage control, it may be conservative, especially when the number of treatment groups is large or the test statistics are strongly correlated.

For multiple comparisons with a common control, Dunnett-type adjustment provides a more efficient alternative. Let

$$\widehat{\pi}_i = \frac{m_{1i} + 2m_{2i}}{2m_i}, \quad i = 1, \ldots, g,$$

and

$$\widehat{\Delta}_i = \widehat{\pi}_i - \widehat{\pi}_1, \quad i = 2, \ldots, g.$$

The treatment-control contrasts

$$\widehat{\Delta}_2, \ldots, \widehat{\Delta}_g$$

are correlated because they share the same control estimator $\widehat{\pi}_1$.

Under an asymptotic multivariate normal approximation, the correlation between two treatment-control contrasts can be approximated by

$$\rho_{ij} = w_i w_j, \quad i, j = 2, \ldots, g, \quad i \neq j,$$

where

$$w_i = \left[1 + \frac{m_1 \pi_i (1 - \pi_i)}{m_i \pi_1 (1 - \pi_1)}\right]^{-1/2}.$$

In practice, the unknown $\pi_i$'s are replaced by their consistent estimators. Let

$$\mathbf{R}_D = (\rho_{ij})_{i,j=2}^{g}$$

denote the resulting $(g-1) \times (g-1)$ correlation matrix, with diagonal entries equal to one.

The Dunnett-type critical value $c_D$ is defined by

$$P\left(\max_{2 \leq i \leq g} |Z_i| \leq c_D\right) = 1 - \alpha,$$

where

$$(Z_2, \ldots, Z_g)^\top \sim N_{g-1}(\mathbf{0}, \mathbf{R}_D).$$

Compared with Bonferroni adjustment, Dunnett-type adjustment explicitly accounts for the correlation among treatment-control comparisons and is usually less conservative.

In this paper, the Dunnett-type dependence structure plays a central role. The existing methods reviewed in the next section use this structure through asymptotic normal or chi-square approximations, whereas the proposed method replaces such first-order approximations with saddlepoint approximation to improve finite-sample accuracy.

## 3. Simultaneous confidence interval methods

This section reviews existing simultaneous confidence interval methods for many-to-one comparisons of proportion differences in correlated paired binary data. These methods include method of variance estimates recovery (MOVER)-type intervals, Wald-type intervals, profile likelihood intervals, and score intervals. They provide the methodological baseline for the proposed Dunnett-type saddlepoint procedure developed in the next section.

Throughout this section, the parameter of interest is

$$\Delta_i = \pi_i - \pi_1, \quad i = 2, \ldots, g.$$

The resulting simultaneous confidence intervals are restricted to the natural parameter space

$$-1 \leq \Delta_i \leq 1.$$

### 3.1. MOVER simultaneous confidence intervals

The method of variance estimates recovery, abbreviated as MOVER, constructs a confidence interval for the difference between two parameters by recovering variance information from the individual confidence intervals of each parameter.

For the comparison between group $i$ and the control group, suppose that individual confidence intervals for $\pi_i$ and $\pi_1$ are

$$(l_i, u_i) \quad \text{and} \quad (l_1, u_1),$$

respectively. The estimator of the proportion difference is

$$\widehat{\Delta}_i = \widehat{\pi}_i - \widehat{\pi}_1,$$

where

$$\widehat{\pi}_i = \frac{m_{1i} + 2m_{2i}}{2m_i}.$$

The MOVER lower limit for $\Delta_i$ is

$$L_i^{\text{MOVER}} = \widehat{\pi}_i - \widehat{\pi}_1 - \sqrt{(\widehat{\pi}_i - l_i)^2 + (u_1 - \widehat{\pi}_1)^2},$$

and the upper limit is

$$U_i^{\text{MOVER}} = \widehat{\pi}_i - \widehat{\pi}_1 + \sqrt{(u_i - \widehat{\pi}_i)^2 + (\widehat{\pi}_1 - l_1)^2}.$$

Therefore, the MOVER simultaneous confidence interval is

$$CI_i^{\text{MOVER}} = \left[\max\{-1, L_i^{\text{MOVER}}\}, \min\{1, U_i^{\text{MOVER}}\}\right], \quad i = 2, \ldots, g.$$

Two types of individual confidence intervals for $\pi_i$ are commonly used.

First, the Wilson score interval is obtained by defining

$$\widetilde{n}_i = 2m_i + c^2,$$

$$\widehat{\pi}_i = \frac{m_{1i} + 2m_{2i}}{2m_i},$$

and

$$\widetilde{\pi}_i = \frac{m_{1i} + 2m_{2i} + c^2/2}{2m_i + c^2},$$

where $c$ is the multiplicity-adjusted critical value. The Wilson score interval for $\pi_i$ is

$$\widetilde{\pi}_i \mp \frac{c}{\widetilde{n}_i} \sqrt{2m_i\widehat{\pi}_i(1 - \widehat{\pi}_i) + \frac{c^2}{4}}.$$

The MOVER simultaneous confidence interval based on this individual interval is referred to as MOVER1.

Second, the Agresti–Coull interval is

$$\widetilde{\pi}_i \mp c \sqrt{\frac{\widetilde{\pi}_i(1 - \widetilde{\pi}_i)}{\widetilde{n}_i}}.$$

The MOVER simultaneous confidence interval based on the Agresti–Coull interval is referred to as MOVER2.

The MOVER methods are simple and computationally convenient. However, they are constructed from marginal intervals and do not directly use the likelihood structure of the correlated paired data. Their performance may therefore be sensitive to the strength of within-subject dependence.

### 3.2. Wald-type simultaneous confidence intervals

#### 3.2.1. Wald-type interval under the null hypothesis

An intuitive estimator of $\Delta_i$ is

$$\widehat{\Delta}_i = \frac{m_{1i} + 2m_{2i}}{2m_i} - \frac{m_{11} + 2m_{21}}{2m_1} = \widehat{\pi}_i - \widehat{\pi}_1.$$

Under the null hypothesis

$$H_{0i} : \Delta_i = 0,$$

the asymptotic variance of $\widehat{\Delta}_i$ can be estimated by

$$\widehat{\mathrm{Var}}(\widehat{\Delta}_i) = \frac{\widehat{\pi}_i\{1 + (\widehat{R}_i - 2)\widehat{\pi}_i\}}{2m_i} + \frac{\widehat{\pi}_1\{1 + (\widehat{R}_i - 2)\widehat{\pi}_1\}}{2m_1},$$

where

$$\widehat{R}_i = \frac{4(m_1 + m_i)(m_{21} + m_{2i})}{(m_{11} + m_{1i} + 2m_{21} + 2m_{2i})^2}.$$

The Wald-type simultaneous confidence interval under the null hypothesis is then

$$CI_i^{\mathrm{Wald1}} = \left[\max\left\{-1, \widehat{\Delta}_i - c\sqrt{\widehat{\mathrm{Var}}(\widehat{\Delta}_i)}\right\}, \min\left\{1, \widehat{\Delta}_i + c\sqrt{\widehat{\mathrm{Var}}(\widehat{\Delta}_i)}\right\}\right], \quad i = 2, \ldots, g.$$

Here, $c$ may be chosen as the Bonferroni or Dunnett-type critical value.

#### 3.2.2. Wald-type interval under the dependence assumption

The second Wald-type method directly uses the maximum likelihood estimators under the constant-$R$ dependence model. Let

$$\boldsymbol{\beta} = (\pi_1, \ldots, \pi_g, R),$$

and let

$$\widehat{\boldsymbol{\beta}} = (\widehat{\pi}_1, \ldots, \widehat{\pi}_g, \widehat{R})$$

be the unconstrained maximum likelihood estimator.

Define the contrast matrix

$$\mathbf{K} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 & 0 \\ -1 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ -1 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix}.$$

For $i = 2, \ldots, g$, the treatment-control contrast can be written as

$$\Delta_i = \mathbf{K}_{i\cdot}\boldsymbol{\beta}^\top,$$

where $\mathbf{K}_{i\cdot}$ denotes the $i$th row of $\mathbf{K}$. Therefore,

$$\widehat{\Delta}_i = \mathbf{K}_{i\cdot}\widehat{\boldsymbol{\beta}}^\top.$$

By the asymptotic normality of the maximum likelihood estimator,

$$\widehat{\boldsymbol{\beta}} \approx N\left(\boldsymbol{\beta}, \mathbf{I}^{-1}(\boldsymbol{\beta})\right),$$

where $\mathbf{I}(\boldsymbol{\beta})$ is the Fisher information matrix. Hence,

$$\widehat{\mathrm{Var}}(\widehat{\Delta}_i) = \mathbf{K}_{i\cdot}\widehat{\mathbf{I}}^{-1}\mathbf{K}_{i\cdot}^{\top}.$$

The Wald-type simultaneous confidence interval under the dependence assumption is

$$CI_i^{\mathrm{Wald2}} = \left[\max\left\{-1, \mathbf{K}_{i\cdot}\widehat{\boldsymbol{\beta}}^{\top} - c\sqrt{\mathbf{K}_{i\cdot}\widehat{\mathbf{I}}^{-1}\mathbf{K}_{i\cdot}^{\top}}\right\}, \min\left\{1, \mathbf{K}_{i\cdot}\widehat{\boldsymbol{\beta}}^{\top} + c\sqrt{\mathbf{K}_{i\cdot}\widehat{\mathbf{I}}^{-1}\mathbf{K}_{i\cdot}^{\top}}\right\}\right].$$

This method incorporates the dependence structure through the maximum likelihood estimation of the constant-$R$ model. Nevertheless, it still relies on a first-order normal approximation.

### *3.3. Profile likelihood simultaneous confidence intervals*

Profile likelihood intervals are obtained by inverting likelihood ratio tests. For a fixed comparison $i$, consider the null hypothesis

$$H_{0i} : \Delta_i = \delta$$

against the two-sided alternative. Let

$$\widetilde{\boldsymbol{\eta}}_i(\delta)$$

be the constrained maximum likelihood estimator of the nuisance parameters under $H_{0i}$, and let

$$\widehat{\boldsymbol{\theta}}$$

be the unconstrained maximum likelihood estimator under the full model.

The likelihood ratio statistic is

$$LR_i(\delta) = 2\left[\ell(\widehat{\boldsymbol{\theta}}) - \ell\{\delta, \widetilde{\boldsymbol{\eta}}_i(\delta)\}\right].$$

Under regularity conditions,

$$LR_i(\delta) \overset{a}{\sim} \chi_1^2.$$

Therefore, the profile likelihood confidence interval for $\Delta_i$ is

$$CI_i^{\mathrm{PL}} = \left\{\delta \in [-1, 1] : LR_i(\delta) \le \chi^2_{1-\alpha/(g-1)}(1)\right\}.$$

Equivalently,

$$\ell(\widehat{\boldsymbol{\theta}}) - \ell\{\delta, \widetilde{\boldsymbol{\eta}}_i(\delta)\} \le \frac{1}{2}\chi^2_{1-\alpha/(g-1)}(1).$$

In practical computation, the lower and upper endpoints are obtained by solving

$$\ell(\widehat{\boldsymbol{\theta}}) - \ell\{\delta, \widetilde{\boldsymbol{\eta}}_i(\delta)\} = \frac{1}{2}\chi^2_{1-\alpha/(g-1)}(1)$$

over the admissible range

$$-1 \le \delta \le 1.$$

The constrained nuisance parameters are usually obtained by numerical maximization or Fisher scoring.

The profile likelihood method is often more stable than simple Wald-type intervals, especially when the likelihood is asymmetric. However, its calibration is still based on the first-order chi-square approximation to the likelihood ratio statistic.

### 3.4. Score simultaneous confidence intervals

Score intervals are obtained by inverting score tests. For a fixed comparison $i$, consider again

$$H_{0i} : \Delta_i = \delta.$$

Let

$$\boldsymbol{\theta}_i = (\Delta_i, \boldsymbol{\eta}_i^\top)^\top$$

be the reparameterized vector, where $\boldsymbol{\eta}_i$ contains the nuisance parameters. Let

$$\widetilde{\boldsymbol{\theta}}_i(\delta) = \{\delta, \widetilde{\boldsymbol{\eta}}_i(\delta)\}$$

be the constrained maximum likelihood estimator under $H_{0i}$.

The score component for $\Delta_i$ is

$$U_{\Delta_i}(\delta) = \left.\frac{\partial \ell(\Delta_i, \boldsymbol{\eta}_i)}{\partial \Delta_i}\right|_{\boldsymbol{\theta}_i = \widetilde{\boldsymbol{\theta}}_i(\delta)}.$$

Let

$$\mathbf{I}\{\widetilde{\boldsymbol{\theta}}_i(\delta)\}$$

be the Fisher information matrix evaluated at the constrained maximum likelihood estimator. If $\Delta_i$ is arranged as the first component of $\boldsymbol{\theta}_i$, then the score test statistic can be written as

$$T_i^{\mathrm{SC}}(\delta) = \{U_{\Delta_i}(\delta)\}^2 \left[\mathbf{I}^{-1}\{\widetilde{\boldsymbol{\theta}}_i(\delta)\}\right]_{11}.$$

Here,

$$\left[\mathbf{I}^{-1}\{\widetilde{\boldsymbol{\theta}}_i(\delta)\}\right]_{11}$$

is the first diagonal element of the inverse Fisher information matrix.

Under regularity conditions,

$$T_i^{\mathrm{SC}}(\delta) \overset{a}{\sim} \chi_1^2.$$

Therefore, the score confidence interval for $\Delta_i$ is

$$CI_i^{\mathrm{SC}} = \left\{\delta \in [-1, 1] : T_i^{\mathrm{SC}}(\delta) \le \chi^2_{1-\alpha/(g-1)}(1)\right\}.$$

The endpoints can be found by solving

$$T_i^{\mathrm{SC}}(\delta) = \chi^2_{1-\alpha/(g-1)}(1)$$

over $\delta \in [-1, 1]$.

The score method is particularly important for the development of the proposed saddlepoint procedure. The reason is that the score statistic can be expressed as a sum of subject-level or cell-level contributions. This additive structure makes it possible to derive its cumulant generating function, which is the key ingredient of saddlepoint approximation.

## 4. Dunnett-type saddlepoint simultaneous confidence intervals

This section develops a Dunnett-type saddlepoint simultaneous confidence interval procedure for many-to-one comparisons of proportion differences in correlated paired binary data. The proposed method is motivated by two observations. First, the treatment-control contrasts

$$\widehat{\Delta}_i = \widehat{\pi}_i - \widehat{\pi}_1, \qquad i = 2, \ldots, g,$$

are mutually correlated because all contrasts share the same control group. Therefore, a simultaneous confidence interval procedure should account for the dependence among these contrasts. Second, under the constant-$R$ model, the number of positive responses from each subject takes values in $\{0, 1, 2\}$, which allows the cumulant generating function of the treatment-control contrast vector to be derived explicitly. This provides a natural basis for saddlepoint approximation.

Let

$$q = g - 1$$

be the number of many-to-one comparisons, and define

$$\widehat{\mathbf{\Delta}} = (\widehat{\Delta}_2, \ldots, \widehat{\Delta}_g)^\top, \quad \mathbf{\Delta} = (\Delta_2, \ldots, \Delta_g)^\top.$$

The goal is to construct simultaneous confidence intervals

$$CI_2, \ldots, CI_g$$

such that

$$P\left(\Delta_i \in CI_i,\ i = 2, \ldots, g\right) \approx 1 - \alpha.$$

### *4.1. Cumulant generating function of the contrast vector*

For group $i$, recall that

$$L_{ij} = Z_{ij1} + Z_{ij2} \in \{0, 1, 2\}.$$

Under the constant-$R$ model,

$$\begin{aligned} P(L_{ij} = 0) &= p_{0i} = R\pi_i^2 - 2\pi_i + 1, \\ P(L_{ij} = 1) &= p_{1i} = 2\pi_i(1 - R\pi_i), \end{aligned}$$

and

$$P(L_{ij} = 2) = p_{2i} = R\pi_i^2.$$

Define the moment generating function of $L_{ij}$ in group $i$ by

$$M_i(u) = E\{\exp(uL_{ij})\} = p_{0i} + p_{1i}e^u + p_{2i}e^{2u}.$$

The corresponding cumulant generating function is

$$\kappa_i(u) = \log M_i(u) = \log\left(p_{0i} + p_{1i}e^u + p_{2i}e^{2u}\right).$$

The sample proportion estimator in group $i$ is

$$\widehat{\pi}_i = \frac{1}{2m_i}\sum_{j=1}^{m_i} L_{ij} = \frac{m_{1i} + 2m_{2i}}{2m_i}.$$

Therefore, for $i = 2, \ldots, g$,

$$\widehat{\Delta}_i = \frac{1}{2m_i}\sum_{j=1}^{m_i} L_{ij} - \frac{1}{2m_1}\sum_{j=1}^{m_1} L_{1j}.$$

Let

$$\mathbf{t} = (t_2, \ldots, t_g)^\top \in \mathbb{R}^q.$$

The cumulant generating function of

$$\widehat{\boldsymbol{\Delta}} = (\widehat{\Delta}_2, \ldots, \widehat{\Delta}_g)^\top$$

is

$$K_{\widehat{\Delta}}(\mathbf{t}; \boldsymbol{\theta}) = \log E_{\boldsymbol{\theta}}\left[\exp\left\{\sum_{i=2}^{g} t_i \widehat{\Delta}_i\right\}\right],$$

where

$$\boldsymbol{\theta} = (\pi_1, \ldots, \pi_g, R)^\top.$$

Since the treatment groups are independent and the common control group contributes to every contrast, we obtain

$$K_{\widehat{\Delta}}(\mathbf{t}; \boldsymbol{\theta}) = m_1 \kappa_1\left(-\frac{1}{2m_1}\sum_{i=2}^{g} t_i\right) + \sum_{i=2}^{g} m_i \kappa_i\left(\frac{t_i}{2m_i}\right).$$

Equivalently,

$$K_{\widehat{\Delta}}(\mathbf{t}; \boldsymbol{\theta}) = m_1 \log\left[p_{01} + p_{11}\exp\left\{-\frac{1}{2m_1}\sum_{i=2}^{g} t_i\right\} + p_{21}\exp\left\{-\frac{1}{m_1}\sum_{i=2}^{g} t_i\right\}\right]$$

$$+ \sum_{i=2}^{g} m_i \log\left[p_{0i} + p_{1i}\exp\left(\frac{t_i}{2m_i}\right) + p_{2i}\exp\left(\frac{t_i}{m_i}\right)\right].$$

This joint cumulant generating function is the key ingredient of the proposed method. It automatically incorporates the dependence among $\widehat{\Delta}_2, \ldots, \widehat{\Delta}_g$ because the control group term depends on the sum $\sum_{i=2}^{g} t_i$. Thus, the correlation among multiple treatment-control comparisons is not imposed through a first-order multivariate normal approximation, but is directly induced by the exact cumulant generating function of the contrast vector.

### *4.2. Mean and variance-covariance matrix*

From the cumulant generating function, the mean vector of $\widehat{\boldsymbol{\Delta}}$ is

$$E(\widehat{\boldsymbol{\Delta}}) = \nabla K_{\widehat{\Delta}}(\mathbf{0}; \boldsymbol{\theta}) = \boldsymbol{\Delta}.$$

The variance-covariance matrix is

$$\mathbf{V}(\boldsymbol{\theta}) = \nabla^2 K_{\widehat{\Delta}}(\mathbf{0}; \boldsymbol{\theta}).$$

For $i = 2, \ldots, g$, the diagonal elements are

$$V_{ii} = \mathrm{Var}(\widehat{\Delta}_i) = \frac{\pi_i\{1 + (R-2)\pi_i\}}{2m_i} + \frac{\pi_1\{1 + (R-2)\pi_1\}}{2m_1}.$$

For $i \neq j$, $i, j = 2, \ldots, g$, the off-diagonal elements are

$$V_{ij} = \mathrm{Cov}(\widehat{\Delta}_i, \widehat{\Delta}_j) = \frac{\pi_1\{1 + (R-2)\pi_1\}}{2m_1}.$$

The off-diagonal covariance arises because all treatment-control contrasts share the same control estimator $\widehat{\pi}_1$.

Let

$$\sigma_i^2 = V_{ii}, \quad i = 2, \ldots, g,$$

and define

$$\mathbf{D} = \mathrm{diag}(\sigma_2^{-1}, \ldots, \sigma_g^{-1}).$$

The standardized contrast vector is

$$\mathbf{X} = \mathbf{D}\left(\widehat{\boldsymbol{\Delta}} - \boldsymbol{\Delta}\right).$$

The cumulant generating function of $\mathbf{X}$ is then

$$K_X(\mathbf{t}; \boldsymbol{\theta}) = K_{\widehat{\Delta}}(\mathbf{Dt}; \boldsymbol{\theta}) - (\mathbf{Dt})^\top \boldsymbol{\Delta}.$$

By construction,

$$E(\mathbf{X}) = \mathbf{0},$$

and

$$\mathrm{Var}(\mathbf{X}) = \mathbf{DV}(\boldsymbol{\theta})\mathbf{D}.$$

The corresponding correlation matrix has entries

$$\mathrm{corr}(X_i, X_j) = \frac{V_{ij}}{\sqrt{V_{ii}V_{jj}}}, \quad i, j = 2, \ldots, g.$$

### *4.3. Multivariate saddlepoint approximation*

Let $f_X(\mathbf{x}; \boldsymbol{\theta})$ denote the joint density or probability mass approximation of $\mathbf{X}$. For a given point

$$\mathbf{x} = (x_2, \ldots, x_g)^\top,$$

the multivariate saddlepoint $\widehat{\mathbf{t}} = \widehat{\mathbf{t}}(\mathbf{x})$ is defined as the solution to

$$\nabla K_X(\widehat{\mathbf{t}}; \boldsymbol{\theta}) = \mathbf{x}.$$

That is,

$$\left.\frac{\partial K_X(\mathbf{t}; \boldsymbol{\theta})}{\partial \mathbf{t}}\right|_{\mathbf{t}=\widehat{\mathbf{t}}} = \mathbf{x}.$$

Once the saddlepoint $\widehat{\mathbf{t}}$ is obtained, the multivariate saddlepoint approximation to the joint density of $\mathbf{X}$ is

$$\widehat{f}_{\mathrm{SPA}}(\mathbf{x};\boldsymbol{\theta}) = (2\pi)^{-q/2}\left|\nabla^2 K_X(\widehat{\mathbf{t}};\boldsymbol{\theta})\right|^{-1/2}\exp\left\{K_X(\widehat{\mathbf{t}};\boldsymbol{\theta}) - \widehat{\mathbf{t}}^{\top}\mathbf{x}\right\}.$$

Here,

$$\nabla^2 K_X(\widehat{\mathbf{t}};\boldsymbol{\theta})$$

is the Hessian matrix of the cumulant generating function evaluated at the saddlepoint. This approximation uses the full cumulant generating function and therefore incorporates higher-order information beyond the mean and covariance matrix.

For numerical implementation, the saddlepoint equation can be solved by Newton–Raphson iterations:

$$\mathbf{t}^{(r+1)} = \mathbf{t}^{(r)} - \left[\nabla^2 K_X(\mathbf{t}^{(r)};\boldsymbol{\theta})\right]^{-1}\left[\nabla K_X(\mathbf{t}^{(r)};\boldsymbol{\theta}) - \mathbf{x}\right].$$

A natural starting value is $\mathbf{t}^{(0)} = \mathbf{0}$, especially when $\mathbf{x}$ is close to the mean. If $\mathbf{x} = \mathbf{0}$, then $\widehat{\mathbf{t}} = \mathbf{0}$.

### *4.4. Dunnett-type saddlepoint critical value*

The classical Dunnett procedure constructs simultaneous confidence intervals by finding a critical value for

$$\max_{2\le i\le g}|X_i|.$$

Under a multivariate normal approximation, the critical value $c_D$ is determined by

$$P\left(\max_{2\le i\le g}|X_i| \le c_D\right) = 1-\alpha,$$

where $\mathbf{X}$ is approximated by a multivariate normal distribution.

In the proposed method, we replace the multivariate normal approximation with a saddlepoint approximation. For any $c > 0$, define the hyper-rectangle

$$A(c) = [-c, c]^q.$$

The Dunnett-type saddlepoint approximation to the joint probability is

$$P_{\mathrm{SPA}}(c;\boldsymbol{\theta}) = P_{\mathrm{SPA}}\{\mathbf{X}\in A(c)\} = \int_{A(c)}\widehat{f}_{\mathrm{SPA}}(\mathbf{x};\boldsymbol{\theta})\,d\mathbf{x}.$$

That is,

$$P_{\mathrm{SPA}}(c;\boldsymbol{\theta}) = \int_{-c}^{c}\cdots\int_{-c}^{c}\widehat{f}_{\mathrm{SPA}}(x_2,\ldots,x_g;\boldsymbol{\theta})\,dx_2\cdots dx_g.$$

The proposed saddlepoint critical value $c_{\mathrm{SPA}}$ is defined as the solution to

$$P_{\mathrm{SPA}}(c_{\mathrm{SPA}};\widehat{\boldsymbol{\theta}}) = 1-\alpha,$$

where

$$\widehat{\boldsymbol{\theta}} = (\widehat{\pi}_1,\ldots,\widehat{\pi}_g,\widehat{R})^{\top}$$

is the unrestricted maximum likelihood estimator under the constant-$R$ model.

Equivalently,

$$c_{\mathrm{SPA}} = \inf\left\{c > 0 : P_{\mathrm{SPA}}(c; \widehat{\boldsymbol{\theta}}) \geq 1 - \alpha\right\}.$$

The integral over $A(c)$ can be evaluated by adaptive numerical integration when $q$ is small or moderate. For larger $q$, quasi-Monte Carlo integration can be used. Specifically, if

$$\mathbf{u}^{(b)} \in [0,1]^q, \quad b = 1, \ldots, B,$$

are quasi-Monte Carlo points, define

$$\mathbf{x}^{(b)} = -c\mathbf{1}_q + 2c\mathbf{u}^{(b)}.$$

Then

$$P_{\mathrm{SPA}}(c; \widehat{\boldsymbol{\theta}}) \approx (2c)^q \frac{1}{B} \sum_{b=1}^{B} \widehat{f}_{\mathrm{SPA}}\left(\mathbf{x}^{(b)}; \widehat{\boldsymbol{\theta}}\right).$$

The critical value $c_{\mathrm{SPA}}$ can then be obtained by a one-dimensional root-finding method, such as bisection.

### *4.5. Construction of simultaneous confidence intervals*

Using the saddlepoint critical value $c_{\mathrm{SPA}}$, we construct the proposed Dunnett-type saddlepoint simultaneous confidence interval for

$$\Delta_i = \pi_i - \pi_1$$

as

$$CI_i^{\mathrm{DSPA}} = \left[L_i^{\mathrm{DSPA}}, U_i^{\mathrm{DSPA}}\right], \quad i = 2, \ldots, g,$$

where

$$L_i^{\mathrm{DSPA}} = \max\left\{-1, \widehat{\Delta}_i - c_{\mathrm{SPA}}\widehat{\sigma}_i\right\}$$

and

$$U_i^{\mathrm{DSPA}} = \min\left\{1, \widehat{\Delta}_i + c_{\mathrm{SPA}}\widehat{\sigma}_i\right\}.$$

Here,

$$\widehat{\sigma}_i^2 = V_{ii}(\widehat{\boldsymbol{\theta}}) = \frac{\widehat{\pi}_i\{1 + (\widehat{R} - 2)\widehat{\pi}_i\}}{2m_i} + \frac{\widehat{\pi}_1\{1 + (\widehat{R} - 2)\widehat{\pi}_1\}}{2m_1}.$$

Thus, the proposed interval has the same rectangular form as a Dunnett-type simultaneous confidence interval, but the critical value is calibrated by saddlepoint approximation rather than by a multivariate normal approximation. This is the central difference between the proposed method and the existing Dunnett-type Wald procedure.

The proposed simultaneous confidence intervals satisfy

$$P\left(L_i^{\mathrm{DSPA}} \leq \Delta_i \leq U_i^{\mathrm{DSPA}},\ i = 2, \ldots, g\right) \approx 1 - \alpha.$$

The approximation is expected to be more accurate than first-order normal calibration because $c_{\mathrm{SPA}}$ is determined from the cumulant generating function of the exact contrast vector.

### *4.6. Computational implementation*

For reproducibility, we provide the main numerical details used to implement the proposed DSPA procedure. The saddlepoint equations were solved using a bracketing root-finding algorithm. The convergence tolerance was set to $10^{-8}$, and the maximum number of iterations was set to 100. Convergence was declared when the absolute residual of the saddlepoint equation was smaller than $10^{-8}$ or when the absolute change in the root between two successive iterations was smaller than $10^{-8}$. To obtain the Dunnett-type simultaneous confidence intervals, the simultaneous critical value was computed by a bisection search. The iteration was stopped when the absolute difference between the target family-wise confidence level and the computed simultaneous probability was less than $10^{-6}$ or when the length of the bisection interval was smaller than $10^{-6}$.

The simultaneous tail probability required in the Dunnett-type calibration was evaluated by randomized quasi-Monte Carlo integration based on Sobol sequences. In all simulation studies and the real data analysis, 20,000 quasi-Monte Carlo points were used. To improve numerical stability, the cumulant generating function and its derivatives were evaluated on the logarithmic scale whenever necessary, and boundary values of estimated probabilities were truncated to lie in a small compact subset of $(0, 1)$. All computations were implemented in R version 4.3.2. The root-finding and bisection procedures were implemented using standard R numerical routines, and the quasi-Monte Carlo integration was carried out using Sobol sequence-based numerical integration.

### *4.7. Boundary and sparse-count cases.*

In sparse-count settings, some observed cell counts may be zero. Such zero counts do not directly invalidate the proposed DSPA construction, since the cumulant generating function is determined by the fitted probability model rather than by the empirical cell frequencies alone. However, sparse data may lead to maximum likelihood estimates close to the boundary of the admissible parameter space. In particular, when the marginal response probabilities are close to 0 or 1, or when the within-subject dependence is strong, the likelihood surface may become flat and the estimate of the dependence parameter in the constant-$R$ model may become unstable. To improve numerical stability, all parameter estimates are constrained to remain in the admissible parameter space. In the numerical implementation, marginal probabilities are restricted to a small compact subset of $(0, 1)$, and the dependence parameter is constrained so that all joint probabilities are nonnegative. If an unconstrained estimate approaches the boundary, it is replaced by the nearest admissible value within the constrained parameter space. This treatment avoids degenerate probabilities when evaluating the cumulant generating function, its derivatives, and the saddlepoint equations. In addition, convergence of the likelihood maximization and saddlepoint root-finding procedures is monitored. These safeguards are particularly useful in small-sample or sparse-count scenarios, where first-order normal or chi-square approximations may be unreliable.

## 5. Simulation studies

This section evaluates the finite-sample performance of the proposed method through Monte Carlo simulations. The methods compared include the MOVER method based on the Wilson score interval (MOVER1), the MOVER method based on the Agresti–Coull interval (MOVER2), the Wald-type

interval under the null hypothesis (Wald1), the Wald-type interval under the constant-$R$ dependence assumption (Wald2), the profile likelihood interval (profile), the score interval (score), the classical ND simultaneous confidence interval (ND), and the proposed Dunnett-type saddlepoint simultaneous confidence interval (DSPA).

The target parameters are the many-to-one proportion differences

$$\Delta_i = \pi_i - \pi_1, \quad i = 2, \ldots, g.$$

The nominal simultaneous confidence level is set to 0.95. For each simulation setting, 500 Monte Carlo replications are conducted. All constructed confidence intervals are truncated to the natural parameter space

$$-1 \leq \Delta_i \leq 1.$$

### *5.1. Simulation settings*

We consider two groups of simulation settings with $g = 3, 5, 10$. The first group of settings is summarized in Table 2, and the second group of settings is summarized in Table 3.

**Table 2.** Simulation settings I.

| Scenario | $g$ | $(m_1, \ldots, m_g)$ | $(\pi_1, \ldots, \pi_g), R$ |
|---|---|---|---|
| F3 | 3 | $(12, 8, 20)$ | $(0.04, 0.08, 0.14),\ R = 5.0$ |
| F5 | 5 | $(12, 10, 15, 12, 20)$ | $(0.03, 0.04, 0.06, 0.09, 0.13),\ R = 5.5$ |
| F10 | 10 | $(12, 8, 10, 14, 12, 16, 20, 18, 10, 15)$ | $(0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.10, 0.11),\ R = 6.0$ |

**Table 3.** Simulation settings II.

| Scenario | $g$ | $(m_1, \ldots, m_g)$ | $(\pi_1, \ldots, \pi_g), R$ |
|---|---|---|---|
| M3 | 3 | $(40, 45, 50)$ | $(0.25, 0.30, 0.35),\ R = 1.5$ |
| M5 | 5 | $(40, 45, 50, 55, 60)$ | $(0.25, 0.28, 0.32, 0.36, 0.40),\ R = 1.5$ |
| M10 | 10 | $(50, 50, 50, 50, 50, 50, 50, 50, 50, 50)$ | $(0.20, 0.22, 0.24, 0.26, 0.28, 0.30, 0.32, 0.34, 0.36, 0.38),\ R = 1.2$ |

### *5.2. Evaluation criteria*

For the $b$th simulated dataset, let

$$CI_i^{(b)} = [L_i^{(b)}, U_i^{(b)}], \quad i = 2, \ldots, g,$$

be the confidence interval for $\Delta_i$. The empirical simultaneous coverage probability is defined as

$$\text{SCI\_ECP} = \frac{1}{500} \sum_{b=1}^{500} I\left(\Delta_i \in CI_i^{(b)},\ i = 2, \ldots, g\right).$$

This quantity estimates the family-wise coverage probability.

The average marginal coverage probability is defined as

$$\text{MeanECP} = \frac{1}{g-1} \sum_{i=2}^{g} \frac{1}{500} \sum_{b=1}^{500} I\left(\Delta_i \in CI_i^{(b)}\right),$$

and the minimum marginal coverage probability is defined as

$$\text{MinECP} = \min_{2\le i\le g} \frac{1}{500}\sum_{b=1}^{500} I\left(\Delta_i \in CI_i^{(b)}\right).$$

The mean interval width is calculated by

$$\text{MIW} = \frac{1}{g-1}\sum_{i=2}^{g}\frac{1}{500}\sum_{b=1}^{500}\left(U_i^{(b)} - L_i^{(b)}\right).$$

To assess the Monte Carlo uncertainty of the empirical simulation results, we also report the Monte Carlo standard error for the empirical simultaneous coverage probability. Let $B$ denote the number of simulation replications and let $\widehat{p}$ be the empirical simultaneous coverage probability. The Monte Carlo standard error is computed as

$$\text{MCSE}(\widehat{p}) = \left(\frac{\widehat{p}(1-\widehat{p})}{B}\right)^{1/2}.$$

### *5.3. Simulation results*

The simulation results under setting I are reported in Tables 4–6, and those under setting II are reported in Tables 7–9. For the empirical coverage probabilities $SCI\_ECP$, MeanECP, and MinECP, the values in parentheses below the estimates are the corresponding MCSEs defined in Section 5.2. The MIW column reports the mean interval width.

**Table 4.** Simulation results for scenario F3 under setting I based on 500 replications.

| Method | SCI_ECP | MeanECP | MinECP | MIW |
|---|---|---|---|---|
| MOVER1 | 0.962 | 0.981 | 0.964 | 0.4166 |
| | (0.009) | (0.006) | (0.008) | |
| MOVER2 | 0.982 | 0.991 | 0.982 | 0.4568 |
| | (0.006) | (0.004) | (0.006) | |
| Wald1 | 0.830 | 0.911 | 0.858 | 0.3714 |
| | (0.017) | (0.013) | (0.016) | |
| Wald2 | 0.992 | 0.996 | 0.992 | 0.4395 |
| | (0.004) | (0.003) | (0.004) | |
| Profile | 0.580 | 0.758 | 0.660 | 0.1994 |
| | (0.022) | (0.019) | (0.021) | |
| Score | 0.580 | 0.758 | 0.660 | 0.2003 |
| | (0.022) | (0.019) | (0.021) | |
| ND | 0.944 | 0.972 | 0.966 | 0.3839 |
| | (0.010) | (0.007) | (0.008) | |
| DSPA | 0.950 | 0.975 | 0.970 | 0.4033 |
| | (0.010) | (0.007) | (0.008) | |

**Table 5.** Simulation results for scenario F5 under setting I based on 500 replications.

| Method | SCI_ECP | MeanECP | MinECP | MIW |
|---|---|---|---|---|
| MOVER1 | 0.978 | 0.995 | 0.980 | 0.4428 |
| | (0.007) | (0.003) | (0.006) | |
| MOVER2 | 0.990 | 0.998 | 0.990 | 0.4922 |
| | (0.004) | (0.002) | (0.004) | |
| Wald1 | 0.652 | 0.882 | 0.766 | 0.3359 |
| | (0.021) | (0.014) | (0.019) | |
| Wald2 | 0.988 | 0.997 | 0.988 | 0.4139 |
| | (0.005) | (0.002) | (0.005) | |
| Profile | 0.348 | 0.742 | 0.568 | 0.1682 |
| | (0.021) | (0.020) | (0.022) | |
| Score | 0.344 | 0.741 | 0.564 | 0.1704 |
| | (0.021) | (0.020) | (0.022) | |
| ND | 0.938 | 0.985 | 0.974 | 0.3462 |
| | (0.011) | (0.005) | (0.007) | |
| DSPA | 0.958 | 0.990 | 0.986 | 0.3579 |
| | (0.009) | (0.004) | (0.005) | |

**Table 6.** Simulation results for scenario F10 under setting I based on 500 replications.

| Method | SCI_ECP | MeanECP | MinECP | MIW |
|---|---|---|---|---|
| MOVER1 | 1.000 | 1.000 | 1.000 | 0.5084 |
| | (0.000) | (0.000) | (0.000) | |
| MOVER2 | 1.000 | 1.000 | 1.000 | 0.5690 |
| | (0.000) | (0.000) | (0.000) | |
| Wald1 | 0.362 | 0.826 | 0.592 | 0.3319 |
| | (0.021) | (0.017) | (0.022) | |
| Wald2 | 1.000 | 1.000 | 1.000 | 0.4370 |
| | (0.000) | (0.000) | (0.000) | |
| Profile | 0.068 | 0.668 | 0.558 | 0.1475 |
| | (0.011) | (0.021) | (0.022) | |
| Score | 0.068 | 0.668 | 0.558 | 0.1475 |
| | (0.011) | (0.021) | (0.022) | |
| ND | 0.920 | 0.990 | 0.984 | 0.3571 |
| | (0.012) | (0.004) | (0.006) | |
| DSPA | 0.944 | 0.993 | 0.984 | 0.4107 |
| | (0.010) | (0.004) | (0.006) | |

**Table 7.** Simulation results for scenario M3 under setting II based on 500 replications.

| Method | SCI_ECP | MeanECP | MinECP | MIW |
|---|---|---|---|---|
| MOVER1 | 0.908 | 0.946 | 0.944 | 0.2969 |
| | (0.013) | (0.010) | (0.010) | |
| MOVER2 | 0.910 | 0.948 | 0.948 | 0.2988 |
| | (0.013) | (0.010) | (0.010) | |
| Wald1 | 0.944 | 0.968 | 0.966 | 0.3317 |
| | (0.010) | (0.008) | (0.008) | |
| Wald2 | 0.946 | 0.969 | 0.966 | 0.3330 |
| | (0.010) | (0.008) | (0.008) | |
| Profile | 0.944 | 0.969 | 0.966 | 0.3303 |
| | (0.010) | (0.008) | (0.008) | |
| Score | 0.942 | 0.968 | 0.966 | 0.3291 |
| | (0.010) | (0.008) | (0.008) | |
| ND | 0.946 | 0.969 | 0.964 | 0.3322 |
| | (0.010) | (0.008) | (0.008) | |
| DSPA | 0.946 | 0.969 | 0.964 | 0.3292 |
| | (0.010) | (0.008) | (0.008) | |

**Table 8.** Simulation results for scenario M5 under setting II based on 500 replications.

| Method | SCI_ECP | MeanECP | MinECP | MIW |
|---|---|---|---|---|
| MOVER1 | 0.926 | 0.979 | 0.976 | 0.3180 |
| | (0.012) | (0.006) | (0.007) | |
| MOVER2 | 0.932 | 0.980 | 0.978 | 0.3204 |
| | (0.011) | (0.006) | (0.007) | |
| Wald1 | 0.958 | 0.988 | 0.984 | 0.3567 |
| | (0.009) | (0.005) | (0.006) | |
| Wald2 | 0.958 | 0.988 | 0.984 | 0.3580 |
| | (0.009) | (0.005) | (0.006) | |
| Profile | 0.962 | 0.989 | 0.984 | 0.3607 |
| | (0.009) | (0.005) | (0.006) | |
| Score | 0.960 | 0.989 | 0.982 | 0.3594 |
| | (0.009) | (0.005) | (0.006) | |
| ND | 0.960 | 0.989 | 0.984 | 0.3586 |
| | (0.009) | (0.005) | (0.006) | |
| DSPA | 0.960 | 0.989 | 0.984 | 0.3587 |
| | (0.009) | (0.005) | (0.006) | |

**Table 9.** Simulation results for scenario M10 under setting II based on 500 replications.

| Method | SCI_ECP | MeanECP | MinECP | MIW |
|---|---|---|---|---|
| MOVER1 | 0.946 | 0.994 | 0.988 | 0.3229 |
| | (0.010) | (0.003) | (0.005) | |
| MOVER2 | 0.952 | 0.994 | 0.988 | 0.3266 |
| | (0.010) | (0.003) | (0.005) | |
| Wald1 | 0.954 | 0.994 | 0.992 | 0.3381 |
| | (0.009) | (0.003) | (0.004) | |
| Wald2 | 0.956 | 0.995 | 0.992 | 0.3395 |
| | (0.009) | (0.003) | (0.004) | |
| Profile | 0.962 | 0.995 | 0.992 | 0.3399 |
| | (0.009) | (0.003) | (0.004) | |
| Score | 0.962 | 0.995 | 0.992 | 0.3377 |
| | (0.009) | (0.003) | (0.004) | |
| ND | 0.962 | 0.995 | 0.992 | 0.3397 |
| | (0.009) | (0.003) | (0.004) | |
| DSPA | 0.960 | 0.995 | 0.992 | 0.3396 |
| | (0.009) | (0.003) | (0.004) | |

Tables 4–6 show that the proposed DSPA method improves the simultaneous coverage accuracy of the classical ND method under setting I. In scenario F3, the $SCI\_ECP$ of ND is (0.944), slightly below the nominal level (0.95), whereas DSPA increases it to (0.950). In scenario F5, the $SCI\_ECP$ of ND is (0.938), while DSPA improves it to (0.958). In scenario F10, as the number of groups increases, the simultaneous coverage probability of ND decreases to (0.920), whereas DSPA still achieves a coverage probability of (0.944). These results indicate that the first-order multivariate normal approximation used by ND may underestimate the joint tail probability when multiple treatment groups are compared with a common control. In contrast, DSPA uses the cumulant generating function of the contrast vector and therefore provides a more accurate calibration of the finite-sample joint distribution.

In terms of interval width, DSPA produces slightly wider intervals than ND under setting I, but the increase in width is accompanied by improved coverage accuracy. For example, in scenario F5, the MIW of ND is (0.3462), whereas that of DSPA is (0.3579). Although the interval width increases only modestly, the $SCI\_ECP$ improves from (0.938) to (0.958). In scenario F10, the MIW of DSPA is (0.4107), larger than the MIW of ND, which is (0.3571), but the simultaneous coverage probability is also improved from (0.920) to (0.944). Therefore, DSPA provides a more reasonable trade-off between coverage accuracy and interval length in finite-sample settings.

Compared with MOVER1 and MOVER2, the main advantage of DSPA is that it avoids excessive conservativeness. The MOVER methods generally produce high simultaneous coverage probabilities under setting I. In particular, in scenario F10, both MOVER1 and MOVER2 attain $SCI\_ECP$ values equal to (1.000). However, this high coverage is achieved at the cost of much wider intervals. In F10, the MIWs of MOVER1 and MOVER2 are (0.5084) and (0.5690), respectively, whereas the MIW of DSPA is (0.4107). This suggests that the MOVER methods may ensure coverage by producing overly

wide intervals, while DSPA achieves a better balance between coverage probability and interval width.

The Wald1 method performs poorly under setting I. Its $SCI\_ECP$ decreases from (0.830) in F3 to (0.362) in F10, indicating that the simple null-based Wald variance approximation can lead to severe undercoverage when the sample size is small, the number of groups is large, or the within-subject dependence is strong. The profile and score methods also show substantial undercoverage under setting I, especially in F10, where their $SCI\_ECP$ values are only (0.068). This reflects the difficulty of likelihood-based and score-based inversion in highly sparse finite-sample settings. By contrast, DSPA maintains simultaneous coverage probabilities much closer to the nominal level across F3, F5, and F10, demonstrating the advantage of saddlepoint calibration.

Tables 7–9 show that, under Setting II, the differences among the competing methods become much smaller. In scenario M3, DSPA gives an $SCI\_ECP$ of (0.946), which is essentially the same as Wald2 and ND. In scenario M5, DSPA gives an $SCI\_ECP$ of (0.960), identical to ND and very close to the profile and score methods. In scenario M10, DSPA gives an $SCI\_ECP$ of (0.960), comparable to the best-performing existing methods. The reported MCSEs are around (0.009)–(0.010) for most $SCI\_ECP$ values close to the nominal level, indicating that small differences among these methods under setting II should be interpreted cautiously. These results indicate that when the sample size is moderate and the data are less sparse, DSPA remains competitive with standard asymptotic and likelihood-based procedures.

In terms of MIW under setting II, DSPA also shows competitive performance. In scenario M3, the MIW of DSPA is (0.3292), which is slightly smaller than those of ND, Wald1, Wald2, and Profile, and is almost identical to that of score. In scenarios M5 and M10, the MIWs of DSPA are very close to those of ND, Wald2, profile, and score. These findings suggest that when conventional asymptotic approximations are already reliable, the proposed saddlepoint method does not introduce noticeable extra conservativeness.

Overall, the simulation results demonstrate the main advantage of the proposed DSPA method. Compared with ND, DSPA improves simultaneous coverage accuracy, especially when the number of groups increases or the finite-sample distribution deviates from the multivariate normal approximation. Compared with MOVER-type methods, DSPA avoids overly conservative intervals. Compared with Wald-type, profile, and score methods, DSPA is more stable in difficult finite-sample settings. The reported MCSEs further help readers assess whether small differences in empirical coverage probabilities are practically meaningful under 500 simulation replications. These results support the use of Dunnett-type saddlepoint calibration as a higher-order simultaneous inference procedure for many-to-one comparisons of proportion differences in correlated paired binary data.

### *5.4. Computational efficiency*

We also considered the computational cost of the proposed DSPA method relative to the competing procedures. The Wald-type and MOVER-type methods are the fastest approaches because they only require closed-form variance estimates and direct interval construction. The ND method is also computationally efficient, although it requires evaluating a multivariate normal probability to obtain the simultaneous critical value. The Profile and Score methods are more computationally intensive because they involve numerical inversion of likelihood-based or score-based test statistics over a grid of candidate parameter values.

The proposed DSPA method requires additional numerical calculations, including repeated

evaluation of the cumulant generating function and its derivatives, solution of the saddlepoint equations, and calibration of the Dunnett-type simultaneous tail probability. Therefore, DSPA is computationally more demanding than the simple Wald, MOVER, and ND procedures. However, the computational burden remains moderate in the present setting because the saddlepoint equations are solved for low-dimensional treatment-control contrast statistics, and the numerical root-finding procedure converges rapidly in most simulated datasets. The computational cost increases with the number of treatment groups, but the method remains feasible for the scenarios considered in the simulation study. Thus, DSPA provides improved finite-sample coverage accuracy at the price of a moderate increase in computational cost.

## 6. Real data analysis

In this section, we apply the proposed method to the retinitis pigmentosa dataset analyzed in the existing study. The data were obtained from an outpatient population of subjects with retinitis pigmentosa. Each subject has two eyes, and the response for each subject is recorded as the number of affected eyes, which can be 0, 1, or 2. The subjects were classified into four genetic types: dominant (DOM), autosomal recessive (AR), sex-linked (SL), and isolate (ISO). Following the original analysis, DOM is treated as the control group, and AR, SL, and ISO are treated as comparison groups. Thus, we write

$$\pi_1 = \pi_{\mathrm{DOM}}, \quad \pi_2 = \pi_{\mathrm{AR}}, \quad \pi_3 = \pi_{\mathrm{SL}}, \quad \pi_4 = \pi_{\mathrm{ISO}}.$$

The parameters of interest are the many-to-one proportion differences

$$\Delta_2 = \pi_2 - \pi_1, \quad \Delta_3 = \pi_3 - \pi_1, \quad \Delta_4 = \pi_4 - \pi_1.$$

The observed data are summarized in Table 10. The same dataset was used in the previous study to illustrate simultaneous confidence interval construction for many-to-one comparisons of proportion differences under correlated paired binary data.

**Table 10.** Prevalence of affected eyes in the retinitis pigmentosa dataset.

| Number of affected eyes | DOM | AR | SL | ISO |
|---|---|---|---|---|
| 0 | 15 | 7 | 3 | 67 |
| 1 | 6 | 5 | 2 | 24 |
| 2 | 7 | 9 | 14 | 57 |
| Total | 28 | 21 | 19 | 148 |

Under the constant-$R$ model, the estimated dependence parameter is

$$\widehat{R} = 1.664,$$

which indicates a relatively strong positive correlation between the two eyes from the same subject. The estimated marginal response probabilities are

$$\widehat{\pi}_1 = 0.393, \qquad \widehat{\pi}_2 = 0.480, \qquad \widehat{\pi}_3 = 0.563, \qquad \widehat{\pi}_4 = 0.493.$$

Therefore, the estimated proportion differences are

$$\widehat{\Delta}_2 = 0.087, \qquad \widehat{\Delta}_3 = 0.170, \qquad \widehat{\Delta}_4 = 0.100.$$

Table 11 reports the simultaneous confidence intervals obtained from the existing methods, including MOVER1, MOVER2, Wald1, Wald2, Profile likelihood, and Score. These results are used as benchmark methods for comparison with the proposed Dunnett-type saddlepoint simultaneous confidence interval.

**Table 11.** Simultaneous confidence intervals obtained from existing methods.

| | MOVER1 | MOVER2 | Wald1 | Wald2 | Profile likelihood | Score |
|---|---|---|---|---|---|---|
| $\pi_2 - \pi_1$ | (0.013, 0.344) | (−0.044, 0.400) | (−0.110, 0.491) | (−0.114, 0.287) | (−0.121, 0.295) | (−0.111, 0.291) |
| $\pi_3 - \pi_1$ | (0.236, 0.558) | (0.187, 0.608) | (0.108, 0.757) | (−0.002, 0.341) | (−0.001, 0.320) | (0.004, 0.319) |
| $\pi_4 - \pi_1$ | (−0.019, 0.215) | (−0.061, 0.257) | (−0.092, 0.310) | (−0.057, 0.257) | (−0.055, 0.195) | (−0.067, 0.282) |

For comparison, Table 12 reports the result of the proposed Dunnett-type saddlepoint simultaneous confidence interval procedure. The proposed method uses saddlepoint calibration for the joint distribution of the treatment-control contrast vector and therefore accounts for both the correlation induced by the common control group and the finite-sample discreteness of the paired binary data.

**Table 12.** Simultaneous confidence intervals obtained from the proposed method.

| Method | $\pi_2 - \pi_1$ | $\pi_3 - \pi_1$ | $\pi_4 - \pi_1$ |
|---|---|---|---|
| DSPA | (−0.123, 0.297) | (−0.019, 0.359) | (−0.080, 0.280) |

From Table 11, MOVER1 and MOVER2 give positive lower limits for $\pi_3 - \pi_1$, and MOVER1 also gives a positive lower limit for $\pi_2 - \pi_1$. Wald1 also gives a positive lower limit for $\pi_3 - \pi_1$. These methods therefore suggest possible significant differences for some comparisons. However, these conclusions are not fully consistent with the methods that explicitly incorporate the dependence structure under the constant-$R$ model. In particular, Wald2, profile likelihood, and score produce intervals that are much closer to each other and generally include zero for most comparisons.

The difference between Wald1 and Wald2 is particularly informative. Wald1 is based on the null-hypothesis variance approximation, whereas Wald2 uses the dependence model more directly. For $\pi_3 - \pi_1$, Wald1 gives (0.108, 0.757), while Wald2 gives (-0.002, 0.341).

The former suggests a significant difference, whereas the latter is borderline and includes zero. This shows that ignoring or inadequately accounting for the within-subject dependence may lead to overly optimistic conclusions.

The proposed DSPA intervals are close to the dependence-adjusted likelihood-based and score-based intervals, but they are slightly more conservative in this dataset. For example, for $\pi_3 - \pi_1$, DSPA gives (-0.019, 0.359) which includes zero and is slightly wider than the Profile likelihood interval (-0.001, 0.320) and the score interval (0.004, 0.319). This behavior is consistent with the purpose of saddlepoint calibration: the method uses higher-order information from the cumulant generating function and gives a more cautious finite-sample adjustment for the joint distribution of multiple treatment-control contrasts.

For $\pi_2 - \pi_1$ and $\pi_4 - \pi_1$, the DSPA intervals also contain zero: (-0.123, 0.297) and (-0.080, 0.280), respectively. Therefore, based on the proposed method, there is no strong evidence that the AR, SL, or ISO groups differ significantly from the DOM control group after simultaneous adjustment and after accounting for the within-subject correlation between paired eyes.

The retinitis pigmentosa dataset was originally reported by Rosner [1] and has subsequently been used in studies of correlated paired binary data and many-to-one simultaneous confidence interval construction [11]. Therefore, this dataset provides a suitable benchmark example for evaluating simultaneous inference methods that account for paired-eye correlation. Compared with separate pairwise analyses or procedures that rely only on first-order approximations, the proposed DSPA method accounts simultaneously for the multiplicity induced by comparing several genetic types with the DOM control group and for the intra-subject correlation between paired eyes. As a result, the DSPA intervals lead to a more cautious interpretation of the group differences. In particular, when the multiplicity-adjusted and correlation-adjusted simultaneous confidence intervals contain zero, there is no statistically significant evidence that the corresponding genetic type differs from the DOM group in the proportion of affected eyes. This more conservative conclusion is appropriate because ignoring either the paired-eye correlation or the many-to-one comparison structure may overstate the evidence for group differences.

Overall, the real data analysis shows that the conclusion depends on whether the within-subject dependence and simultaneous comparison structure are adequately handled. The MOVER and Wald1 methods may suggest significance for some comparisons, whereas the dependence-adjusted methods and the proposed DSPA method lead to a more conservative conclusion. The proposed DSPA procedure provides a higher-order simultaneous inference approach and supports the conclusion that the differences in affection rates among the genetic types are not statistically significant after adjusting for multiplicity and paired-eye correlation.

## 7. Conclusions

This paper proposed a Dunnett-type saddlepoint simultaneous confidence interval procedure for many-to-one comparisons of proportion differences in correlated paired binary data. The method was motivated by the need for more accurate finite-sample simultaneous inference when multiple treatment or disease groups are compared with a common control group and the binary responses from paired organs are correlated.

Under the constant-$R$ model, the number of positive responses from each subject follows a three-category distribution. This structure allows the cumulant generating function of the treatment-control contrast vector to be derived explicitly. Based on this cumulant generating function, the proposed DSPA method replaces the first-order multivariate normal calibration used in the classical ND procedure with a saddlepoint calibration. Therefore, the method accounts not only for the dependence induced by the common control group, but also for higher-order finite-sample features of the joint distribution.

The simulation studies showed that the proposed method has clear advantages in sparse, unbalanced, and strongly dependent settings. In these settings, the ND method may undercover, while the DSPA method improves simultaneous coverage accuracy. Compared with MOVER-type methods, DSPA avoids excessive conservativeness and produces shorter intervals when the MOVER intervals become

overly wide. In moderate settings, DSPA remains comparable to existing Wald-type, profile likelihood, score, and ND methods in terms of both coverage probability and mean interval width.

The real data analysis further illustrated the practical usefulness of the proposed method. Applied to the retinitis pigmentosa dataset, DSPA produced simultaneous confidence intervals that were broadly consistent with the dependence-adjusted likelihood-based and score-based methods. The analysis indicated no strong evidence of significant differences in affection rates between the comparison groups and the DOM control group after accounting for multiplicity and within-subject correlation.

In summary, the proposed Dunnett-type saddlepoint simultaneous confidence interval procedure provides a useful higher-order alternative for many-to-one comparisons of proportion differences in correlated paired binary data. It is particularly suitable for finite-sample biomedical studies involving paired organs, where the data may be sparse, the group sizes may be unbalanced, and the within-subject dependence may be substantial. Future work may extend the proposed framework to other effect measures, such as proportion ratios, odds ratios, and relative risks, and to more general correlated binary data models, including stratified bilateral and unilateral data settings.

## Author contributions

Juanjuan Zhang: conceptualization, methodology, formal analysis, software, validation, writing – original draft, visualization; Yujuan Luo: methodology, formal analysis, software, data curation, validation, writing – review and editing; Weixian Wang: conceptualization, methodology, supervision, project administration, writing – review and editing. All authors have read and approved the final manuscript.

## Use of Generative-AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

## Acknowledgments

The authors gratefully acknowledge the support of the Talent Program of Guangxi Normal University, grant number DC2600000938. Weixian Wang was supported by this program.

## Conflict of interest

The authors declare that there are no conflicts of interest.

## References

1. B. Rosner, Statistical methods in ophthalmology: an adjustment for the intraclass correlation between eyes, *Biometrics*, **38** (1982), 105–114.
2. G. E. Dallal, Paired Bernoulli trials, *Biometrics*, **44** (1988), 253–257. https://doi.org/10.2307/2531913

3. A. Donner, Statistical methods in ophthalmology: an adjusted chi-square approach, *Biometrics*, **45** (1989), 605–611. https://doi.org/10.2307/2531501
4. M. L. Tang, N. S. Tang, B. Rosner, Statistical inference for correlated data in ophthalmologic studies, *Stat. Med.*, **25** (2006), 2771–2783. https://doi.org/10.1002/sim.2425
5. N. S. Tang, M. L. Tang, S. F. Qiu, Testing the equality of proportions for correlated otolaryngologic data, *Comput. Stat. Data Anal.*, **52** (2008), 3719–3729. https://doi.org/10.1016/j.csda.2007.12.017
6. C. X. Ma, G. Shan, S. Liu, Homogeneity test for correlated binary data, *PLOS One*, **10** (2015), e0124337. https://doi.org/10.1371/journal.pone.0124337
7. N. S. Tang, S. F. Qiu, M. L. Tang, Y. B. Pei, Asymptotic confidence interval construction for proportion difference in medical studies with bilateral data, *Stat. Methods Med. Res.*, **20** (2011), 233–259. https://doi.org/10.1177/0962280209358135
8. X. Peng, C. Liu, S. Liu, C. X. Ma, Asymptotic confidence interval construction for proportion ratio based on correlated paired data, *J. Biopharm. Stat.*, **29** (2019), 1137–1152. https://doi.org/10.1080/10543406.2019.1584629
9. C. W. Dunnett, A multiple comparison procedure for comparing several treatments with a control, *J. Amer. Stat. Assoc.*, **50** (1955), 1096–1121. https://doi.org/10.1080/01621459.1955.10501294
10. W. W. Piegorsch, Multiple comparisons for analyzing dichotomous response, *Biometrics*, **47** (1991), 45–52. https://doi.org/10.2307/2532494
11. Z. Yang, G. L. Tian, X. Liu, C. X. Ma, Simultaneous confidence interval construction for many-to-one comparisons of proportion differences based on correlated paired data, *J. Appl. Stat.*, **48** (2021), 1442–1456. https://doi.org/10.1080/02664763.2020.1795815
12. X. Zhang, C. Ma, Testing the homogeneity of differences between two proportions for stratified bilateral and unilateral data across strata, *Mathematics*, **11** (2023), 4156. https://doi.org/10.3390/math11194156
13. S. Liang, K. T. Fang, X. W. Huang, Y. Xin, C. X. Ma, Homogeneity tests and interval estimations of risk differences for stratified bilateral and unilateral correlated data, *Stat. Papers*, **65** (2024), 3499–3543. https://doi.org/10.1007/s00362-024-01532-6
14. Z. Zhang, C. X. Ma, Simultaneous confidence interval construction for many-to-one of proportion ratios of bilateral correlated data, *PLOS One*, **19** (2024), e0311850. https://doi.org/10.1371/journal.pone.0311850
15. S. Hua, C. Ma, Common odds ratio test and interval estimation for stratified bilateral and unilateral data, *Stat. Methods Med. Res.*, **33** (2024), 1559–1576. https://doi.org/10.1177/09622802241267357
16. K. Wang, C. X. Ma, Interval estimation of relative risks for combined unilateral and bilateral correlated data, *J. Biopharm. Stat.*, **35** (2025), 163–186. https://doi.org/10.1080/10543406.2023.2297789
17. H. E. Daniels, Saddlepoint approximations in statistics, *Ann. Math. Stat.*, **25** (1954), 631–650. https://doi.org/10.1214/aoms/1177728652
18. R. Lugannani, S. O. Rice, Saddle point approximation for the distribution of the sum of independent random variables, *Adv. Appl. Probab.*, **12** (1980), 475–490. https://doi.org/10.2307/1426607

19. N. Reid, Saddlepoint methods and statistical inference, *Stat. Sci.*, **3** (1988), 213–238. https://doi.org/10.1214/ss/1177012906

20. J. L. Jensen, *Saddlepoint approximations*, Oxford University Press, 1995.

21. J. E. Kolassa, *Series approximation methods in statistics*, Springer, 2006. https://doi.org/10.1007/0-387-32227-2

22. W. Zhou, J. B. Nielsen, L. G. Fritsche, R. Dey, M. E. Gabrielsen, B. N. Wolford, et al., Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies, *Nat. Genet.*, **50** (2018), 1335–1341. https://doi.org/10.1038/s41588-018-0184-y

23. W. Bi, L. Fritsche, B. Mukherjee, M. Kim, S. Lee, A fast and accurate method for genome-wide time-to-event data analysis and its application to UK Biobank, *Amer. J. Hum. Genet.*, **107** (2020), 222–233. https://doi.org/10.1016/j.ajhg.2020.06.003

24. P. V. B. Johnsen, Ø. Bakke, T. Bjørnland, A. T. DeWan, M. Langaas, Saddlepoint approximations to score test statistics in logistic regression for analyzing genome-wide association studies, *Stat. Med.*, **42** (2023), 2746–2759. https://doi.org/10.1002/sim.9746

25. H. A. Newer, Saddlepoint approximation $p$-values of weighted log-rank tests based on censored clustered data under block Efron's biased-coin design, *Stat. Methods Med. Res.*, **32** (2023), 465–473. https://doi.org/10.1177/09622802221143498

26. Y. Ma, H. Xu, Y. Li, H. Kim, L. L. Xu, L. Miao, et al., SPAmix: a scalable, accurate, and universal analysis framework for large-scale genetic association studies in admixed populations, *Genome Biol.*, **26** (2025), 356. https://doi.org/10.1186/s13059-025-03827-9


AIMS Press