

Article

Statistical Analysis of a Generalized Linear Model for Bilateral Correlated Data Under Donner's Model

Jinlong Cheng, Zhiming Li *  and Keyi Mou

College of Mathematics and System Science, Xinjiang University, Urumqi 830046, China;
107552303595@stu.xju.edu.cn (J.C.); 107556521095@stu.xju.edu.cn (K.M.)

* Correspondence: zml@xju.edu.cn

Abstract

Paired data often arise in medical studies, with a correlation between responses of paired organs or parts. Under an intra-correlated model, this paper proposes a generalized linear model to investigate probable confounding factors of the individual response rates in paired data. The main link functions include logistic, log-log, complementary log-log, probit, and double exponential. The estimators of model parameters are calculated through the Newton-Raphson, quadratic lower bound, and Fisher bounded algorithms. Then, three tests (i.e., likelihood ratio test, Wald-type test, and score test) are constructed to analyze whether covariates significantly affect the response rate. Finally, the proposed methods are illustrated by numerical simulation and visual impairment data from Iran.

Keywords: paired data; generalized linear model; parameter estimation; hypothesis test

MSC: 62F03; 62F05; 62F10



Academic Editor: Matieyendou Lamboni

Received: 1 May 2025

Revised: 18 June 2025

Accepted: 24 June 2025

Published: 26 June 2025

Citation: Cheng, J.; Li, Z.; Mou, K. Statistical Analysis of a Generalized Linear Model for Bilateral Correlated Data Under Donner's Model. *Axioms* **2025**, *14*, 500. <https://doi.org/10.3390/axioms14070500>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In medical studies, paired data (i.e., eyes, ears, and arms) are often encountered from paired organs or parts of patients. For each individual, the outcomes for paired organs may be neither, only one, or both responses. A specific correlation is usually considered to avoid biased results. Until now, various intra-correlated models have been proposed to investigate dependent properties, such as Ronser's model [1], Dallal's model [2], and Donner's model [3]. The first two models reflect response dependence through conditional probabilities, which may be inapplicable in some extreme cases, while Donner's model is based on a correlation coefficient ρ . Thompson [4] considered that Donner's model is more effective than others. Significant progress has been made in the study of response rates under these intra-correlated models [5–10]. However, existing research primarily focuses on whether confounding factors significantly impact response rate risk measures. For example, Zhang and Ma [11] analyzed the homogeneity of differences between two proportions for stratified data across strata. Hua and Ma [12] assessed the significance of treatment effects on the ratio of response rates. Sun et al. [13] proposed three testing methods to evaluate the homogeneity of the risk difference in stratified data. These results cannot provide the effect of various factors on response rates.

The generalized linear model is an extension of the traditional linear regression model, designed to accommodate a broader range of data types and distributions [14–16]. Logistic regression is widely used to investigate binary data [17,18]. The classical logistic regression model assumes that the observations are independent. Taking into account the correlated

data [19–22], Lin et al. [23] constructed a logistic regression model to explore the relationship between the incidence of ophthalmic disease and the covariates under Donner’s model. However, the dependency curve between the response rate and confounding effects is often unknown. Although logistic regression can solve such correlated data, it may result in inadequate model fitting in the context of complex relationships, failing to fully capture the nonlinear associations between these variables. Based on the preceding analysis, the innovations of the article are summarized below:

- (i) A generalized linear model with correlated data is proposed to investigate the relationship between the response rate and confounding effects, so the logistic model is a special case [23].
- (ii) The Newton–Raphson (NR) iterative algorithm calculates maximum likelihood estimators (MLEs) of unknown parameters. Considering the complexity of the link function, the quadratic lower bound (QLB) and the Fisher bounded algorithm are further introduced.
- (iii) The likelihood ratio test, Wald-type test, and score test are constructed to analyze whether the confounding effects significantly impacted the response rate.
- (iv) Numerical simulations compare the performance of algorithms and statistics. Real data is studied to illustrate the proposed methods.

The rest of the paper is organized as follows. In Section 2, we first introduce the generalized linear model and the log-likelihood under the Donner framework. Then, three iterative algorithms for parameter estimation and hypothesis testing are conducted. The main results are provided in Section 3, and an application is used to illustrate the theoretical results in Section 4. Section 5 gives the discussion and a brief conclusion.

2. Methods

2.1. Generalized Linear Model with Correlated Data

When each individual in the study contributes the measurements of paired organs, the overall outcome may be neither, only one, or both responses; this is called bilateral data. If a certain correlation exists between paired organs’ responses, it is called bilateral correlated data. Let $Z_{ik} = 1$ if the k -th organ of the i -th individual has a response, and $Z_{ik} = 0$ otherwise for $k = 1, 2, i = 1, \dots, n$. Suppose that Z_{ik} are independent and Bernoulli-distributed with probabilities π_i , that is, $Z_{ik} \sim \text{Bernoulli}(1, \pi_i)$. Thus, $E[Z_{ik}] = P(Z_{ik} = 1) = \pi_i (i = 1, \dots, n)$. Under Donner’s model, we have $\text{Corr}(Z_{ik}, Z_{i(3-k)}) = \rho$ for $k = 1, 2$, and $i = 1, \dots, n$, where $\rho (|\rho| \leq 1)$ reflects the correlation of the paired data. Denote by p_{li} the corresponding probability of $l = (0, 1, 2)$ response(s) of the i -th individual. By calculation, the probabilities of neither, one, or both responses can be obtained by

$$p_{0i} = \rho(1 - \pi_i) + (1 - \rho)(1 - \pi_i)^2, p_{1i} = 2\pi_i(1 - \rho)(1 - \pi_i), p_{2i} = \rho\pi_i + (1 - \rho)\pi_i^2,$$

for $i = 1, 2, \dots, n$, satisfying $0 \leq p_{li} \leq 1$, and $\sum_{l=0}^2 p_{li} = 1$. The detailed calculations are provided in the Appendix A.1. Let Y_i be the response variable of paired organs of the i th individual, and take the values 0, 1, 2, where $Y_i = 0$ indicates that neither of the two organs has the response, $Y_i = 1$ indicates that only one organ has the response, and $Y_i = 2$ indicates that both organs have the responses. Denote $\boldsymbol{\pi} = (\pi_1, \dots, \pi_n)$. Based on the probabilities $p_{li} (l = 0, 1, 2)$, we obtain the likelihood function

$$L(\boldsymbol{\pi}, \rho) = \prod_{i=1}^n p_{0i}^{I(Y_i=0)} p_{1i}^{I(Y_i=1)} p_{2i}^{I(Y_i=2)},$$

where $I(\cdot)$ denotes the indicator function. Take the natural logarithm of both sides of $L(\boldsymbol{\pi}, \rho)$, thereby obtaining the log-likelihood function:

$$\ell(\boldsymbol{\pi}, \rho) = \sum_{i=1}^n \left\{ I(Y_i = 0) \log[(1 - \pi_i)(\rho\pi_i - \pi_i + 1)] + I(Y_i = 1) \log[2\pi_i(1 - \rho)(1 - \pi_i)] \right. \\ \left. + I(Y_i = 2) \log[\pi_i^2 + \rho\pi_i(1 - \pi_i)] \right\}. \quad (1)$$

To investigate the relationship between response rates of n different individuals and confounding factors (e.g., age), we introduce a generalized linear model (GLM) as follows:

$$g(\pi_i) = \mathbf{x}_i^T \boldsymbol{\beta}, \quad i = 1, \dots, n, \quad (2)$$

where $\mathbf{x}_i = (1, x_{i(1)}, \dots, x_{i(q-1)})^T$ is the covariate vector of interest, $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_{q-1})^T$ is the vector of unknown regression coefficients, and $g(\cdot)$ is a monotonic and differentiable link function. In this work, we mainly use the following link functions:

- (i) *Logistic link function.* The logistic link function is typically used in the model (2) to analyze the relationship between a binary response variable and linear predictor variables. Take the link function $g(\pi_i) = \ln(\pi_i / (1 - \pi_i))$.
- (ii) *Log-log and complementary log-log.* When the response probability exhibits asymmetry or there is a high likelihood of extreme events, the log-log or complementary log-log link function can better capture the characteristics of the data. The log-log link function $g(\pi_i) = \log(-\log(1 - \pi_i))$, and the complementary log-log link function $g(\pi_i) = \log(-\log(\pi_i))$.
- (iii) *Power family.* Considering that the response rate falls within the $(0, 1)$ range, a power family distribution with a shape parameter is appropriate as a link function [24]. Its distribution is indexed by shape parameters as follows:

$$F_p(x) = \frac{p\omega_p^{1/2p}}{\Gamma(1/2p)} \int_{-\infty}^x e^{-\omega_p|t|^{2p}} dt,$$

where p is the shape parameter, $\omega_p = 1/2p$, and $\Gamma(\cdot)$ is the gamma function. Specifically, the power family derives two link functions: the double exponential (DE) link function $\pi_i = \frac{1}{2} \int_{-\infty}^{\mathbf{x}_i^T \boldsymbol{\beta}} e^{-|t|} dt$ and the probit (Pt) link function $g(\pi_i) = \Phi^{-1}(\pi_i)$, which correspond to $p = 1/2$ and $p = 1$, respectively. As the distribution is symmetric, it can be used as a symmetric link function to describe symmetric response characteristics. Table 1 provides a summary of the five link functions.

Table 1. Comparison of five link functions.

Link Function	π_i	Applicable Data	Advantages	Algorithm
Logistic (L)	$\pi_i = \frac{e^{\mathbf{x}_i^T \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}}$	Symmetric response probabilities	Interpretable, widely used	QLB-NR
Log-log (LL)	$\pi_i = 1 - e^{-e^{\mathbf{x}_i^T \boldsymbol{\beta}}}$	Right-extreme (near 0 to increasing)	Asymmetric right-skewed data	FB-NR
Complementary log-log (CLL)	$\pi_i = e^{-e^{\mathbf{x}_i^T \boldsymbol{\beta}}}$	Left-extreme (near 1 to decreasing)	Asymmetric left-skewed data	FB-NR
Double exponential (DE)	$\pi_i = \frac{1}{2} \int_{-\infty}^{\mathbf{x}_i^T \boldsymbol{\beta}} e^{- t } dt$	Symmetric, extreme events	Strong symmetry	FB-NR
Probit (Pt)	$\pi_i = \Phi(\mathbf{x}_i^T \boldsymbol{\beta})$	Symmetric, near-normal	High accuracy, low errors	FB-NR

Note. QLB-NR: Quadratic lower-bound-based NR algorithm; FB-NR: Fisher-bound-based NR algorithm.

Denote $h(\mathbf{x}_i^T \boldsymbol{\beta}) = \pi_i = g^{-1}(\mathbf{x}_i^T \boldsymbol{\beta})$; the function (1) can be rewritten as

$$\begin{aligned} \ell(\boldsymbol{\beta}, \rho) = & \sum_{i=1}^n \{ I(Y_i = 0) \log[(1 - h(\mathbf{x}_i^T \boldsymbol{\beta}))(\rho h(\mathbf{x}_i^T \boldsymbol{\beta}) - h(\mathbf{x}_i^T \boldsymbol{\beta}) + 1)] \\ & + I(Y_i = 1) \log[2h(\mathbf{x}_i^T \boldsymbol{\beta})(1 - \rho)(1 - h(\mathbf{x}_i^T \boldsymbol{\beta}))] \\ & + I(Y_i = 2) \log[h^2(\mathbf{x}_i^T \boldsymbol{\beta}) + \rho h(\mathbf{x}_i^T \boldsymbol{\beta})(1 - h(\mathbf{x}_i^T \boldsymbol{\beta}))] \}. \end{aligned} \quad (3)$$

Given the link function $g(\pi_i)$, we can estimate the unknown coefficient vector $\boldsymbol{\beta}$ and the correlation coefficient ρ . Some studies have demonstrated that maximum likelihood estimation (MLE) remains valid even when there is correlation between parameters [25]. By fixing one parameter and computing the partial derivative of the log-likelihood function (3) for the other parameter [26], we obtain easily computable derivatives to $\boldsymbol{\beta}$ and ρ . Setting these derivatives to zero, that is, $\frac{\partial \ell}{\partial \boldsymbol{\beta}} = 0$, $\frac{\partial \ell}{\partial \rho} = 0$. Thus, the MLEs are the solution of the above equations. However, since closed-form solutions do not exist, specific algorithms must be proposed to address this issue. In the following section, we focus on these algorithms to estimate the values of unknown parameters for different link functions.

2.2. Parameter Estimation

Based on the link functions, the NR algorithm is first used to estimate the correlation coefficient ρ . For the regression coefficient vector $\boldsymbol{\beta}$, the NR algorithm may fail to converge in certain situations, particularly for complex link functions. In such cases, the QLB algorithm is proposed for the logistic link function, while the Fisher-bounded algorithm is proposed for the remaining link functions.

Newton–Raphson algorithm. Given the t -th approximations $\pi^{(t)}$ and $\rho^{(t)}$, the $(t + 1)$ th approximation of the MLE $\hat{\rho}$ can be calculated via the NR algorithm:

$$\rho^{(t+1)} = \rho^{(t)} - \left[\frac{\partial^2 \ell(\boldsymbol{\pi}^{(t)}, \rho^{(t)})}{\partial \rho^2} \right]^{-1} \frac{\partial \ell(\boldsymbol{\pi}^{(t)}, \rho^{(t)})}{\partial \rho},$$

where

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\pi}, \rho)}{\partial \rho} &= \sum_{i=1}^n \left[\frac{\pi_i I(Y_i = 0)}{\pi_i(\rho - 1) + 1} + \frac{I(Y_i = 1)}{\rho - 1} + \frac{(1 - \pi_i)I(Y_i = 2)}{\pi_i + \rho(1 - \pi_i)} \right], \\ \frac{\partial^2 \ell(\boldsymbol{\pi}, \rho)}{\partial \rho^2} &= - \sum_{i=1}^n \left[\frac{\pi_i^2 I(Y_i = 0)}{(\rho \pi_i - \pi_i + 1)^2} + \frac{I(Y_i = 1)}{(\rho - 1)^2} + \frac{(\pi_i - 1)^2 I(Y_i = 2)}{(\rho + \pi_i - \rho \pi_i)^2} \right]. \end{aligned}$$

Although the MLE of ρ can be obtained, the parameter vector $\boldsymbol{\beta}$ cannot be directly calculated by the NR algorithm. We use the assembly and decomposition (AD) approach proposed by [27] to construct a surrogate function $Q(\boldsymbol{\pi} | \boldsymbol{\pi}^{(t)}, \rho)$ as follows:

$$Q(\boldsymbol{\pi} | \boldsymbol{\pi}^{(t)}, \rho) = \sum_{i=1}^n \left\{ a_i(\pi_i^{(t)}, \rho) \log(\pi_i) + b_i(\pi_i^{(t)}, \rho) \log(1 - \pi_i) + c_i^{(t)}(\rho) \right\},$$

where $\pi_i^{(t)}$ is the t th approximation of π_i , $c_i^{(t)}(\rho)$ is independent of π_i ($i = 1, \dots, n$), and

$$\begin{aligned} a_i(\pi_i^{(t)}, \rho) &= I(Y_i = 1) + I(Y_i = 2) + \frac{I(Y_i = 2)\pi_i^{(t)}}{\pi_i^{(t)} + \rho(1 - \pi_i^{(t)})} + \frac{I(Y_i = 0)\rho\pi_i^{(t)}}{\rho\pi_i^{(t)} + (1 - \pi_i^{(t)})}, \\ b_i(\pi_i^{(t)}, \rho) &= I(Y_i = 0) + I(Y_i = 1) + \frac{I(Y_i = 2)\rho(1 - \pi_i^{(t)})}{\pi_i^{(t)} + \rho(1 - \pi_i^{(t)})} + \frac{I(Y_i = 0)(1 - \pi_i^{(t)})}{\rho\pi_i^{(t)} + (1 - \pi_i^{(t)})}. \end{aligned}$$

The detailed construction process is provided in Appendix A.2. The function $Q(\pi|\pi^{(t)}, \rho)$ satisfies the condition $\ell(\pi, \rho) - Q(\pi|\pi^{(t)}, \rho) \geq 0$. Replacing π_i in the surrogate function with $\pi_i = h(\mathbf{x}_i^T \boldsymbol{\beta})$, we have

$$Q(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}, \rho) = \sum_{i=1}^n \left\{ a_i(\pi_i^{(t)}, \rho) \log[h(\mathbf{x}_i^T \boldsymbol{\beta})] + b_i(\pi_i^{(t)}, \rho) \log[1 - h(\mathbf{x}_i^T \boldsymbol{\beta})] + c_i^{(t)}(\rho) \right\}.$$

Given $(\boldsymbol{\beta}^{(t)}, \rho^{(t)})$, the $(t+1)$ th approximation of the MLEs $\boldsymbol{\beta}$ can be obtained by

$$\boldsymbol{\beta}^{(t+1)} = \arg \max_{\boldsymbol{\beta} \in \mathbb{R}^q} Q(\boldsymbol{\beta} | \boldsymbol{\beta}^{(t)}, \rho^{(t)}). \quad (4)$$

Thus, we can use the NR algorithm to calculate Equation (4) by

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} - \left[\frac{\partial^2 Q(\boldsymbol{\beta} | \boldsymbol{\beta}^{(t)}, \rho)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \right]_{\boldsymbol{\beta}=\boldsymbol{\beta}^{(t)}}^{-1} \left[\frac{\partial Q(\boldsymbol{\beta} | \boldsymbol{\beta}^{(t)}, \rho)}{\partial \boldsymbol{\beta}} \right]_{\boldsymbol{\beta}=\boldsymbol{\beta}^{(t)}}.$$

In practice, the NR algorithm may be unsuitable for all link functions since it may fail to guarantee the monotonic increase in the objective function $Q(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}, \rho)$ in each iteration. Based on this, we can adopt the QLB algorithm to estimate the MLEs.

Quadratic lower-bound-based NR algorithm. The QLB algorithm improves the NR algorithm using a parameter-independent quadratic lower-bound matrix, replacing the negative observed information matrix [28]. The key idea is to find a negative definite matrix B , independent of the unknown parameters, such that $\frac{\partial^2 Q(\boldsymbol{\beta}^{(t)})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \geq B$ for solving the extreme value problem (4) as follows:

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} - B^{-1} \left(\frac{\partial Q(\boldsymbol{\beta}^{(t)} | \boldsymbol{\beta}^{(t)}, \rho)}{\partial \boldsymbol{\beta}} \right).$$

Thus, the algorithm avoids computing the inverse of the observed information matrix at each iteration to reach a maximum. Compared with other algorithms, such as the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm, the QLB may be computationally more straightforward and efficient. The detailed process of the QLB-based NR algorithm is given in Algorithm 1.

Algorithm 1: QLB-based NR algorithm

Input: Initial value $\boldsymbol{\beta}^{(0)}$; Initial value $\rho^{(0)}$; Sample size n ; Covariate x_i , for $i, \dots, n$; thresholds $\varepsilon_1, \varepsilon_2$

Output: MLEs $\boldsymbol{\beta}^{(k)}$ and $\rho^{(k)}$

1 $k \leftarrow 0$;

2 Compute a negative definite B ;

3 **while** $\|\nabla Q(\boldsymbol{\beta}^{(k)}, \rho^{(k)})\| > \varepsilon_1$ **or** $\|\boldsymbol{\beta}^{(k)} - \boldsymbol{\beta}^{(k-1)}\| > \varepsilon_2$ **do**

4 $\boldsymbol{\beta}^{(k+1)} = \boldsymbol{\beta}^{(k)} - B^{-1} \left(\frac{\partial Q(\boldsymbol{\beta}^{(k)} | \boldsymbol{\beta}^{(k)}, \rho^{(k)})}{\partial \boldsymbol{\beta}} \right)$;

5 **for** $i \leftarrow 1$ **to** n **do**

6 $\pi_i^{(k)} = h(\mathbf{x}_i^T \boldsymbol{\beta}^{(k)})$;

7 Update Likelihood function $\ell(\boldsymbol{\pi}^{(k)}, \rho^{(k)})$, where $\boldsymbol{\pi}^{(t)} = (\pi_1^{(t)}, \dots, \pi_n^{(t)})^T$;

8 $\rho^{(k+1)} = \rho^{(k)} - \left[\frac{\partial^2 \ell(\boldsymbol{\pi}^{(k)}, \rho^{(k)})}{\partial \rho^2} \right]^{-1} \frac{\partial \ell(\boldsymbol{\pi}^{(k)}, \rho^{(k)})}{\partial \rho}$;

9 $k \leftarrow k + 1$;

10 **return** $\boldsymbol{\beta}^{(k)}$ and $\rho^{(k)}$

For the logistic link function, the MLEs of β can be calculated:

$$Q(\beta|\beta^{(t)}, \rho)_{\text{logistic}} = \sum_{i=1}^n \left\{ a_i(\pi_i^{(t)}, \rho) \mathbf{x}_i^T \beta - [a_i(\pi_i^{(t)}, \rho) + b_i(\pi_i^{(t)}, \rho)] \log(1 + e^{\mathbf{x}_i^T \beta}) + c_i^{(t)}(\rho) \right\}.$$

Then

$$\left. \frac{\partial Q(\beta|\beta^{(t)}, \rho^{(t)})_{\text{logistic}}}{\partial \beta} \right|_{\beta=\beta^{(t)}} = \sum_{i=1}^n \left\{ [a_i(\pi_i^{(t)}, \rho) - 2\pi_i^{(t)}] \mathbf{x}_i \right\} = \mathbf{X}^T [\mathbf{y}^{(t)} - 2\boldsymbol{\pi}^{(t)}],$$

where $\mathbf{y}^{(t)} = (a_1(\pi_1^{(t)}, \rho^{(t)}), \dots, a_n(\pi_n^{(t)}, \rho^{(t)}))^T$, $\mathbf{X}_{n \times q} = (x_1, \dots, x_n)^T$, and $\boldsymbol{\pi}^{(t)} = (\pi_1^{(t)}, \dots, \pi_n^{(t)})^T$. Moreover,

$$\left. \frac{\partial^2 Q(\beta|\beta^{(t)}, \rho^{(t)})_{\text{logistic}}}{\partial \beta \partial \beta'} \right|_{\beta=\beta^{(t)}} = \sum_{i=1}^n - \left\{ 2\pi_i^{(t)}(1 - \pi_i^{(t)}) \mathbf{x}_i \mathbf{x}_i^T \right\} = -2\mathbf{X}^T \mathbf{N} \mathbf{X},$$

where $\mathbf{N} = \text{diag}(\pi_1^{(t)}(1 - \pi_1^{(t)}), \pi_2^{(t)}(1 - \pi_2^{(t)}), \dots, \pi_n^{(t)}(1 - \pi_n^{(t)}))$. For the inequality $\pi_i(1 - \pi_i) \leq 1/4$ ($i = 1, \dots, n$), the bound value of the logistic link function is 0.25. So, $B = -(1/2)\mathbf{X}^T \mathbf{X}$. Further, the MLEs of β are calculated as follows:

$$\beta^{(t+1)} = \beta^{(t)} + 2(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T [\mathbf{y}^{(t)} - 2\boldsymbol{\pi}^{(t)}].$$

Note that the above iterative formula applies to low- and high-dimensional scenarios. Regularization techniques can be used to avoid overfitting and promote stability in high-dimensional cases. For other link functions, $a_i(\pi_i^{(t)}, \rho)$ and $b(\pi_i^{(t)}, \rho)$ are not constant, making $\frac{\partial^2 Q(\beta|\beta^{(t)}, \rho^{(t)})}{\partial \beta \partial \beta'}$ variable and the lower bound matrix B hard to determine. Therefore, the Fisher-bound algorithm is proposed to estimate regression coefficients.

Fisher-bound-based NR algorithm. The Fisher scoring method is usually sensitive to initial values, and consequently, there may be difficulties in obtaining convergence. Under certain general conditions on the information matrix, the bounded algorithm [27] has been shown to consistently converge to the MLE, independent of the initial values. The specific iterative process of this algorithm is as follows:

$$\beta^{(t+1)} = \beta^{(t)} + B^{-1} S_c(\beta^{(t)}).$$

For a random variable $\eta \in (-\infty, +\infty)$, with the inverse function of the link function $h(\eta)$, where $\frac{\partial h(\eta)}{\partial \eta} = f(\eta)$, the vector $S_c(\beta^{(t)})$ satisfies

$$S_c(\beta^{(t)}) = \begin{pmatrix} \sum_{i=1}^n \frac{f(\mathbf{x}_i^T \beta)}{h(\mathbf{x}_i^T \beta)[1-h(\mathbf{x}_i^T \beta)]} [a_i(\pi_i^{(t)}, \rho) - 2h(\mathbf{x}_i^T \beta)] \\ \vdots \\ \sum_{i=1}^n \frac{f(\mathbf{x}_i^T \beta)}{h(\mathbf{x}_i^T \beta)[1-h(\mathbf{x}_i^T \beta)]} [a_i(\pi_i^{(t)}, \rho) - 2h(\mathbf{x}_i^T \beta)] x_{i(q-1)} \end{pmatrix},$$

and B is an upper bound positive definite matrix satisfying $B \geq I(\beta^{(t)})$, defined as follows:

$$B = nb \begin{bmatrix} 1 & \cdots & \sum_{i=1}^n x_{i(s)} & \cdots & \sum_{i=1}^n x_{i(q-1)} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n x_{i(s)} & \cdots & \sum_{i=1}^n x_{i(s)}^2 & \cdots & \sum_{i=1}^n x_{i(s)} x_{i(q-1)} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n x_{i(q-1)} & \cdots & \sum_{i=1}^n x_{i(s)} x_{i(q-1)} & \cdots & \sum_{i=1}^n x_{i(q-1)}^2 \end{bmatrix}.$$

The pseudocode of the Fisher-bound-based-NR algorithm is shown in Algorithm 2.

The bound b can be obtained by bounding the function $C(\eta) = \frac{2f_p^2(\eta)}{h(\eta)[1-h(\eta)]}$. For both the log-log and complementary log-log link functions, the bound b is 1.2952. For the power family, Devidas and Georege [29] proved that the least upper bound of the function $C(\eta)$ is given by $C(0) = 4c_p^2$, where $c_p = p^{\omega_p/2p} [\Gamma(1/2p)]^{-1}$ and $\omega_p = 1/2p$, for $0 < p \leq 1$. For $p > 1$, the bound is obtained by solving the equation

$$\frac{4f_p'(\eta)}{1 - 2F_p(\eta)} = \frac{2f_p^2(\eta)}{F_p(\eta)[1 - F_p(\eta)]},$$

where $\partial F_p(\eta)/\partial \eta = f_p(\eta)$. With shape parameters $p = \frac{1}{2}$ and $p = 1$, the corresponding link functions are double exponential and probit, with b values of 1.2732 and 2, respectively.

Algorithm 2: Fisher-bound-based NR algorithm

Input: Initial value $\beta^{(0)}$; Initial value $\rho^{(0)}$; Sample size n ; Covariate x_i , for $i, \dots, n$; thresholds $\varepsilon_1, \varepsilon_2$
Output: MLEs $\beta^{(k)}$ and $\rho^{(k)}$

```

1  $k \leftarrow 0$ ;
2 Compute a negative definite  $B$ ;
3 while  $\|\nabla Q(\beta^{(k)}, \rho^{(k)})\| > \varepsilon_1$  or  $\|\beta^{(k)} - \beta^{(k-1)}\| > \varepsilon_2$  do
4    $\beta^{(t+1)} = \beta^{(t)} + B^{-1}S_c(\beta^{(t)})$ ;
5   for  $i \leftarrow 1$  to  $n$  do
6      $\pi_i^{(k)} = h(\mathbf{x}_i^T \beta^{(k)})$ ;
7   Update likelihood function  $\ell(\pi^{(k)}, \rho^{(k)})$ , where  $\pi^{(t)} = (\pi_1^{(k)}, \dots, \pi_n^{(k)})^T$ ;
8    $\rho^{(k+1)} = \rho^{(k)} - \left[ \frac{\partial^2 \ell(\pi^{(k)}, \rho^{(k)})}{\partial \rho^2} \right]^{-1} \frac{\partial \ell(\pi^{(k)}, \rho^{(k)})}{\partial \rho}$ ;
9    $k \leftarrow k + 1$ ;
10 return  $\beta^{(k)}$  and  $\rho^{(k)}$ 

```

When the global maximum likelihood estimates $\hat{\theta} = (\hat{\beta}^T, \hat{\rho})^T$ are calculated, the standard errors (σ) of each parameter can be obtained using the square root of the diagonal elements of the inverse Fisher information matrix $I^{-1}(\hat{\theta})$ in the Appendix A. Hence, the asymptotic 95% confidence interval can be expressed as $\hat{\theta} \pm 1.96 \times \sigma$.

In large sample settings, jointly estimating β and ρ with the NR algorithm incurs an overall time complexity of $O(2k(nq^2 + q^3))$, where n is the sample size, q is the covariate dimension, and k is the number of iterations. The QLB-based NR algorithm only inverts the matrix B once during the initial estimation of β , which reduces the total time complexity to $O((k+1)(nq^2 + q^3))$. The time complexity of the Fisher-bound-based NR is approximately the same as that of QLB-based NR. The three algorithms have consistent performance in terms of space complexity. We list the computational complexity of the three algorithms in Table 2.

Table 2. Comparison of algorithmic computational complexity.

Algorithms	Space Complexity	Time Complexity
NR	$O(nq + q^2)$	$O(2k(nq^2 + q^3))$
QLB-NR	$O(nq + q^2)$	$O((k+1)(nq^2 + q^3))$
FB-NR	$O(nq + q^2)$	$O((k+1)(nq^2 + q^3))$

Note. QLB-NR: Quadratic lower-bound-based NR algorithm; FB-NR: Fisher-bound-based NR algorithm.

2.3. Hypothesis Test

In this section, we propose three statistics to test the following hypothesis:

$$H_0 : C\beta = 0_m, \quad \text{vs.} \quad H_1 : C\beta \neq 0_m, \quad (5)$$

where C is an $m \times q$ matrix with $\text{rank}(C) = r$ and $r < q$.

Likelihood ratio test. The likelihood ratio statistic can be constructed through the global and constrained MLEs as follows:

$$T_L = 2[\ell(\tilde{\beta}, \tilde{\rho}) - \ell(\hat{\beta}_{H_0}, \hat{\rho}_{H_0})],$$

where $(\tilde{\beta}, \tilde{\rho})$ is the global MLEs of β, ρ through the algorithm described above. And $(\hat{\beta}_{H_0}, \hat{\rho}_{H_0})$ are the constrained MLEs of (β, ρ) under H_0 , which can be obtained by

$$\beta_{H_0}^{(t+1)} = \arg \max_{\beta} Q(\beta | \beta_{H_0}^{(t)}, \rho_{H_0}^{(t)}), \quad \rho_{H_0}^{(t+1)} = \rho_{H_0}^{(t)} - \left[\frac{\partial^2 \ell(\pi_{H_0}^{(t)}, \rho_{H_0}^{(t)})}{\partial \rho^2} \right]^{-1} \frac{\partial \ell(\pi_{H_0}^{(t)}, \rho_{H_0}^{(t)})}{\partial \rho}. \quad (6)$$

We employ the built-in `fmincon` function in MATLAB (Version R2023b) to solve Equation (6). The likelihood ratio test is asymptotically distributed as a chi-square distribution with r degrees of freedom.

Wald test. Denote $A = [C, 0]$ and $\theta = (\beta^T, \rho)^T$. Then the null hypothesis H_0 can be changed to $A\theta = 0$. The Wald test T_W is given by

$$T_W = (A\hat{\theta})^T [AI^{-1}(\hat{\theta})A^T]^{-1} (A\hat{\theta}),$$

where $\hat{\theta} = (\hat{\beta}^T, \hat{\rho})^T$ denote the global MLEs of θ and $I(\theta)$ denotes the Fisher information matrix in the Appendix A. The Wald test is asymptotically distributed as a chi-square distribution with r degrees of freedom.

Score test. Define a score function $s(\theta) = (\partial \ell(\beta, \rho) / \partial \beta_0, \dots, \partial \ell(\beta, \rho) / \partial \beta_{q-1}, 0)^T$. The score statistic is given by

$$T_S = [s(\hat{\theta}_{H_0})]^T I^{-1}(\hat{\theta}_{H_0}) s(\hat{\theta}_{H_0}),$$

where $\hat{\theta}_{H_0} = (\hat{\beta}_{H_0}^T, \hat{\rho}_{H_0})^T$ denote the constrained MLEs of θ . Under H_0 , the score test is asymptotically distributed as a chi-square distribution with r degrees of freedom.

According to the convergence analysis [5], the MLEs obtained using QLB and Fisher bounded algorithms are asymptotically equivalent to the exact MLEs. Therefore, the proposed test statistics still follow asymptotic chi-square distributions.

3. Main Results

First, we analyze the estimators' accuracy for given parameters under various link functions. Subsequently, we compare the algorithms. Then, we investigate the likelihood ratio test, Wald test, and score test under different link functions in terms of the type I error rates (TIEs) and powers.

3.1. Accuracy of the Estimators

In this subsection, we apply bias, average standard errors (ASE), and coverage probability (CP) to assess the point and interval estimates of the parameters (β, ρ) . The bias measures the difference between the average MLEs of the parameters (β, ρ) and the corresponding actual values. The average standard errors mean the average of the standard errors of the MLEs. The coverage probability is a proportion of the asymptotic 95% confidence intervals for the parameters (β, ρ) .

Take the sample size $n = 200, 400$ and correlation coefficient $\rho = 0.3, 0.5, 0.7$. Other parameter settings are given according to the cases below:

Case (1). $q = 2$, $\beta = (-1, 2)^T$, $\mathbf{x}_i = (1, x_{i(1)})^T$, where $\{x_{i(1)}\}_{i=1}^n \stackrel{i.i.d.}{\sim} N(0.4, 2.25 \times 10^{-2})$.

Case (2). $q = 4$, $\beta = (-1, 2, -1, 2)^T$, $\mathbf{x}_i = (1, x_{i(1)}, x_{i(2)}, x_{i(3)})^T$, where $\{x_{i(1)}\}_{i=1}^n \stackrel{i.i.d.}{\sim} N(0.4, 2.25 \times 10^{-2})$, $\{x_{i(2)}\}_{i=1}^n \stackrel{i.i.d.}{\sim} N(0.45, 0.01)$, and $\{x_{i(3)}\}_{i=1}^n \stackrel{i.i.d.}{\sim} 0.3 + 0.06 \times t(5)$.

The samples $\{Y_i\}$ are generated randomly from the discrete distribution $Y_i \sim F_{Discrete}((0, 1, 2)^T, \mathbf{p}_i)$, where $\mathbf{p}_i = (p_{i0}, p_{i1}, p_{i2})^T$ depends on the specified link function. Once the samples are generated, the MLEs of the parameters (β, ρ) are computed following the estimation procedures detailed in Section 2 for each corresponding link function. In addition, the standard errors associated with the MLEs are also calculated to assess the variability of the parameter estimates. This simulation process is repeated 10,000 times for each link function, and the resulting values for bias, average standard error (ASE), and coverage probability (CP) are presented in Tables 3 and 4.

Table 3 shows that the values of the three evaluation metrics, bias, ASE, and CP, fall within acceptable and reasonable ranges. As the sample size increases, while keeping other parameters constant, both the bias and ASE of the parameters (β, ρ) decrease under all five link functions. A closer analysis of the three metrics across different link functions reveals that, for a fixed sample size and correlation coefficient, the logistic link function tends to exhibit the highest ASE for the parameter β . However, when using the Fisher bounded algorithm to estimate β , the ASE is lower than the other link functions, suggesting that the Fisher bounded algorithm performs better than the QLB algorithm in estimation accuracy.

Further examination of the probit link function shows that the estimated parameters (β, ρ) demonstrate smaller bias and lower ASE, with the CP ranging from 94% to 96%, which is within an acceptable range. In contrast, under the DE link function, when $\rho = 0.7$ and $n = 400$, the bias and ASE for the parameters (β, ρ) are higher. Additionally, for smaller sample sizes, under the log–log and complementary log–log link functions, the CP for the parameters (β, ρ) ranges from 93% to 97%, which is slightly wider than the CP observed for the probit link function.

As shown in Table 4, the overall performance for Case (2) is quite similar to that observed in Case (1). The logistic link function shows the largest ASE for β . The three metrics demonstrate a more pronounced advantage for the probit link function than other link functions. While the CP for (β, ρ) improves under the complementary log–log link function, the bias and ASE are still higher than those for the probit link function. The performance of the double exponential and log–log link functions remains similar to that observed in Case (1), with comparable results for the three metrics.

In summary, all five link functions perform within an acceptable range, validating the algorithm's convergence. The probit link function demonstrates the most robust performance and can be considered the preferred choice. The logistic link function exhibits a higher ASE related to the algorithm. The performance of the CLL and LL link functions fluctuates significantly, reflecting each link function's characteristic suitability for extreme event probabilities. The DE link function shows poor stability under high correlation and large sample sizes.

Table 3. Three metrics (%) of MLEs under the different link functions for Case (1).

ρ	MLEs	Index	$n = 200$					$n = 400$				
			LL	CLL	L	DE	Pt	LL	CLL	L	DE	Pt
0.3	$\hat{\beta}_0$	Bias	−2.17	−1.15	−2.58	−2.13	−1.09	−2.13	0.05	−2.54	−1.23	−0.65
		ASE	31.8	23.7	111	24.0	20.8	24.8	17.3	85.2	18.2	15.9
		CP	95.5	95.4	95.5	95.6	95.3	95.1	94.7	95.3	94.9	95.2
	$\hat{\beta}_1$	Bias	5.27	2.65	6.94	4.43	2.33	4.97	0.25	6.30	3.16	1.91
		ASE	67.9	54.5	277	52.9	47.1	59.4	39.1	210	40.1	36.1
		CP	93.6	95.3	95.6	95.7	95.3	94.9	94.3	95.4	95.1	95.4
	$\hat{\rho}$	Bias	−0.76	−0.71	−0.52	−0.53	−0.49	−0.44	−0.31	−0.29	−0.33	−0.32
		ASE	10.3	7.18	6.94	6.97	6.96	6.90	4.71	4.86	4.94	5.00
		CP	95.2	93.0	94.3	94.7	94.6	95.5	95.3	94.2	94.7	94.4
0.5	$\hat{\beta}_0$	Bias	−2.03	−0.84	0.56	−2.41	−1.10	−0.03	−0.07	−0.06	−0.65	−0.16
		ASE	26.9	26.7	145	30.1	26.6	17.6	18.5	90.5	18.7	16.3
		CP	95.2	94.8	95.1	94.6	94.8	95.3	95.2	95.1	94.4	94.8
	$\hat{\beta}_1$	Bias	4.28	1.76	−0.29	5.04	2.31	−0.34	0.35	−1.29	1.20	0.25
		ASE	60.1	59.5	356	65.0	59.2	38.5	40.9	224	41.0	37.2
		CP	94.9	94.8	94.5	94.6	95.0	95.2	95.2	94.7	95.2	94.4
	$\hat{\rho}$	Bias	−3.30	−0.36	−0.35	−0.22	−0.22	−0.1	−0.23	−0.23	−0.17	−0.21
		ASE	6.35	6.36	6.22	6.24	6.19	4.45	4.46	4.32	4.47	4.47
		CP	93.3	94.7	94.3	95.6	95.6	95.2	94.8	94.5	94.7	94.9
0.7	$\hat{\beta}_0$	Bias	−1.19	−1.74	−1.40	−0.80	−1.36	−0.65	−1.44	−0.98	−1.81	−0.59
		ASE	26.5	29.3	39.3	27.7	25.4	19.6	19.3	29.3	20.1	17.3
		CP	94.5	93.9	95.2	94.0	94.9	94.3	95.2	94.9	95.0	95.1
	$\hat{\beta}_1$	Bias	2.88	3.78	2.67	4.07	2.61	1.80	2.91	2.22	2.24	0.95
		ASE	60.4	64.9	90.1	61.2	59.7	43.8	42.5	68.7	44.8	39.3
		CP	94.5	95.3	95.2	93.9	94.7	94.2	94.7	95.1	94.7	94.9
	$\hat{\rho}$	Bias	−0.32	0.05	−0.23	−0.21	−0.16	−0.25	0.04	0.01	−0.02	−0.07
		ASE	5.19	5.13	6.00	6.63	5.21	3.61	3.68	3.59	4.59	3.79
		CP	94.8	94.3	94.4	93.9	94.6	95.0	95.1	94.5	94.9	94.5

Note. LL: Log–log, CLL: Complementary log–log, L: Logistic, DE: Double exponential, Pt: Probit.

Table 4. Three metrics (%) of MLEs under the different link functions for Case (2).

ρ	MLEs	Index	$n = 200$					$n = 400$				
			LL	CLL	L	DE	Pt	LL	CLL	L	DE	Pt
0.3	$\hat{\beta}_0$	Bias	−2.62	−3.03	−8.37	−3.53	−1.17	−1.46	−1.54	−1.85	−3.3	−0.77
		ASE	59.1	45.2	74.3	53.5	53.4	37.3	36.9	55.5	34.2	33.9
		CP	93.7	95.2	94.6	94.9	94.5	95.6	95.1	95.3	94.9	94.8
	$\hat{\beta}_1$	Bias	2.15	3.59	4.42	6.59	3.23	1.43	1.33	3.35	2.04	2.68
		ASE	61.7	53.9	73.1	57.7	51.1	36.9	56.4	57.2	37.8	34.7
		CP	94.5	94.8	95.1	94.9	94.6	95.1	94.5	94.7	94.5	94.7
	$\hat{\beta}_2$	Bias	1.14	0.94	−2.08	−2.79	−3.01	−0.32	−0.87	−2.11	−0.63	−0.06
		ASE	79.5	74.9	124	72.6	76.9	56.4	52.1	83.3	49.9	54.6
		CP	94.8	94.8	94.3	95.3	94.6	95.3	94.7	95.4	95.7	94.5
	$\hat{\beta}_3$	Bias	6.98	7.65	23.9	6.92	2.48	0.75	2.09	5.64	3.83	1.99
		ASE	105	96.4	137	100	97.6	71.7	70.7	111	70.4	68.4
		CP	95.0	95.1	95.8	95.3	94.8	95.1	95.6	95.7	95.2	94.7
	$\hat{\rho}$	Bias	−1.22	−1.19	−0.87	−0.97	−0.70	−0.45	−0.28	−0.45	−0.69	−0.34
		ASE	7.29	6.96	6.79	7.15	7.10	5.21	5.17	4.78	5.05	4.88
		CP	94.4	95.3	95.6	94.3	94.9	93.8	94.7	95.2	94.9	94.8

Table 4. Cont.

ρ	MLEs	Index	$n = 200$					$n = 400$				
			LL	CLL	L	DE	Pt	LL	CLL	L	DE	Pt
0.5	$\hat{\beta}_0$	Bias	−3.8	−2.83	−5.50	−5.95	−1.56	−2.16	−2.79	−1.14	−1.47	−0.11
		ASE	53.2	56.3	81.5	54.2	54.2	38.7	39.3	60.0	39.8	39.1
		CP	94.9	95.4	94.2	95.5	95.1	95.4	94.7	95.6	95.3	95.0
	$\hat{\beta}_1$	Bias	5.26	4.11	5.66	10.5	4.0	1.90	3.94	1.61	5.35	3.42
		ASE	60.0	56.6	79.7	58.3	55.8	37.0	44.0	61.1	43.8	38.9
		CP	94.2	94.8	95.4	94.3	94.5	95.8	94.4	95.1	94.3	94.6
	$\hat{\beta}_2$	Bias	−1.68	−1.64	1.02	−4.85	−4.24	0.50	−1.37	−0.83	−3.67	−1.03
		ASE	78.8	86.8	136	86.2	80.9	57.4	57.3	86.1	58.7	55.3
		CP	94.2	95.0	94.9	95.1	95.3	95.2	95.3	94.7	95.3	94.8
	$\hat{\beta}_3$	Bias	6.69	6.25	12.5	12.3	2.33	4.94	5.86	0.19	3.05	1.29
		ASE	106	99.2	152	107	103	75.9	73.5	123	77.1	74.4
		CP	94.4	95.2	95.4	96.0	94.5	94.6	94.7	94.2	95.5	94.8
	$\hat{\rho}$	Bias	−0.56	−0.72	−0.66	−0.47	−0.43	−0.51	−0.32	−0.32	−0.40	−0.20
		ASE	6.64	6.42	6.45	6.44	6.43	4.76	4.55	4.42	4.54	4.48
		CP	95.1	95.6	93.6	94.4	94.7	95.0	95.1	95.3	94.8	95.3
0.7	$\hat{\beta}_0$	Bias	−4.37	−5.69	−9.83	−5.41	−4.15	−1.13	−1.57	−2.52	−1.83	−0.33
		ASE	60.3	59.5	84.0	57.6	56.9	42.9	38.6	63.8	42.6	38.1
		CP	94.6	94.8	94.5	95.0	94.9	94.5	95.2	95.9	95.2	94.5
	$\hat{\beta}_1$	Bias	6.30	5.67	6.33	9.61	4.16	2.21	2.35	5.06	2.91	2.43
		ASE	57.7	64.1	84.9	66.7	65.3	45.0	42.3	65.1	42.8	40.3
		CP	95.2	94.3	95.1	94.5	94.9	94.9	95.4	95.2	95.0	94.9
	$\hat{\beta}_2$	Bias	−2.79	−0.19	−4.13	−2.61	−2.38	0.12	−1.34	−3.43	0.60	0.27
		ASE	98.0	98.4	139	91.6	87.5	63.8	59.5	95.9	61.3	57.8
		CP	94.4	94.5	95.3	95.0	94.8	95.3	95.7	96.2	94.9	95.7
	$\hat{\beta}_3$	Bias	5.76	11.3	29.1	6.93	5.97	5.71	4.58	7.35	3.91	1.19
		ASE	117	121	157	119	112	78.9	79.1	127	88.4	78.9
		CP	94.5	95.4	96.4	95.2	94.8	95.8	94.7	95.6	95.0	95.0
	$\hat{\rho}$	Bias	−0.38	−0.51	−0.54	−0.71	−0.57	−0.19	−0.21	−0.27	−0.37	−0.21
		ASE	5.56	5.36	5.24	5.91	5.27	3.77	3.75	3.36	4.29	3.78
		CP	94.1	94.6	94.7	95.0	95.1	95.4	95.2	95.1	94.5	94.7

Note. LL: Log–log, CLL: Complementary log–log, L: Logistic, DE: Double exponential, Pt: Probit.

3.2. Comparison of Algorithms

Under various link functions, we compare the effectiveness of algorithms by the average iteration count and the computing time. An algorithm is considered more efficient if it achieves a desired level of accuracy with fewer iterations or in less computation time.

Let $m = 200, 400$, and $\rho = 0.3, 0.5, 0.7$, and the other parameter settings refer to Case (1). The samples $\{Y_i\}$ are randomly generated from the discrete distribution $Y_i \sim F_{Discrete}((0, 1, 2)^T, \mathbf{p}_i)$, where $\mathbf{p}_i = (p_{i0}, p_{i1}, p_{i2})^T$ depends on the specified link function. Regarding parameter estimations, we use the QLB-based NR algorithm with the logistic link function, while for the other four link functions, we employ the Fisher-bound-based NR algorithm. All experiments were repeated 500 times.

Figures 1 and 2 provide the frequency distributions of the average iteration numbers under the different link functions. The average number of iterations under the logistic link function is significantly higher than that of the other four link functions across all parameter settings. On the other hand, we fixed the iteration number to 100 and recorded their time cost. The average iteration time for five link functions is shown in Table 5. Under all parameter settings, the average iteration time for the logistic link function is also notably longer than for the other link functions.

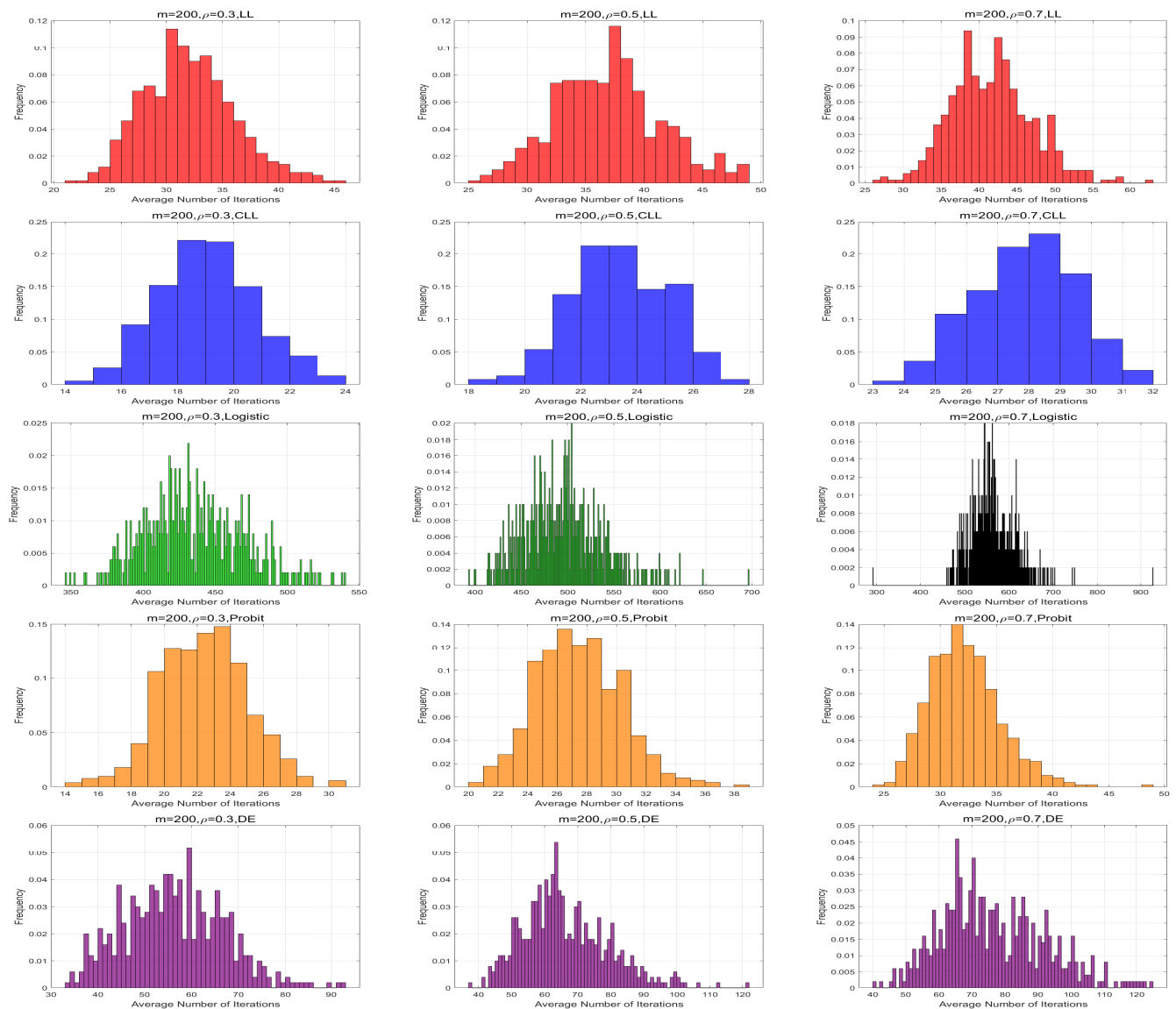


Figure 1. Comparison of iteration frequencies under different link functions when the $m = 200$.

These findings suggest that the computational performance of the Fisher-bound-based NR algorithm is substantially more efficient than the QLB-based NR algorithm. Through comparisons across different link functions, we observe that both the average number of iterations and the average computation time increase with the sample size and the correlation coefficient. This demonstrates that both factors have a significant impact on the computational burden of the algorithms. Moreover, it is worth noting that the Fisher-bound-based NR algorithm with the CLL link function consistently exhibits the best performance across all parameter settings.

Table 5. Comparison of the average iteration time (ms) under five link functions.

n	ρ	LL	CLL	L	DE	Pt
200	0.3	6.19	5.30	11.9	7.75	5.56
	0.5	7.26	6.24	12.3	9.25	6.34
	0.7	8.15	7.06	13.1	10.1	7.28
400	0.3	54.2	46.7	104	70.8	47.2
	0.5	63.1	49.3	115	80.2	57.1
	0.7	74.1	50.7	146	87.4	68.3

Note. LL: Log-log, CLL: Complementary log-log, L: Logistic, DE: Double exponential, Pt: Probit.

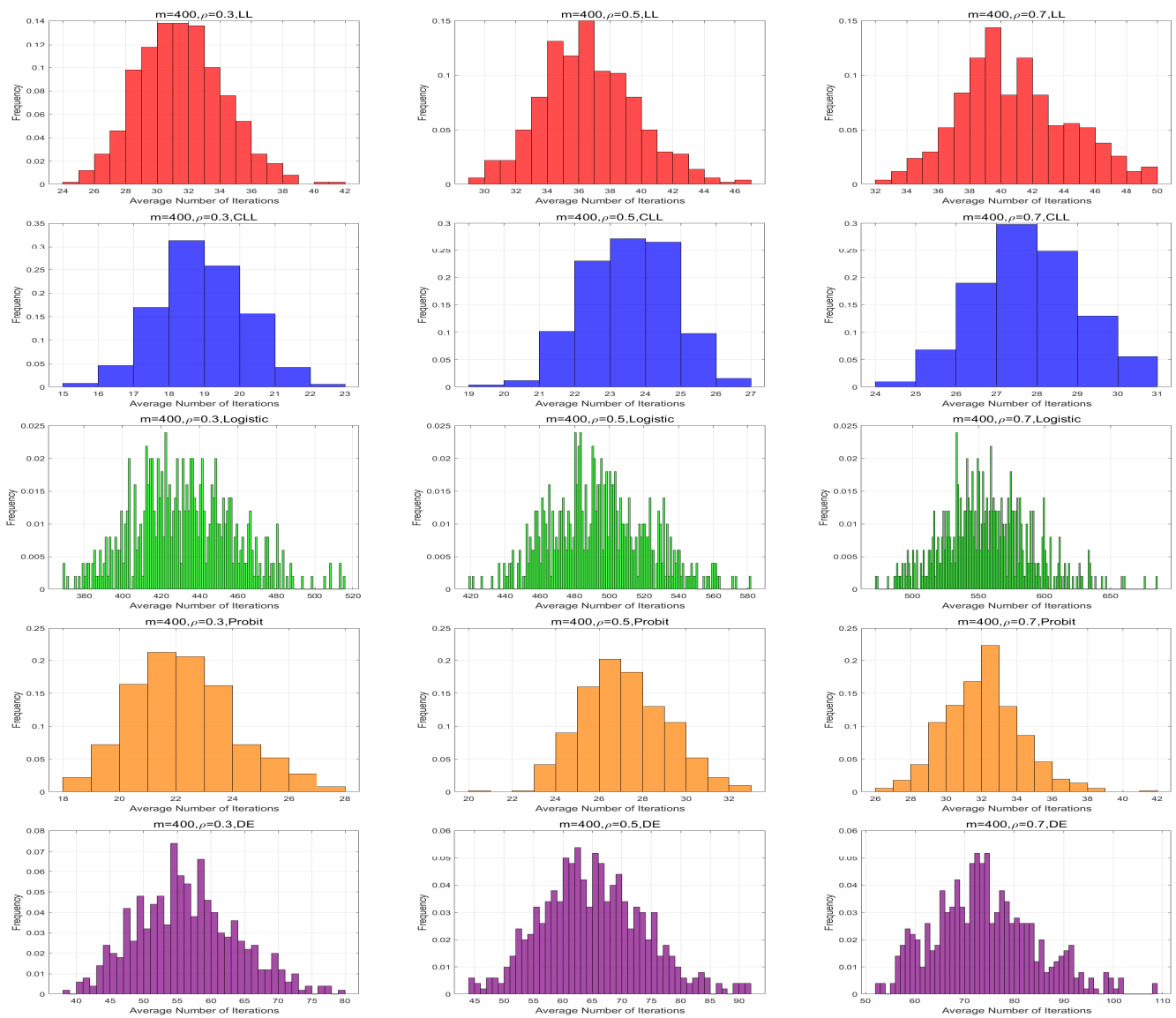


Figure 2. Comparison of iteration frequencies under different link functions when the $m = 400$.

3.3. Comparison of Test Statistics

Under the hypothesis test (5), we evaluate the performance of the three statistics for given link functions with parameter settings. Take the sample size $n = 200, 400$ and the correlation coefficient $\rho = 0.3, 0.5, 0.7$. The covariate vector x_i is defined in Cases (1) and (2). Other parameter settings are based on the two cases: (I) $q = 2$, $\beta = (\beta_0, \beta_1)^T$, $\beta_0 = \beta_1 = 0$, $C = (1, -1)$. (II) $q = 4$, $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)^T$, $\beta_0 = \beta_1 = \beta_2 = \beta_3 = 0$, and

$$C = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}.$$

Similarly, the samples $\{Y_i\}$ are generated from the discrete distribution $F_{Discrete}((0, 1, 2)^T, \mathbf{p}_i)$ for 10,000 times, where $\mathbf{p}_i = (p_{i0}, p_{i1}, p_{i2})^T$. The empirical type I error rate is calculated by dividing the number of rejections of the null hypothesis H_0 by the total number of simulations, which is 10,000. The empirical TIEs are summarized in Table 6. Power refers to the probability of correctly rejecting the null hypothesis H_0 when it is false. We consider sample sizes $n = 50, 80, 100, 200, 400$, and $\rho = 0.3, 0.5, 0.7$. The following two cases of regression coefficient vectors are considered: (A) $\beta = (-1, 2)^T$ and (B) $\beta = (-1, 2, -1, 2)^T$.

The other parameter configurations remain consistent with Cases (I) and (II). For each $i = 1, \dots, n$, we generate $Y_i \sim F_{Discrete}((0, 1, 2)^T, \mathbf{p}_i)$ 10,000 times, where $\mathbf{p}_i = (p_{i0}, p_{i1}, p_{i2})^T$. The empirical power of each test is computed as the proportion of rejections at the 0.05 significance level over 10,000 simulations. Figures 3 and 4 display the empirical powers of the three test statistics across five different link functions. A comparison of the results shows that both the T_L and T_{SC} tests consistently maintain empirical type I error rates (TIEs) close to the nominal level of 0.05 across all link functions considered (LL, CLL, L, DE, and Pt), indicating robust performance. In contrast, the T_W test tends to yield lower empirical TIEs under the double exponential link, though its TIEs for other link functions fluctuate around the nominal level. Furthermore, empirical TIEs are generally higher when $q = 4$ compared to $q = 2$, with this effect being more noticeable for smaller sample sizes.

Table 6. Empirical type I error rates (%) for Case (I) and Case (II) under the different link functions.

n	ρ	Statistics	Case (I)					Case (II)				
			LL	CLL	L	DE	Pt	LL	CLL	L	DE	Pt
200	0.3	T_L	5.0	5.3	4.8	4.7	5.6	5.4	4.7	4.8	5.2	5.0
		T_W	5.3	5.4	4.8	3.3	5.0	4.8	4.8	4.7	3.7	4.7
		T_{SC}	5.3	5.4	4.7	4.8	4.8	4.7	5.0	4.9	4.9	4.6
	0.5	T_L	5.4	4.5	5.2	5.0	5.1	5.3	4.8	5.0	5.2	4.8
		T_W	5.4	5.1	4.9	3.7	5.3	5.3	4.8	4.9	4.0	4.9
		T_{SC}	5.1	5.2	4.7	4.9	4.6	5.1	4.6	4.8	5.1	5.1
	0.7	T_L	4.8	5.4	4.7	4.7	5.2	5.2	5.1	4.8	5.5	4.8
		T_W	5.2	4.6	5.5	2.9	5.2	5.2	5.0	4.8	4.1	4.8
		T_{SC}	5.0	5.0	5.4	4.8	5.2	4.7	5.5	4.7	5.4	4.7
400	0.3	T_L	4.7	4.8	5.2	5.0	5.2	4.7	4.8	5.2	5.0	5.2
		T_W	4.8	4.8	4.7	3.7	4.7	4.8	4.8	4.7	3.7	4.7
		T_{SC}	4.7	5.0	4.9	4.9	4.6	4.7	5.0	4.9	4.9	4.6
	0.5	T_L	5.3	4.8	5.0	5.2	4.8	5.3	4.8	5.0	5.2	4.8
		T_W	5.4	5.1	4.9	3.7	5.3	5.3	4.8	4.9	4.0	4.9
		T_{SC}	5.1	4.6	4.8	5.1	5.1	5.1	4.6	4.8	5.1	5.1
	0.7	T_L	5.2	5.1	4.8	5.5	4.8	5.2	5.1	4.8	5.5	4.8
		T_W	5.2	5.0	4.8	4.1	4.8	5.2	5.0	4.8	4.1	4.8
		T_{SC}	4.7	5.5	4.7	5.4	4.7	4.7	5.5	4.7	5.4	4.7

Note. LL: Log-log, CLL: Complementary log-log, L: Logistic, DE: Double exponential, Pt: Probit.

Figure 3 illustrates that increasing the sample size enhances the power of all three test statistics. Power decreases slightly for a fixed sample size as ρ increases. The T_W statistic demonstrates the best performance, while T_L and T_{SC} produce comparable results across all parameter settings. Among the link functions, the probit link generally yields the highest power, especially for small sample sizes, whereas the logistic link results in the lowest. The log-log, complementary log-log, and double exponential links exhibit similar power levels.

From Figure 4, we observe that the performance of the three test statistics for $q = 4$ is generally similar to that for $q = 2$, though there are some differences. Specifically, compared to $q = 2$, the power of all test statistics decreases when $q = 4$, with the T_L and T_{SC} test statistics showing a more pronounced decline across all parameter settings and link functions.

In addition to the empirical findings, theoretical analysis suggests that the Wald test achieves higher power under the probit link due to its derivative structure, which yields greater Fisher information and a larger non-centrality parameter λ [30].

Moreover, considering that the T_W test statistic under the probit link function shows robust empirical TIEs, it is recommended as the most effective test for assessing the significance of confounding effects on the response rate.

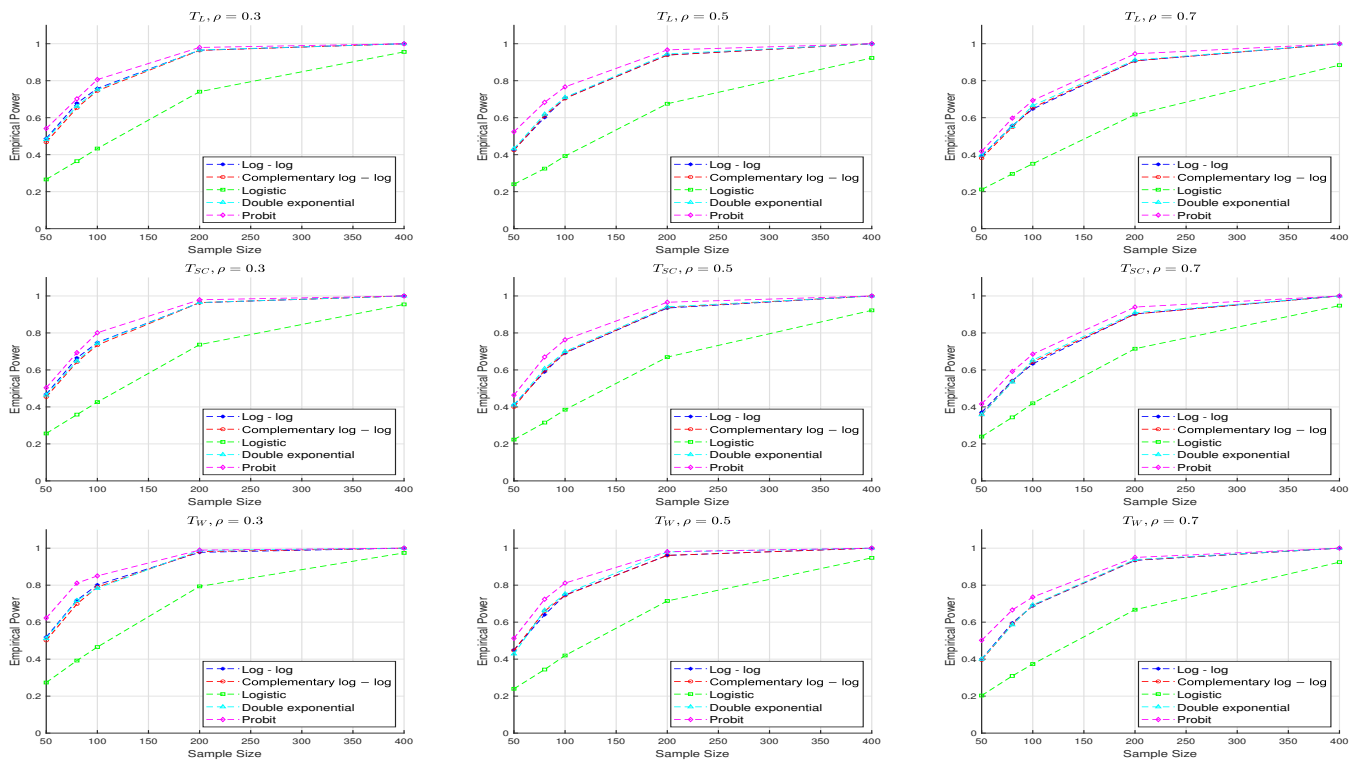


Figure 3. Comparison of the power (%) under different link functions in Case (A).

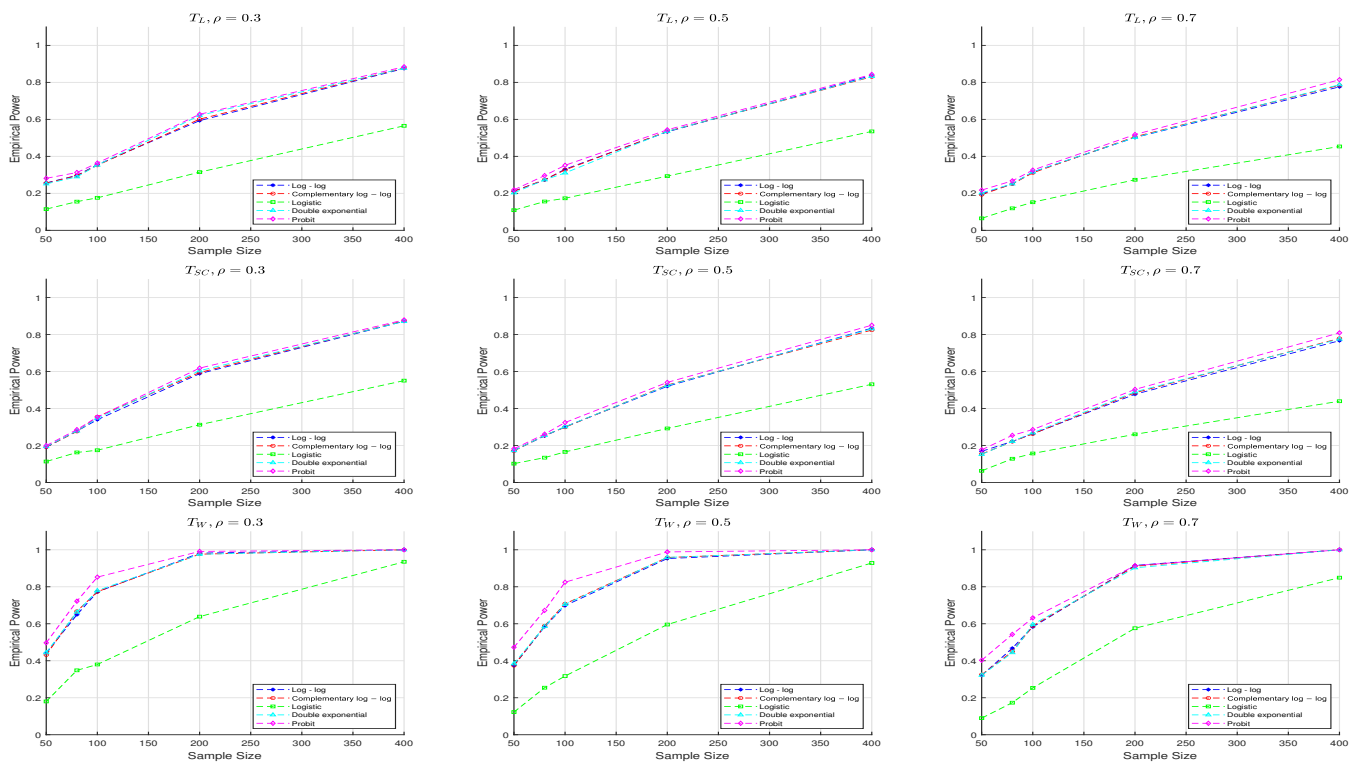


Figure 4. Comparison of the power (%) under different link functions in Case (B).

4. An Application

In this section, we analyze a dataset from a survey conducted in the Varamin region of Iran, focusing on individuals aged 50 and above with visual impairment [31]. The dataset includes nearly 3000 participants, categorized into seven groups ($g = 1, 2, \dots, 7$). Responses are recorded as no blindness, unilateral blindness, or bilateral blindness ($l = 0, 1, 2$). The

dataset's structure is presented in Table 7. Since the response probabilities in our model are defined at the individual level, we preprocess the data as follows to ensure analytical accuracy and consistency:

- (1) Use the midpoints of the age groups as a proxy for age to account for potential confounding.
- (2) Assume that individuals within the same group share identical covariates, resulting in uniform response rates within each group;
- (3) Scale all data values by a factor of 1/100 to reduce the impact of magnitude differences on computation.

Table 7. Prevalence of blindness by age groups ($g = 7$).

Age (g)	Blindness (l)		
	None (0)	Unilateral (1)	Bilateral (2)
50–54 yrs	964	23	2
55–59 yrs	541	17	8
60–64 yrs	469	18	4
65–69 yrs	257	16	5
70–74 yrs	242	32	3
75–79 yrs	127	30	9
80+ yrs	104	29	10

We construct a generalized linear model to explore the relationship between age and disease rates. The model incorporates five link functions specified in Section 2. The MLEs of (β, ρ) , their standard errors (SE), and the asymptotic 95% confidence interval widths (CIW) can be computed using the method described in Section 3 under the corresponding link function.

As shown in Table 8, $\hat{\rho}$ is approximately 0.273 under all link functions, indicating a positive correlation between the responses of both eyes of a patient. Among them, the Pt model exhibits the lowest standard errors and the narrowest confidence intervals, demonstrating the most robust performance. The LL link function and the DE link function yield similar results. Still, the LL link function is slightly better because it is more suitable for modeling low probability events, which aligns with the dataset's characteristics [32]. In contrast, the CLL link function is designed for modeling high-probability events, leading to inconsistencies between the model assumptions and the data characteristics, resulting in a negative regression coefficient. The L link function has the largest standard error for the intercept coefficient. As shown in Table 9, the Q-Q plots and residual plots under all link functions show little difference. Overall, the models appear to provide similar fits. Furthermore, according to the AIC and BIC results (Figure 5), the Pt model demonstrates the best fitting efficiency among all the models. Given the low disease probability in this dataset [32], which does not match the overall assumptions of the simulated data, different link functions significantly impact the parameter estimation results, particularly for β_0 and β_1 . However, the conclusions drawn from the model comparisons are consistent with those from the simulation study.

Table 8. Parameter estimates under different link functions.

Index	Parameters	LL	CLL	L	DE	Pt
MLE	β_0	−7.984	2.960	−9.003	−8.842	−4.401
	β_1	0.084	−0.029	0.091	0.088	0.042
	ρ	0.272	0.274	0.273	0.272	0.272
SE	β_0	0.156	0.487	0.506	0.467	0.229
	β_1	0.242	0.453	0.073	0.066	0.034
	ρ	0.041	0.041	0.041	0.077	0.041
CIW	β_0	0.612	1.910	1.980	1.830	0.898
	β_1	0.948	0.178	0.286	0.259	0.133
	ρ	0.159	0.159	0.159	0.159	0.159

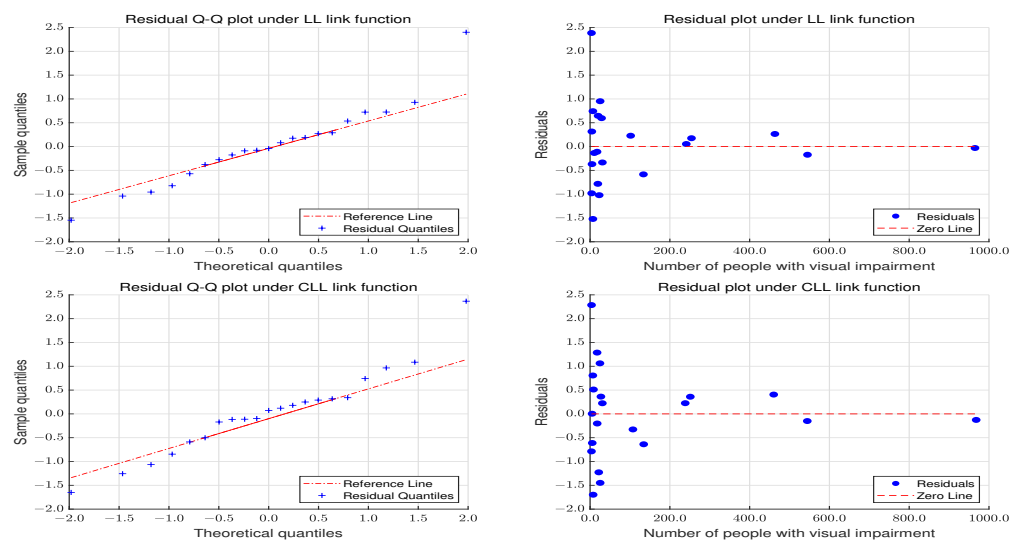
Note. LL: Log–log, CLL: Complementary log–log, L: Logistic, DE: Double exponential, Pt: Probit.

Table 9. Values of AIC and BIC under different link functions.

Tests	LL	CLL	L	DE	Pt
AIC	1540.8	1544.8	1540.9	1542.6	1540.7
BIC	1558.7	1560.6	1562.8	1558.9	1558.6

Note. LL: Log–log, CLL: Complementary log–log, L: Logistic, DE: Double exponential, Pt: Probit.

Lastly, we need to test the hypothesis (5), where $C = (0, 1)$ and $\beta = (\beta_0, \beta_1)^T$. The values of the three statistics T_L , T_W , and T_{SC} are shown in Table 10. The test statistic T_W values vary slightly with different link functions, and the test statistic T_L exhibits a similar pattern. In contrast, the T_{SC} test statistic remains consistent under all link functions. This indicates that the choice of link function significantly impacts some statistics but has little impact on others. Under all link functions, the values of the T_W test statistic are greater than the other two, indicating that the T_W test statistic performs better. Since all p -values are less than 0.05, we reject the null hypothesis and conclude that age significantly affects disease rates. Similarly, this result is consistent with the observations in the simulation.


Figure 5. Cont.

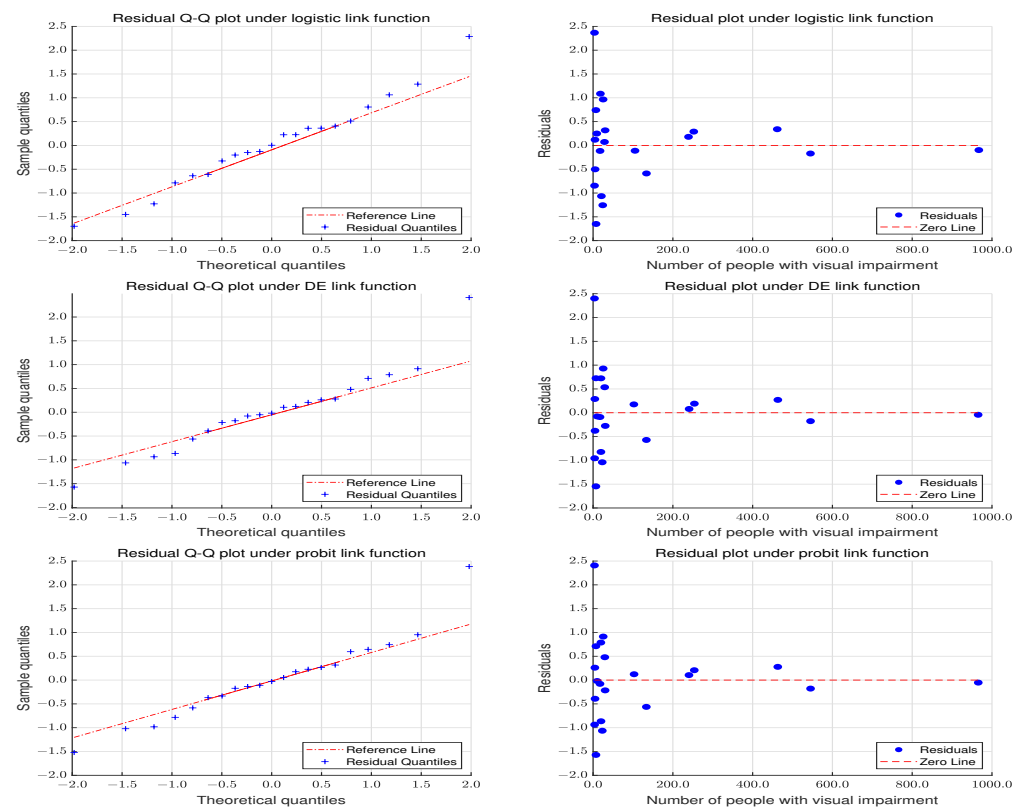


Figure 5. Comparison of Q-Q and residual plots.

Table 10. Values of three statistics under different link functions.

Tests	Results	LL	CLL	L	DE	Pt
T_L	Statistic	152.404	158.871	158.713	158.967	156.832
	p -value	5.17×10^{-35}	2.00×10^{-36}	2.16×10^{-36}	1.90×10^{-36}	5.57×10^{-36}
T_W	Statistic	226.148	223.229	215.056	213.505	232.597
	p -value	4.13×10^{-51}	1.79×10^{-50}	1.08×10^{-48}	1.90×10^{-36}	5.57×10^{-36}
T_S	Statistic	178.743	178.743	178.743	178.742	178.743
	p -value	9.12×10^{-41}	9.12×10^{-41}	9.12×10^{-41}	9.12×10^{-41}	9.12×10^{-41}

Note. LL: Log-log, CLL: Complementary log-log, L: Logistic, DE: Double exponential, Pt: Probit.

5. Discussion and Conclusions

This paper develops a generalized linear model to investigate the relationship between confounding effects and response rates. Since the log-likelihood function involving various link functions is generally non-concave, we adopt a novel MM algorithm to construct a surrogate function that approximates the logistic log-likelihood, ensuring computational feasibility and stability. Five specific link functions are considered: log-log, complementary log-log, logistic, double exponential, and probit. We also describe three parameter estimation methods: the NR, QLB, and Fisher-bounded algorithms. Additionally, three hypothesis testing procedures, the likelihood ratio test, the Wald test, and the score test, are introduced to assess the significance of regression coefficients under different link functions. Simulation results indicate that, regardless of the link function or estimation method, the MLEs closely approximate the actual values, and the corresponding confidence intervals fall within reasonable bounds. In terms of convergence, the Fisher bounded algorithm outperforms the others. Comparison of the test statistics reveals that the point and interval estimates of (β, ρ) are the most accurate under the probit link function. Under the double exponential link, the empirical type I error rate of T_W is notably lower than the nominal level, while T_L and T_{SC} remain close to 0.05. All three tests control type I error for the other

link functions. Furthermore, T_W consistently demonstrates the highest statistical power across all scenarios, increasing power as the sample size grows. We therefore recommend the use of T_W . The logistic link yields the lowest power among the five link functions, while the probit link yields the highest. Finally, the proposed method is applied to real data from a study of visually impaired individuals in Iran, demonstrating its practical applicability and effectiveness.

Lin et al. [23] used a logistic regression to study the relationship between ophthalmic disease prevalence and patient covariates. It combines the fast QLB algorithm with coordinate ascent to compute the unconstrained MLE of the parameters of interest. Compared with the existing results, our proposed model offers greater generalization ability and achieves better performance in terms of computational efficiency and applicability. Simulation studies and real-data analyses further validate this assumption, demonstrating the practical advantages of our method. In addition, Li et al. [5] proposed a flexible beta-kernel correlation model capable of fitting binary data with a wide range of correlation structures. Compared to our method, their model only shows superior performance in certain scenarios.

The method proposed in this paper can be extended to other intra-class models such as Rosner's or Dallal's model and supports both unilateral and bilateral data structures. However, some unsolved problems exist, such as complex correlations, binary-only outcomes, and small sample behavior. Moreover, the model dependence between random vectors is conducive to more comprehensively capturing the information between covariates. However, our current method has not yet constructed the joint distribution of random variables, and thus it is difficult to delve into the specific relationships between variables. Moreover, we have only focused on the individual distributions of random variables. Future research can further construct the joint distribution of random variables to explore the specific relationships between them thoroughly. Based on [33], we can model the dependence between random vectors, thereby comprehensively capturing all the information between covariates. These are important directions for future research, and we will further investigate these interesting issues in the general case.

Author Contributions: J.C.: Methodology, software, writing—original draft preparation; Z.L.: writing—review and editing, supervision, funding acquisition; K.M.: writing—review and editing, K.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the Central Guidance for Local Science and Technology Development Fund (Grant No. ZYYD2025ZY20).

Data Availability Statement: Real data was referred to are from Rajavi et al. [31].

Acknowledgments: We thank the editor and referees for their insightful comments and constructive suggestions.

Conflicts of Interest: The authors declare that they have no conflicts of interest.

Appendix A

Appendix A.1. Calculation of the Three Response Probabilities

Let $Pr(Z_{ik} = 1) = \pi_i (0 \leq \pi_i \leq 1)$ be the probability of the i th patient's k th organ being cured. Thus, the variable Z_{ik} follows a Bernoulli distribution, indicating the cure status of the k th eye of the i th patient. Based on this, the expected value and variance of Z_{ik} satisfy $E(Z_{ik}) = \pi_i$ and $D(Z_{ik}) = \pi_i(1 - \pi_i)$. Under Donner's model, we have

$\text{Corr}(Z_{ik}, Z_{i(3-k)}) = \rho$ for $k = 1, 2$, and $i = 1, \dots, n$, where $\rho(|\rho| \leq 1)$ reflects the correlation of the paired data. Thus,

$$\text{Corr}(Z_{ik}, Z_{i(3-k)}) = \frac{E(Z_{ik}Z_{i(3-k)}) - E(Z_{ik})E(Z_{i(3-k)})}{\sqrt{D(Z_{ik})D(Z_{i(3-k)})}} = \rho.$$

Note that $p_{2i} = P(Z_{i1} = 1, Z_{i2} = 1) = E(Z_{ik}Z_{i(3-k)}) = \rho\sqrt{D(Z_{ik})D(Z_{i(3-k)})} + E(Z_{ik})E(Z_{i(3-k)})$. Then, $p_{2i} = \rho\pi_i + (1 - \rho)\pi_i^2$. Moreover, $p_{1i} = P(Z_{i1} = 1, Z_{i2} = 0) + P(Z_{i1} = 0, Z_{i2} = 1) = 2(\pi_i - p_{2i}) = 2\pi_i(1 - \rho)(1 - \pi_i)$. Since $p_{0ij} + p_{1ij} + p_{2ij} = 1$, we have $p_{0i} = \rho(1 - \pi_i) + (1 - \rho)(1 - \pi_i)^2$.

Appendix A.2. The Construction of the $Q(\pi|\pi^{(t)}, \rho)$

To derive the asymptotic expression of the log-likelihood function, we employ the discrete version of Jensen's inequality for concave functions:

$$h(\alpha^T z) \geq \sum_{i=1}^n \frac{a_i z_i(t)}{\alpha^T z(t)} h\left(\frac{\alpha^T z(t)}{z_i(t)} z_i\right),$$

where $h(\cdot)$ is an arbitrary concave function, $\alpha = (\alpha_1, \dots, \alpha_n)^T$, $z = (z_1, \dots, z_n)^T$ and $z^{(t)} = (z_1^{(t)}, \dots, z_n^{(t)})^T$ are three positive vectors. Define $\alpha_1 = (1, \rho)^T$, $\alpha_2 = (\rho, 1)^T$, $z = (\pi_i, 1 - \pi_i)^T$ and $h(\cdot) = \log(\cdot)$. Based on this inequality and the model structure, p_{0i} and p_{2i} are asymptotically approximated as follows:

$$\begin{aligned} \log p_{0i} &\geq \log(1 - \pi_i) + \frac{\rho\pi_i^{(t)}}{\rho\pi_i^{(t)} + (1 - \pi_i^{(t)})} \log\left[\frac{\rho\pi_i^{(t)} + (1 - \pi_i^{(t)})}{\pi_i^{(t)}} \pi_i\right] \\ &\quad + \frac{(1 - \pi_i^{(t)})}{\rho\pi_i^{(t)} + (1 - \pi_i^{(t)})} \log\left[\frac{\rho\pi_i^{(t)} + (1 - \pi_i^{(t)})}{1 - \pi_i^{(t)}} (1 - \pi_i)\right], \\ \log p_{2i} &\geq \log \pi_i + \frac{\pi_i^{(t)}}{\pi_i^{(t)} + \rho(1 - \pi_i^{(t)})} \log\left[\frac{\pi_i^{(t)} + \rho(1 - \pi_i^{(t)})}{\pi_i^{(t)}} \pi_i\right] \\ &\quad + \frac{\rho(1 - \pi_i^{(t)})}{\pi_i^{(t)} + \rho(1 - \pi_i^{(t)})} \log\left[\frac{\pi_i^{(t)} + \rho(1 - \pi_i^{(t)})}{1 - \pi_i^{(t)}} (1 - \pi_i)\right]. \end{aligned}$$

Hence, the log function (1) is expressed as

$$\begin{aligned} \ell(\pi, \rho) &= \sum_{i=1}^n \left\{ [I(Y_i = 0) + I(Y_i = 1)] \log(1 - \pi_i) + [I(Y_i = 2) + I(Y_i = 1)] \log \pi_i \right. \\ &\quad \left. + I(Y_i = 0) \log(\rho\pi_i - \pi_i + 1) + I(Y_i = 2) \log[\pi_i + \rho(1 - \pi_i)] + c_{i1}^{(t)}(\rho) \right\}, \end{aligned} \quad (\text{A1})$$

where c_{i1} is a function of ρ but not depending on π_i for $i = 1 \dots n$. By the asymptotic approximations of p_{0i} and p_{2i} , we derive

$$\begin{aligned} \ell(\pi, \rho) &\geq \sum_{i=1}^n \left\{ \left[I(Y_i = 1) + I(Y_i = 2) + \frac{I(Y_i = 2)\pi_i^{(t)}}{\pi_i^{(t)} + \rho(1 - \pi_i^{(t)})} + \frac{I(Y_i = 0)\rho\pi_i^{(t)}}{\rho\pi_i^{(t)} + (1 - \pi_i^{(t)})} \right] \log(\pi_i) \right. \\ &\quad \left. + \left[I(Y_i = 0) + I(Y_i = 1) + \frac{I(Y_i = 2)\rho(1 - \pi_i^{(t)})}{\pi_i^{(t)} + \rho(1 - \pi_i^{(t)})} + \frac{I(Y_i = 0)(1 - \pi_i^{(t)})}{\rho\pi_i^{(t)} + (1 - \pi_i^{(t)})} \right] \log(1 - \pi_i) + c_{i1}^{(t)}(\rho) \right\}. \end{aligned}$$

Let

$$a_i(\pi_i^{(t)}, \rho) = I(Y_i = 1) + I(Y_i = 2) + \frac{I(Y_i = 2)\pi_i^{(t)}}{\pi_i^{(t)} + \rho(1 - \pi_i^{(t)})} + \frac{I(Y_i = 0)\rho\pi_i^{(t)}}{\rho\pi_i^{(t)} + (1 - \pi_i^{(t)})},$$

$$b_i(\pi_i^{(t)}, \rho) = I(Y_i = 0) + I(Y_i = 1) + \frac{I(Y_i = 2)\rho(1 - \pi_i^{(t)})}{\pi_i^{(t)} + \rho(1 - \pi_i^{(t)})} + \frac{I(Y_i = 0)(1 - \pi_i^{(t)})}{\rho\pi_i^{(t)} + (1 - \pi_i^{(t)})},$$

We have

$$Q(\beta|\beta^{(t)}, \rho) = \sum_{i=1}^n \left\{ a_i(\pi_i^{(t)}, \rho) \log(\pi_i) + b_i(\pi_i^{(t)}, \rho) \log(1 - \pi_i) + c_i^{(t)}(\rho) \right\}.$$

Appendix A.3. Fisher Information Matrix

The Fisher information matrix $I(\theta)$ is given by

$$I(\theta) = -E \begin{pmatrix} H(\beta) & q(\theta) \\ q(\theta)^T & \frac{\partial^2 \ell}{\partial \rho^2} \end{pmatrix}, \quad (A2)$$

where

$$H(\beta) = \left(\frac{\partial^2 \ell}{\partial \beta_s \partial \beta_t} \right), \quad s, t = 0, 1, \dots, q-1,$$

$$q(\theta) = \left(\frac{\partial^2 \ell}{\partial \beta_0 \partial \rho}, \dots, \frac{\partial^2 \ell}{\partial \beta_{q-1} \partial \rho} \right)^T,$$

$$\frac{\partial^2 \ell}{\partial \rho^2} = - \sum_{i=1}^n \left\{ \frac{I(Y_i = 1)}{(\rho - 1)^2} + \frac{\pi_i^2 I(Y_i = 0)}{(\rho \pi_i - \pi_i + 1)^2} + \frac{(\pi_i - 1)^2 I(Y_i = 2)}{(\rho + \pi_i - \rho \pi_i)^2} \right\}, \quad i = 1, \dots, n.$$

The elements of $q(\theta)$ are expressed by

$$\frac{\partial^2 \ell}{\partial \beta_s \partial \rho} = \frac{\partial}{\partial \rho} \left(\sum_{i=1}^n \frac{\partial \ell}{\partial \pi_i} \frac{\partial \pi_i}{\partial \beta_s} \right) = \sum_{i=1}^n \frac{\partial^2 \ell}{\partial \rho \partial \pi_i} \frac{\partial \pi_i}{\partial \beta_s} + \sum_{i=1}^n \frac{\partial^2 \pi_i}{\partial \beta_s \partial \rho} \frac{\partial \ell}{\partial \pi_i},$$

$$\frac{\partial^2 \ell}{\partial \beta_s^2} = \frac{\partial}{\partial \beta_s} \left(\sum_{i=1}^n \frac{\partial \ell}{\partial \pi_i} \frac{\partial \pi_i}{\partial \beta_s} \right) = \sum_{i=1}^n \left(\frac{\partial \pi_i}{\partial \beta_s} \right)^2 \left(\frac{\partial^2 \ell}{\partial \pi_i^2} \right) + \sum_{i=1}^n \frac{\partial^2 \pi_i}{\partial \beta_s^2} \frac{\partial \ell}{\partial \pi_i},$$

$$\frac{\partial^2 \ell}{\partial \beta_s \partial \beta_t} = \frac{\partial}{\partial \beta_t} \left(\sum_{i=1}^n \frac{\partial \ell}{\partial \pi_i} \frac{\partial \pi_i}{\partial \beta_s} \right) = \sum_{i=1}^n \left(\frac{\partial \pi_i}{\partial \beta_s} \frac{\partial \pi_i}{\partial \beta_t} \right) \left(\frac{\partial^2 \ell}{\partial \pi_i^2} \right) + \sum_{i=1}^n \frac{\partial^2 \pi_i}{\partial \beta_t \partial \beta_s} \frac{\partial \ell}{\partial \pi_i},$$

where

$$\frac{\partial^2 \ell}{\partial \rho \partial \pi_i} = \frac{I(Y_i = 2)(\rho + 2\pi_i - 2\rho\pi_i)(\pi_i - 1)}{\pi_i^2(\rho + \pi_i - \rho\pi_i)^2} + \frac{I(Y_i = 0)(\rho + 2\pi_i - 2\rho\pi_i - 2)(\pi_i^2 - \rho\pi_i - 1)}{(\pi_i - 1)^2(\rho\pi_i - \pi_i + 1)^2}$$

$$+ \frac{I(Y_i = 0)(2\pi_i - 1)}{(\pi_i - 1)(\rho\pi_i - \pi_i + 1)} + \frac{I(Y_i = 2)(1 - 2\pi_i)}{\pi_i(\rho + \pi_i - \rho\pi_i)},$$

$$\frac{\partial^2 \ell}{\partial \pi_i^2} = \frac{2I(Y_i = 1)}{\pi_i(\pi_i - 1)} - \frac{I(Y_i = 1)(2\pi_i - 1)^2}{\pi_i^2(\pi_i - 1)^2} + \frac{2I(Y_i = 2)(1 - \rho)}{\pi_i(\rho + \pi_i - \rho\pi_i)} - \frac{I(Y_i = 2)(\rho + 2\pi_i - 2\rho\pi_i)^2}{\pi_i^2(\rho + \pi_i - \rho\pi_i)^2}$$

$$- \frac{2I(Y_i = 0)(1 - \rho)}{(\pi_i - 1)(\rho\pi_i - \pi_i + 1)} - \frac{I(Y_i = 0)(2\rho\pi_i - 2\pi_i - \rho + 2)^2}{(\pi_i - 1)^2(\rho\pi_i - \pi_i + 1)^2},$$

and $\frac{\partial^2 \pi_i}{\partial \beta_s \partial \rho} = 0$. Moreover, the specific forms of the derivatives $\frac{\partial \pi_i}{\partial \beta_s}$, $\frac{\partial^2 \pi_i}{\partial \beta_s^2}$, and $\frac{\partial^2 \pi_i}{\partial \beta_t \partial \beta_s}$ differ under the log-log, complementary log-log, logistic, double exponential, and probit link functions, as shown below:

(1) Log-log link function

$$\frac{\partial \pi_i}{\partial \beta_s} = -e^{-\mathbf{x}_i^T \boldsymbol{\beta}} e^{\mathbf{x}_i^T \boldsymbol{\beta}} x_{is}, \quad \frac{\partial^2 \pi_i}{\partial \beta_s^2} = x_{is}^2 (e^{2\mathbf{x}_i^T \boldsymbol{\beta}} - e^{\mathbf{x}_i^T \boldsymbol{\beta}}) e^{-\mathbf{x}_i^T \boldsymbol{\beta}}, \quad \frac{\partial^2 \pi_i}{\partial \beta_t \partial \beta_s} = x_{it} x_{is} (e^{2\mathbf{x}_i^T \boldsymbol{\beta}} - e^{\mathbf{x}_i^T \boldsymbol{\beta}}) e^{-\mathbf{x}_i^T \boldsymbol{\beta}}.$$

(2) Complementary log-log link function

$$\frac{\partial \pi_i}{\partial \beta_s} = e^{-\mathbf{x}_i^T \boldsymbol{\beta}} e^{\mathbf{x}_i^T \boldsymbol{\beta}} x_{is}, \quad \frac{\partial^2 \pi_i}{\partial \beta_s^2} = x_{is}^2 (e^{\mathbf{x}_i^T \boldsymbol{\beta}} - e^{2\mathbf{x}_i^T \boldsymbol{\beta}}) e^{-\mathbf{x}_i^T \boldsymbol{\beta}}, \quad \frac{\partial^2 \pi_i}{\partial \beta_t \partial \beta_s} = x_{it} x_{is} (e^{\mathbf{x}_i^T \boldsymbol{\beta}} - e^{2\mathbf{x}_i^T \boldsymbol{\beta}}) e^{-\mathbf{x}_i^T \boldsymbol{\beta}}.$$

(3) Logistic link function

$$\frac{\partial \pi_i}{\partial \beta_s} = \frac{x_{is} e^{-\mathbf{x}_i^T \boldsymbol{\beta}}}{(1 + e^{-\mathbf{x}_i^T \boldsymbol{\beta}})^2}, \quad \frac{\partial^2 \pi_i}{\partial \beta_s^2} = \frac{2x_{is}^2 e^{-2\mathbf{x}_i^T \boldsymbol{\beta}}}{(e^{-\mathbf{x}_i^T \boldsymbol{\beta}} + 1)^3} - \frac{x_{is}^2 e^{-\mathbf{x}_i^T \boldsymbol{\beta}}}{(e^{-\mathbf{x}_i^T \boldsymbol{\beta}} + 1)^2}, \quad \frac{\partial^2 \pi_i}{\partial \beta_t \partial \beta_s} = \frac{2x_{it} x_{is} e^{-2\mathbf{x}_i^T \boldsymbol{\beta}}}{(e^{-\mathbf{x}_i^T \boldsymbol{\beta}} + 1)^3} - \frac{x_{it} x_{is} e^{-\mathbf{x}_i^T \boldsymbol{\beta}}}{(e^{-\mathbf{x}_i^T \boldsymbol{\beta}} + 1)^2}.$$

(4) Double exponential link function

$$\frac{\partial \pi_i}{\partial \beta_s} = \frac{1}{2} e^{-\|\mathbf{x}_i^T \boldsymbol{\beta}\|} x_{is}, \quad \frac{\partial^2 \pi_i}{\partial \beta_t \partial \beta_s} = \begin{cases} -\frac{x_{it} x_{is} e^{-\mathbf{x}_i^T \boldsymbol{\beta}}}{2}, & \text{if } \mathbf{x}_i^T \boldsymbol{\beta} \geq 0, \\ \frac{x_{it} x_{is} e^{\mathbf{x}_i^T \boldsymbol{\beta}}}{2}, & \text{otherwise,} \end{cases} \quad \frac{\partial^2 \pi_i}{\partial \beta_s^2} = \begin{cases} -\frac{x_{is}^2 e^{-\mathbf{x}_i^T \boldsymbol{\beta}}}{2}, & \text{if } \mathbf{x}_i^T \boldsymbol{\beta} \geq 0, \\ \frac{x_{is}^2 e^{\mathbf{x}_i^T \boldsymbol{\beta}}}{2}, & \text{otherwise.} \end{cases}$$

(5) Probit link function

$$\frac{\partial \pi_i}{\partial \beta_s} = \phi(\mathbf{x}_i^T \boldsymbol{\beta}) x_{is}, \quad \frac{\partial^2 \pi_i}{\partial \beta_s^2} = -\phi(\mathbf{x}_i^T \boldsymbol{\beta}) \mathbf{x}_i^T \boldsymbol{\beta} x_{is}^2, \quad \frac{\partial^2 \pi_i}{\partial \beta_t \partial \beta_s} = -\phi(\mathbf{x}_i^T \boldsymbol{\beta}) \mathbf{x}_i^T \boldsymbol{\beta} x_{it} x_{is},$$

where $s \neq t$. Based on these expectations:

$$\begin{aligned} E[I(Y_i = 0)] &= p_{i0} = (1 - \pi_i)(\rho\pi_i - \pi_i + 1), \\ E[I(Y_i = 1)] &= p_{i1} = 2\pi_i(1 - \rho)(1 - \pi_i), \\ E[I(Y_i = 2)] &= p_{i2} = \pi_i^2 + \rho\pi_i(1 - \pi_i), \quad i = 1, \dots, n, \end{aligned}$$

we can obtain the Fisher information matrix (A2).

References

- Rosner, B. Statistical methods in ophthalmology: An adjustment for the intraclass correlation between eyes. *Biometrics* **1982**, *38*, 105–114. [\[CrossRef\]](#) [\[PubMed\]](#)
- Dallal, G.E. Paired Bernoulli trials. *Biometrics* **1988**, *44*, 253–257. [\[CrossRef\]](#) [\[PubMed\]](#)
- Donner, A. Statistical methods in ophthalmology: An adjusted chi-square approach. *Biometrics* **1989**, *45*, 605–661. [\[CrossRef\]](#) [\[PubMed\]](#)
- Thompson, J.R. The χ^2 test for data collected on eyes. *Br. J. Ophthalmol.* **1993**, *77*, 115–142. [\[CrossRef\]](#)
- Li, X.J.; Li, S.; Tian, G.L.; Shi, J. Modeling paired binary data by a new bivariate Bernoulli model with flexible beta kernel correlation. *TEST* **2024**, *33*, 1180–1224. [\[CrossRef\]](#)
- Tian, W.; Ma, C. Testing homogeneity of proportion ratios for stratified bilateral correlated data. *Math. Comput. Appl.* **2024**, *29*, 26. [\[CrossRef\]](#)
- Li, Y.; Li, Z.; Mou, K. Homogeneity test of many-to-one relative risk ratios in unilateral and bilateral data with multiple groups. *Axioms* **2023**, *12*, 333. [\[CrossRef\]](#)
- Li, Z.; Ma, C.; Mou, K. Testing the common risk difference of proportions for stratified uni-and bilateral correlated data. *Stat. Neerl.* **2023**, *77*, 340–364. [\[CrossRef\]](#)
- Sun, S.; Li, Z.; Mou, K. Interval estimation of common risk difference for stratified unilateral and bilateral data. *J. Biopharm. Stat.* **2025**, *35*, 85–105. [\[CrossRef\]](#)
- Wang, K.; Ma, C.X. Interval estimation of relative risks for combined unilateral and bilateral correlated data. *J. Biopharm. Stat.* **2025**, *35*, 163–186. [\[CrossRef\]](#)
- Zhang, X.; Ma, C. Testing the homogeneity of differences between two proportions for stratified bilateral and unilateral data across strata. *Mathematics* **2023**, *11*, 4156. [\[CrossRef\]](#)

12. Hua, S.; Ma, C. Common odds ratio test and interval estimation for stratified bilateral and unilateral data. *Stat. Methods Med. Res.* **2024**, *33*, 1559–1576. [[CrossRef](#)] [[PubMed](#)]
13. Sun, S.; Li, Z.; Jiang, H. Homogeneity test and sample size of risk difference for stratified unilateral and bilateral data. *Commun. Stat.-Simul. Comput.* **2024**, *53*, 4209–4232. [[CrossRef](#)]
14. Geng, S.; Zhang, L. Decorrelated empirical likelihood for generalized linear models with high-dimensional longitudinal data. *Stat. Probab. Lett.* **2024**, *211*, 110135. [[CrossRef](#)]
15. Chen, X.; Tan, X.; Yan, L. Penalized empirical likelihood for high-dimensional generalized linear models with longitudinal data. *J. Stat. Comput. Simul.* **2023**, *93*, 1515–1531. [[CrossRef](#)]
16. Jiang, J.; Shang, J. Feature screening for high-dimensional variable selection in generalized linear models. *Entropy* **2023**, *25*, 851. [[CrossRef](#)]
17. Cherifi, M.; El Korso, M.N.; Fortunati, S.; Mesloub, A.; Ferro-Famil, L. Robust inference with incompleteness for logistic regression model. *Signal Processing* **2025**, *236*, 110027. [[CrossRef](#)]
18. Shin, B.; Lee, S. Robust logistic regression with shift parameter estimation. *J. Stat. Comput. Simul.* **2023**, *93*, 2625–2641. [[CrossRef](#)]
19. Mou, K.; Li, Z.; Cheng, J. Parameter estimation and hypothesis tests in logistic model for complex correlated data. *Stat. Probab. Lett.* **2024**, *217*, 110294. [[CrossRef](#)]
20. Vasconcelos, J.C.S.; Cordeiro, G.M.; Ortega, E.M.M.; Silva, G.O. A random effect regression based on the odd log-logistic generalized inverse Gaussian distribution. *J. Appl. Stat.* **2023**, *50*, 1199–1214. [[CrossRef](#)]
21. de Freitas, J.V.B.; Nobre, J.S.; Espinheira, P.L.; Rêgo, L.C. Unit gamma regression models for correlated bounded data. *Braz. J. Probab. Stat.* **2023**, *37*, 693–719. [[CrossRef](#)]
22. Zhang, Z.; Arellano-Valle, R.B.; Genton, M.G.; Huser, R. Tractable Bayes of skew-elliptical link models for correlated binary data. *Biometrics* **2023**, *79*, 1788–1800. [[CrossRef](#)]
23. Lin, Y.Q.; Zhang, Y.S.; Tian, G.L.; Ma, C.X. Fast QLB algorithm and hypothesis tests in logistic model for ophthalmologic bilateral correlated data. *J. Biopharm. Stat.* **2021**, *31*, 91–107. [[CrossRef](#)]
24. Devidas, M.; George, E.O. Low-dose extrapolation using the power family of response functions. *Comput. Stat. Data Anal.* **2001**, *36*, 311–317. [[CrossRef](#)]
25. Novianti, P.; Rosadi, D. Copula-based Markov chain logistic regression modeling on binomial time series data. *MethodsX* **2024**, *12*, 102509. [[CrossRef](#)]
26. Gu, J.; Kong, X.; Guo, J.; Qi, H.; Wang, Z. Parameter estimation of three-parameter Weibull distribution by hybrid gray genetic algorithm with modified maximum likelihood method with small samples. *J. Mech. Sci. Technol.* **2024**, *38*, 5363–5379. [[CrossRef](#)]
27. Böhning, D.; Lindsay, B.G. Monotonicity of quadratic-approximation algorithms. *Ann. Inst. Stat. Math.* **1988**, *40*, 641–663. [[CrossRef](#)]
28. Tian, G.L.; Huang, X.F.; Xu, J. An assembly and decomposition approach for constructing separable minorizing functions in a class of MM algorithms. *Stat. Sin.* **2019**, *29*, 961–982. [[CrossRef](#)]
29. Devidas, M.; George, E.O. Monotonic algorithms for maximum likelihood estimation in generalized linear models. *Sankhyā Indian J. Stat. Ser.* **1999**, *61*, 382–396.
30. Ly, A.; Marsman, M.; Verhagen, J.; Grasman, R.P.; Wagenmakers, E.J. A tutorial on Fisher information. *J. Math. Psychol.* **2017**, *80*, 40–55. [[CrossRef](#)]
31. Rajavi, Z.; Katibeh, M.; Ziaei, H.; Fardesmaeilpour, N.; Sehat, M.; Ahmadi, H.; Javadi, M.A. Rapid assessment of avoidable blindness in Iran. *Ophthalmology* **2011**, *118*, 1812–1818. [[CrossRef](#)] [[PubMed](#)]
32. Ma, C.X.; Liu, S. Testing equality of proportions for correlated binary data in ophthalmologic studies. *J. Biopharm. Stat.* **2017**, *27*, 611–619. [[CrossRef](#)] [[PubMed](#)]
33. De Keyser, S.; Gijbels, I. High-dimensional copula-based Wasserstein dependence. *Comput. Stat. Data Anal.* **2025**, *204*, 108096. [[CrossRef](#)]

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