

2×2 交叉设计中错误指定工作相关结构矩阵对样本量估计值的影响

张佩瑜, 谢子恒, 庄 严

南方医科大学公共卫生学院生物统计学系, 广东 广州 510515

摘要:目的 在2×2交叉设计中基于广义估计模型,探究错误指定工作相关结构矩阵对样本量估计值的影响。方法 基于蒙特卡罗模拟,控制模拟时总样本量大小 n ,分配到AB序列($s=1$)的受试者比例 θ ,相关系数 ρ 和安慰剂效果OR四个因素后,探究不同条件下错误指定工作相关结构矩阵对样本量估计值的影响。评价样本量估计值和理论值之间差异的指标采用偏差和均方误差。结果 当正确工作相关结构矩阵为独立时,正确指定工作相关结构矩阵的样本量估计效果优于错误指定工作相关结构矩阵的样本量估计效果。但是当正确工作相关结构矩阵为等相关时,相关系数越接近0的前提下,其他因素越小(本文中 $n \leq 50, \theta \leq 0.5, OR = 2$),出现了正确指定工作相关结构矩阵后样本量估计值对理论值的bias大于错误指定工作相关结构矩阵的bias的情况。结论 大多数条件下错误指定工作相关结构矩阵会使样本量估计值远偏离理论值,但是一定条件下错误指定工作相关结构矩阵对样本量估计值的影响较小。

关键词:2×2交叉试验;广义估计设计;工作相关结构矩阵;样本量估计

Impact of incorrect designation of working correlation structure matrix on sample size estimation in 2×2 cross design: a simulation study

ZHANG Peiyu, XIE Ziheng, ZHUANG Yan

Department of Biostatistics, School of Public Health, Southern Medical University, Guangzhou 510515, China

Abstract: Objective To investigate the impact of incorrect specification of the working correlation structure matrix on estimated sample size in a 2×2 crossover design based on the generalized estimating equation (GEE). **Methods** Based on Monte Carlo simulation, the influence of incorrect specification of the work-related structure matrix on the sample size estimation under different conditions was evaluated after controlling the total sample size n , the proportion of subjects assigned to AB sequence ($s=1$) θ , the correlation coefficient ρ , and the placebo effect OR. Bias and mean square error (MSE) were used to assess the difference between the sample size estimates and the theoretical values. **Results** When the correctly specified working correlation structure matrix is independent, the sample size estimation effect of correctly specifying the working correlation structure matrix is better than that of incorrect specification. But when the correctly specified working correlation structure matrix is equal and the correlation coefficient is closer to 0, with other factors being smaller ($n \leq 50, \theta \leq 0.5, OR=2$ in this article), there is a situation where the bias of the sample size estimation value for the correctly specified working correlation structure matrix is greater than the bias for the incorrectly specified working correlation structure matrix. **Conclusion** Under most conditions, incorrectly specifying the working correlation structure matrix can cause the estimated sample size to deviate significantly from the theoretical value, but under certain conditions, the impact of incorrectly specifying the working correlation structure matrix can be small on the estimated sample size.

Keywords: 2×2 crossover trials; generalized estimation equation design; working correlation structure matrix; sample size estimation

在临床研究设计中,交叉设计是一种独特且高效的研究方法。其中,简单的AB/BA或2×2交叉试验被广泛应用于对症药物的开发,特别是在缓解癫痫、哮喘和遗传性血管性水肿(HAE)等不可治愈的慢性疾病的症状研究中。在进行交叉试验时,受试者经过第1次治疗后要经历一段洗脱期,然后再进行第2次治疗,这样受试者以自己作为对照组,大大减少了治疗比较中受试者之间混杂因素的影响^[1-3]。由于交叉试验在对结果感兴趣

的患者差异很大时,相对平行试验需要更少的患者,所以引起了研究者的兴趣,但同时也引出样本量计算的问题。

在实际研究活动中,样本的收集往往是一项会耗费大量人力、物力和财力的工作,同时样本量的大小也会影响研究的结果。为确保研究结果的有效性并节省资源,计算合适的样本量至关重要,已经有文章提出关于2×2交叉设计的样本量计算^[4-6]。然而,交叉设计的样本量计算还需考虑序列间和序列内的相关性,如果忽略了相关性可能会导致样本量大小异常,从而影响研究结果的准确性。传统的样本量计算方法通常假设每个受试者的效应是恒定的,并且在传统回归模型中,受试者的个体效应会被消除,为了解决这一问题,Liang^[7]提出了广义估计方程(GEE)。由于GEE不需要对数据的分布

收稿日期:2024-12-12

基金项目:国家自然科学基金(81773544)

Supported by National Natural Science Foundation of China (81773544).

作者简介:张佩瑜,在读硕士研究生,E-mail: 952398229@qq.com;

谢子恒,在读本科生,E-mail: 2747128090@qq.com

通信作者:庄 严,副教授,博士,E-mail: zhuangy179@126.com

做出严格假设,而是通过工作相关矩阵来描述数据之间的相关性结构,可以用来灵活处理各种离散数据和连续数据,并且相较于广义线性模型(GLM),GEE可以更好地处理响应变量集群内部具有相关性的纵向数据^[8-11]与非正态重复测量数据^[7, 12-14]。

使用GEE模型进行样本量计算已经应用于平行设计^[15-18]、聚类随机试验^[19]、聚类随机交叉试验^[20]、阶梯形聚类随机试验^[21, 22]和前后研究^[23]中。2021年,Tang^[24]提出了用GEE模型对2×2交叉设计的样本量进行计算。然而,实际应用中,研究者常无法准确识别变量间的相关关系(如独立、等相关或其他复杂结构),可能导致工作相关矩阵的错误指定,进而影响样本量估计^[25]。因此,模拟分析错误指定相关关系对样本量计算的影响具有重要现实意义。

本文主要基于GEE模型,探讨错误指定工作相关结构矩阵对样本量计算的影响,并通过模拟研究验证其偏差情况。本研究关注的结局数据类型为二分类结局,考虑的相关矩阵结构为等相关和独立结构。通过系统分析,本研究者提供更准确的样本量计算方法,并为复杂研究设计中的统计推断提供理论支持。

1 理论基础

在2×2交叉设计中受试者被随机分配到AB或BA序列,其中A和B代表被比较的两种治疗(安慰剂和试验药物)。受试者在第1期接受治疗A,在第2期接受治疗B,按顺序为AB,反之则为BA。洗脱期通常在两种治疗方法之间插入,它应该足够长以消除传递效应。在本文中,我们假设没有传递效应^[1-3]。

二项结果的GEE方法,其中在第j周期中的受试者i的结果 y_{ij} 是一个标量。在GLM中线性预测函数可以写成以下形式:

$$\eta_{ij} = x_{ij}\beta = \beta_0 + \beta_p x_{ij1} + \beta_g x_{ij2} \quad (1)$$

其中 $x_{ij} = (1, x_{ij1}, x_{ij2})'$, $\beta = (\beta_0, \beta_p, \beta_g)'$, 如果 $j = 2, 0$ 那么 $x_{ij1} = 1$,反之为0;如果接受治疗A那么 $x_{ij2} = 1$,反之为0。 y_{ij} 的工作边际方差可以具体写为 $\phi A_{ij} = \phi v(\mu_{ij})$ 。令 $y_i = (y_{i1}, \dots, y_{in})'$ 为第i位受试者能观察到的结局, y_i 的均值为 $\mu_i = (\mu_{i1}, \dots, \mu_{in})$,工作协变量矩阵为 $V_{wi} =$

$\phi A_{wi}^{1/2} R_{wi} A_{wi}^{1/2}$,其中 $A_{wi} = \text{diag}(v(\mu_{i1}), \dots, v(\mu_{in}))$, R_{wi} 是工作相关结构矩阵。这里令 $n_i = 2$,那么 $R_{wi} = \begin{bmatrix} 1 & \rho_w \\ \rho_w & 1 \end{bmatrix}$;只有当 $n_i = 1$ 时 $R_{wi} = 1$ 。如果假设一个独立的工作相关性,则 $\rho_w = 0$ 。

令 $B_i = \text{diag}(\partial u_{i1}/\partial \eta_{i1}, \dots, \partial u_{in}/\partial \eta_{in})$, $E_i = x_i'(B_i A_{wi}^{-1/2})$, $F_i = A_{wi}^{-1/2} A_{wi}^{1/2}$ 。对于logit连接 $\partial u_{ij}/\partial \eta_{ij} = \mu_{ij}(1 - \mu_{ij})$ 。 I_0, I_1 计算如下:

$$I_0 = \sum_{i=1}^m w_i E_i R_{wi}^{-1} E_i' \text{ and } I_1 = \sum_{i=1}^m w_i E_i R_{wi}^{-1} F_i R_{wi} F_i R_{wi}^{-1} E_i' \quad (2)$$

GEE回归参数估计的渐近分布:

$$\sqrt{n}[\hat{\beta} - \beta] \sim N(0, \Sigma_\beta) \text{ and } \sqrt{n}(\hat{\beta}_g - \beta_g) \sim N(0, \sigma_g^2) \quad (3)$$

其中 σ_g^2 是 $\Sigma_\beta = I_0^{-1} I_1 I_0^{-1}$ 中的第(p,p)位元素(右下角元素)。

在优效性试验中,目的是证明实验治疗比对照治疗更有效,假设 β_g 越小表示健康状况越好。假设可以写为以下形式:

$$H_0: \beta_g \geq 0 \text{ vs } H_1: \beta_g < 0$$

在Wald Z检验下的样本量可以由以下公式给出

$$n_z = \frac{(z_{1-\alpha/2} - z_{1-\beta})^2 \sigma_g^2}{\beta_g^2} \quad (4)$$

自由度为 $f = n - p$ 的Wald t检验 样本量可以由以下公式给出^[26]

$$n_t = n_z + \frac{z_{1-\alpha/2}^2}{2} \text{ or } n_{t2} = n_t + \frac{z_{1-\alpha/2}^4}{4n_t} \quad (5)$$

假设二元结局 y_{ij} ($y_{ij} = 1$ or 0)的真实分布服从均值为 μ_{ij} 以及方差为 $\mu_{ij}(1 - \mu_{ij})$,并且logit连接为 $\mu_{ij} = \frac{\exp(\eta_{ij})}{1 + \exp(\eta_{ij})}$ 。真实边际方差为 $A_{wi} = \text{diag}[\mu_{i1}(1 - \mu_{i1}), \dots, \mu_{in}(1 - \mu_{in})]$ 。假设所有受试者之间的真实相关性恒为 ρ_t ,使用OR作为估计值^[27, 12]。表1定义了二元结果的协变量搭配和权重。令 μ_{sj} 为受试者 y_{ij} 在周期j中第s序列下的均值。

将以下 ϕ 和 ρ_w 分别代入公式(2),即可得到样本量公式。

$$\phi = \begin{cases} 1 & \text{binomial} \\ \frac{(w_{11} + w_{13})(1 - \mu_{11}) + (w_{12} + w_{13})(1 - \mu_{12}) + (w_{21} + w_{23})(1 - \mu_{21}) + (w_{22} + w_{23})(1 - \mu_{22})}{w_{11} + w_{12} + 2w_{13} + w_{21} + w_{22} + 2w_{23}} & \text{Poisson} \end{cases}$$

$$\rho_w = \begin{cases} 0 & \text{independent correlation structure} \\ \rho_t & \text{dependent structure, binomial} \\ \frac{w_{13}\sqrt{(1 - \mu_{11})(1 - \mu_{12})} + w_{23}\sqrt{(1 - \mu_{21})(1 - \mu_{22})}}{(w_{13} + w_{23})\phi} \rho_t & \text{dependent structure, Poisson} \end{cases}$$

表1 每个时期的单个时间点收集结果的交叉试验中的权重和协变量搭配

Tab.1 Weights and covariate configurations in crossover trials with outcomes collected at a single time point in each period

| Item | Sequence AB (s=1) | | | Sequence BA (s=2) | | |
|---|---|---|---|---|---|---|
| | Period 1 | Period 2 | Both period | Period 1 | Period 2 | Both period |
| Weights ^a | $w_{11} = \theta_1 \pi_1^{(1)}$ | $w_{12} = \theta_1 \pi_2^{(1)}$ | $w_{13} = \theta_1 \pi_{12}^{(1)}$ | $w_{21} = \theta_2 \pi_1^{(2)}$ | $w_{22} = \theta_2 \pi_2^{(2)}$ | $w_{23} = \theta_2 \pi_{12}^{(2)}$ |
| Covariates for binary | | | | | | |
| x_{r1} | [1 0 1] | - | [1 0 1] | [1 0 0] | - | [1 0 0] |
| x_{r2} | - | [1 1 0] | [1 1 0] | - | [1 1 1] | [1 1 1] |
| Covariates for ordinal outcome with L levels ^b | | | | | | |
| $\begin{bmatrix} x_{i1_1} \\ \dots \\ x_{i1_M} \end{bmatrix}$ | $\begin{bmatrix} x_{ij_1}^* & 0 & 1 \\ \dots \\ x_{ij_M}^* & 0 & 1 \end{bmatrix}$ | - | $\begin{bmatrix} x_{ij_1}^* & 0 & 1 \\ \dots \\ x_{ij_M}^* & 0 & 1 \end{bmatrix}$ | $\begin{bmatrix} x_{ij_1}^* & 0 & 0 \\ \dots \\ x_{ij_M}^* & 0 & 0 \end{bmatrix}$ | - | $\begin{bmatrix} x_{ij_1}^* & 0 & 0 \\ \dots \\ x_{ij_M}^* & 0 & 0 \end{bmatrix}$ |
| $\begin{bmatrix} x_{i2_1} \\ \dots \\ x_{i2_M} \end{bmatrix}$ | - | $\begin{bmatrix} x_{ij_1}^* & 1 & 0 \\ \dots \\ x_{ij_M}^* & 1 & 0 \end{bmatrix}$ | $\begin{bmatrix} x_{ij_1}^* & 1 & 0 \\ \dots \\ x_{ij_M}^* & 1 & 0 \end{bmatrix}$ | - | $\begin{bmatrix} x_{ij_1}^* & 1 & 1 \\ \dots \\ x_{ij_M}^* & 1 & 1 \end{bmatrix}$ | $\begin{bmatrix} x_{ij_1}^* & 1 & 1 \\ \dots \\ x_{ij_M}^* & 1 & 1 \end{bmatrix}$ |

^a θ_s is proportion of subjects allocated to sequence $s(\theta_1 + \theta_2 = 1)$; $\pi_1^{(s)}, \pi_2^{(s)}, \pi_{12}^{(s)}$ are proportions of subjects observed in period 1, period 2 and both periods in sequence s ; $x_{ij_m}^*$ is a $1 \times M$ vector with 1 in the m -th entry, and 0 in all other entries, where $M = L - 1$.

2 模拟结果

2.1 模拟结果评价指标

为评价理论样本量 m_i 和样本量估计值 m_r 间的差异,同时比较错误指定工作相关结构矩阵后对样本量影响的趋势,这里采用了两个评价指标作为评价标准。

(1) 偏差(bias) 偏差的计算公式如下:

$$bias(m) = \sum_{i=1}^{1000} \frac{m_r - m_i}{1000} \quad (6)$$

当偏差的绝对值越小,就代表样本量估计值越接近理论值,即估计的效果越好,反之代表估计效果很差。

(2) 均方误差(MSE) 均方误差(MSE)的计算公式如下:

$$MSE(m) = \sum_{i=1}^{1000} \frac{(m_r - m_i)^2}{1000} \quad (7)$$

当均方误差越小时,代表1000次模拟下 m_r 和 m_i 之间的累计误差值越小,即估计值的波动性越小,反之代表估计值的波动性越大。

2.2 模拟设置

有关模拟的设置,分析的是有关交叉设计二项数据的基于GEE模型在错误指定工作相关结构矩阵后对样本量估计值大小的影响,其中通过GEE模型得到关于组内差异的二项系数的方差 σ_g ,对样本量进行模拟并和理论值进行比较。

首先,需要模拟2000个数据集作为后续计算样本量的基础,模拟数据的数据集结构如表2所示。由于交叉设计的特殊性,模拟中假设安慰剂治疗第一期和第二期的发病率 P_1 和 P_2 服从 $U(0.3, 0.6)$ 的均匀分布,常数项 β_0 及一次项系数 β_p 由 P_1 和 P_2 决定,分别为 $\beta_0 =$

$\log(P_1), \beta_p = \log(P_1/P_2)$,二次项系数由OR决定 $\beta_g = \log(OR)$,连接函数选择logit函数,而且假设符合 $logit(p_{ij}) = \beta_0 + \beta_p X_{ij1} + \beta_g X_{ij2}$,其中可以通过R语言中 $inv.logit()$ 函数求有关 $logit(p_{ij})$ 的逆函数得到 p_{ij} 。使用R函数“mvtBinaryEP”的系列函数,通过 $ep()$ 函数基于 $inv.logit()$ 函数求出的 p_{ij} ,给定的相关系数 ρ 以及 $R(\rho)$ 的结构类型来生成数据对应的二元响应变量。

表2 数据结构

Tab.2 Data structure

| Id | Period | Treatment | s | Y |
|----|--------|-----------|---|---|
| 1 | 0 | 1 | 1 | 0 |
| 1 | 1 | 0 | 1 | 0 |
| 2 | 0 | 1 | 1 | 0 |
| 2 | 1 | 0 | 1 | 0 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 49 | 0 | 0 | 2 | 0 |
| 49 | 1 | 1 | 2 | 0 |
| 50 | 0 | 0 | 2 | 0 |
| 50 | 1 | 1 | 2 | 0 |

在生成模拟数据集后,根据模拟生成的数据求出理论的OR值以及 P_1 和 P_2 值,然后使用R包“geepack”中的 $geeglm()$ 函数分析模拟生成的数据,通过GEE模型计算得到估计的OR以及 P_1 和 P_2 值。分别将两组值代入根据文章后给出的SAS程序改写的R代码的 $geesize_bin()$ 函数来计算样本量大小,分别记作样本量理

论值 m_i 和样本量估计值 m_r , 并且将 m_i 和 m_r 进行比较。

本次模拟主要研究的因素:(1)模拟时总样本量的大小 n ; (2)分配到AB序列 ($s = 1$) 的受试者比例 θ ; (3)相关系数 ρ , 并且假设工作相关矩阵 $R(\rho)$ 有等相关结构和独立相关结构; (4)安慰剂效果 OR 。

2.3 模拟结果

2.3.1 正确工作相关结构矩阵为等相关 工作相关结构矩阵选择为等相关, 使用广义估计方程拟合数据得到中间参数和使用R程序计算样本量时分别为等相关、独立。在本次模拟研究中正确工作相关结构矩阵为等相关结构, 错误指定工作相关结构矩阵为独立

在本次模拟生成数据时, 关于模拟时的总样本量为 n , 令 $n = 20, 50, 100, 200$ 四种情况; 关于分配到AB序列的受试者比例为 θ , 令 $\theta = 0.3, 0.5, 0.6$ 三种情况; 关于相关系数为 ρ , 令 $\rho = 0.3, 0.5, 0.8$; 关于安慰剂效果为 OR , 令 $OR = 2, 3, 4, 5$ 四种情况; 关于工作矩阵是等相关矩阵和独立矩阵的对比。已知bias和MSE越小代表样本量理论值和样本量估计值差别越小, 即样本量估计效果越好。由于在模拟时发现多数情况下MSE的值都

比较大, 为了方便比较, 这里有关MSE的变化趋势图的纵坐标用 $\log(MSE)$ 来表示。

模拟时的总样本量 n 和分配到AB序列的受试者比例 θ 固定时, 样本量理论值在正确指定工作相关结构矩阵和错误指定工作相关结构矩阵两种情况下不变。观察图1和图2可以发现错误指定工作相关结构矩阵为独立相关后, 样本量估计值 m_r 和样本量理论值 m_i 之间的bias和MSE总体大于正确指定工作相关矩阵的bias和MSE, 这也意味着在错误指定工作矩阵后随着相关系数 ρ 的变化, 样本量估计值总体大于正确指定工作相关结构矩阵的样本量估计值。并且随着 ρ 的变化, 正确指定工作相关结构矩阵后样本量间bias和MSE的总体趋势是向下的, 这代表 ρ 越大, 样本量估计值越接近理论样本量, 估计效果越好, 估计效果越稳定。相反, 错误指定工作相关结构矩阵后样本量间bias和MSE的总体趋势是向上的, 这代表 ρ 越大, 样本量估计值越远离理论样本量, 估计效果越差, 估计效果越不稳定。

(1) 固定 n 和 θ 后随着bias和MSE的变化趋势

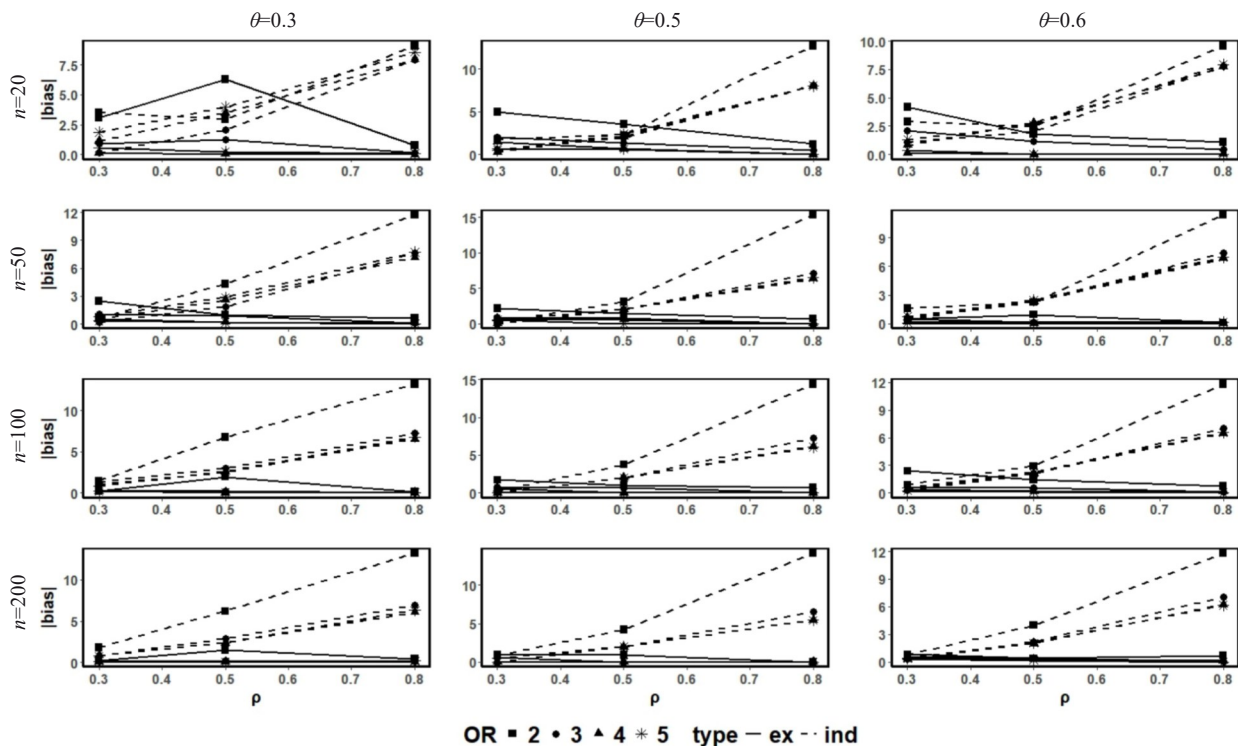


图1 关于分别固定 n 和 θ 后随着 ρ 的变化bias的变化趋势

Fig.1 Changes of bias with ρ after fixing n and θ .

在 $\theta = 0.3, n = 20, \rho = 0.3$ 时正确指定工作相关结构矩阵后的bias和MSE大于错误指定工作相关结构矩阵后的bias和MSE, 同时在多种组合情况下的 $\rho = 0.3, OR = 2$ 时, 也出现了同样的情况, 这说明在参数值偏小时, 更容易出现错误指定工作相关结构矩阵后的估

计效果比正确指定工作相关结构矩阵后的估计效果好的情况, 导致出现这种情况的原因可能有点: 第一, 当样本量 n 以及分配到AB序列的受试者比例 θ 较小时, 模型的估计可能不够稳定, 尤其是对于复杂的工作相关结构(如等相关或自回归结构)。在这种情况下, 错

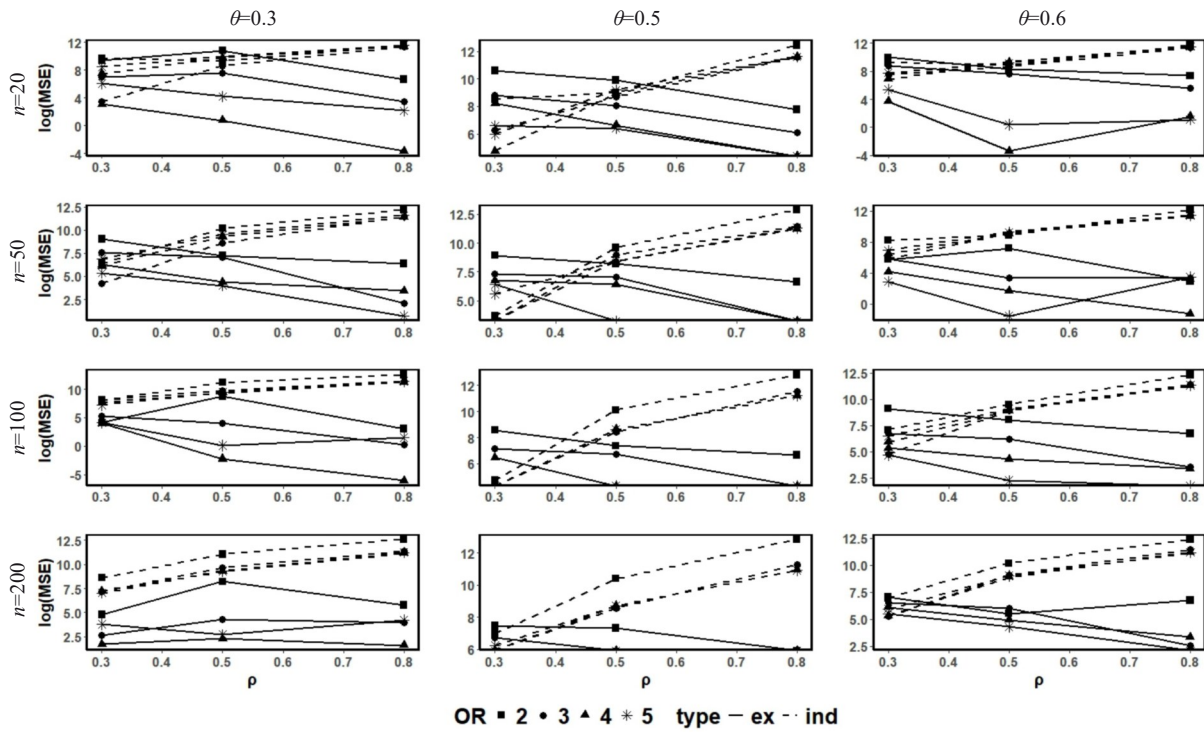


图2 关于分别固定 n 和 θ 后随着 ρ 的变化 $\log(\text{MSE})$ 的变化趋势
 Fig.2 Changes of $\log(\text{MSE})$ with ρ after fixing n and θ .

误指定为独立结构(假设数据之间无相关性)可能会简化模型,减少估计的方差,从而在特定条件下表现出较低的 bias 和 MSE。第二,当相关系数 ρ 较小时,数据之间的相关性较弱,独立结构的假设可能更接近真实情况,正确指定工作相关结构(如等相关结构)可能会过度拟合数据。

(2)固定 ρ 和 OR 后 bias 和 MSE 的变化趋势

模拟时的安慰剂效果 OR 和相关系数 ρ 固定时,样本量理论值在正确指定工作相关结构矩阵和错误指定工作相关结构矩阵两种情况下不变。通过观察图 3 和图 4,随着 θ 的变化,错误指定工作相关结构矩阵时的样本量估计值和理论样本量间的 bias 和 MSE 总体大于正确指定工作相关结构矩阵时的 bias 和 MSE,并且随着 θ 的增大,无论是正确指定工作相关矩阵还是错误指定工作相关矩阵在 $\theta = 0.5$ 处都会出现一个峰值。理论上来说在这个点上计算的样本量是最理想的,因为大多数的研究者都会假设受试者平均分配到两个序列中,所以这可能会导致 θ 对样本量的影响降到最低,其他参数的交互影响达到最大,进而导致 bias 和 MSE 出现最大值或最小值。例如,当 $\theta = 0.5$,在 $\rho = 0.3$ 时,以及 $\rho = 0.5$, $OR = 2$ 且 $n = 20$ 时都出现了错误指定工作相关结构矩阵估计效果优于正确指定工作相关结构矩阵估计效果的情况,而且错误指定工作相关结构矩阵的 bias 和 MSE 为最小值,正确指定工作相关结构矩阵的 bias 和 MSE 为最大值。并且在 $\rho = 0.3$ 时,出现这种情况的频率很高,在不

同的参数组合下基本都会出现这个情况,这也同时说明了在 $\rho = 0.3$ 时正确指定工作相关结构矩阵后样本量估计效果的不稳定性。

2.3.2 正确工作相关结构矩阵为独立 在本次模拟研究中正确工作相关结构矩阵为独立,错误指定工作相关结构矩阵为等相关。在本次模拟中生成模拟数据的具体操作与正确工作相关结构矩阵为等相关时的流程一致,稍有不同的是本次模拟中的正确工作相关结构矩阵为独立,即生成理论样本量时 $\rho = 0$ 。关于错误指定工作相关结构矩阵的情况下的参数设置与上一部分一致。

通过图 5 和图 6 可以观察到随着参数的变化,正确指定工作相关结构矩阵后的样本量估计效果始终优于错误指定工作相关结构矩阵后的样本量估计效果,并且 bias 和 MSE 值都在理想范围内,这说明样本量估计值和理论样本量间的差距很小,估计效果很好。

3 讨论

当正确的工作相关结构矩阵为等相关时,从 bias 和 MSE 的变化趋势来看,随着 OR 和样本量 n 的增大,样本量估计值与理论值间的 bias 和 MSE 逐渐减小,且正确指定与错误指定工作相关矩阵间的 bias 差距也逐渐缩小。然而,随着相关系数 ρ 的增大,正确指定工作相关矩阵后 bias 和 MSE 逐渐减小,而错误指定后 bias 和 MSE 逐渐增大。此外,随着分配到 AB 序列的受试者比例 θ 的增大,错误指定后的 bias 和 MSE 在 $\theta = 0.5$ 时达到最小值,而正确指定后的 bias 和 MSE 在 $\theta = 0.5$ 时达

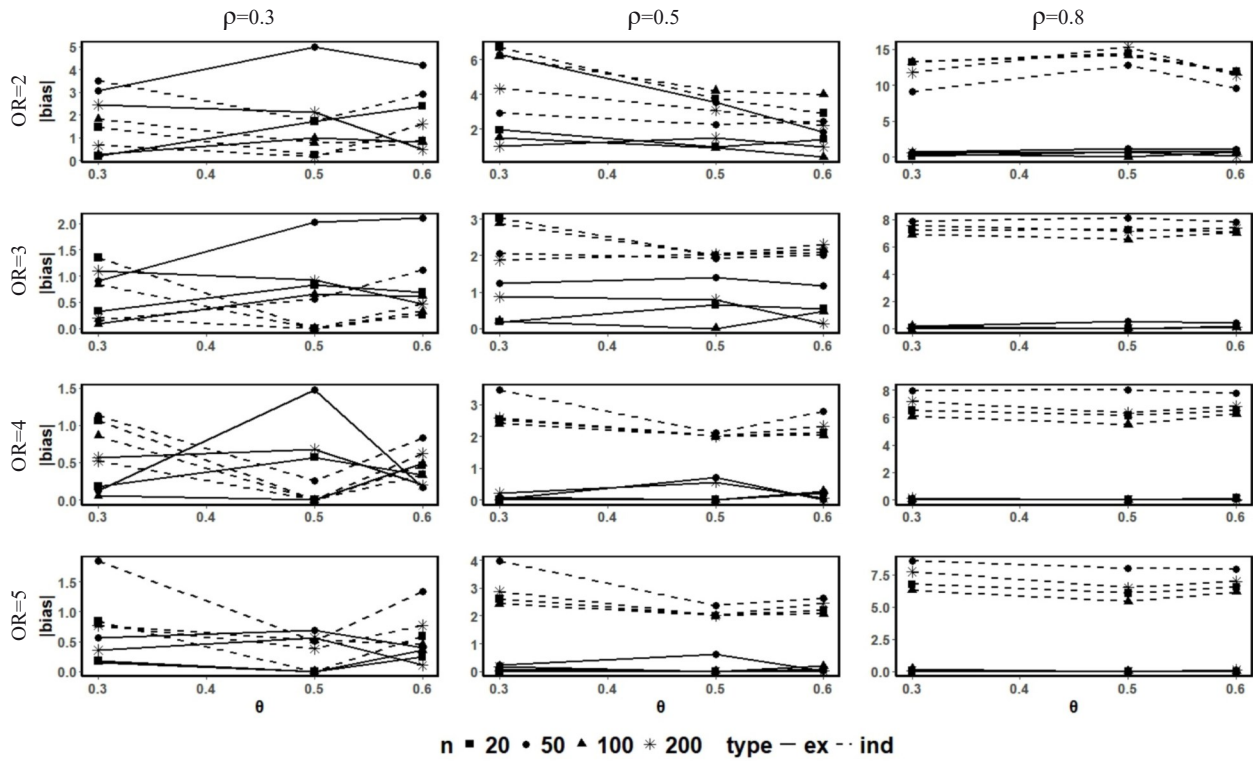


图3 关于分别固定 ρ 和 OR 后随 θ 的变化bias的变化趋势
 Fig.3 Changed of bias with θ after fixing ρ and OR .

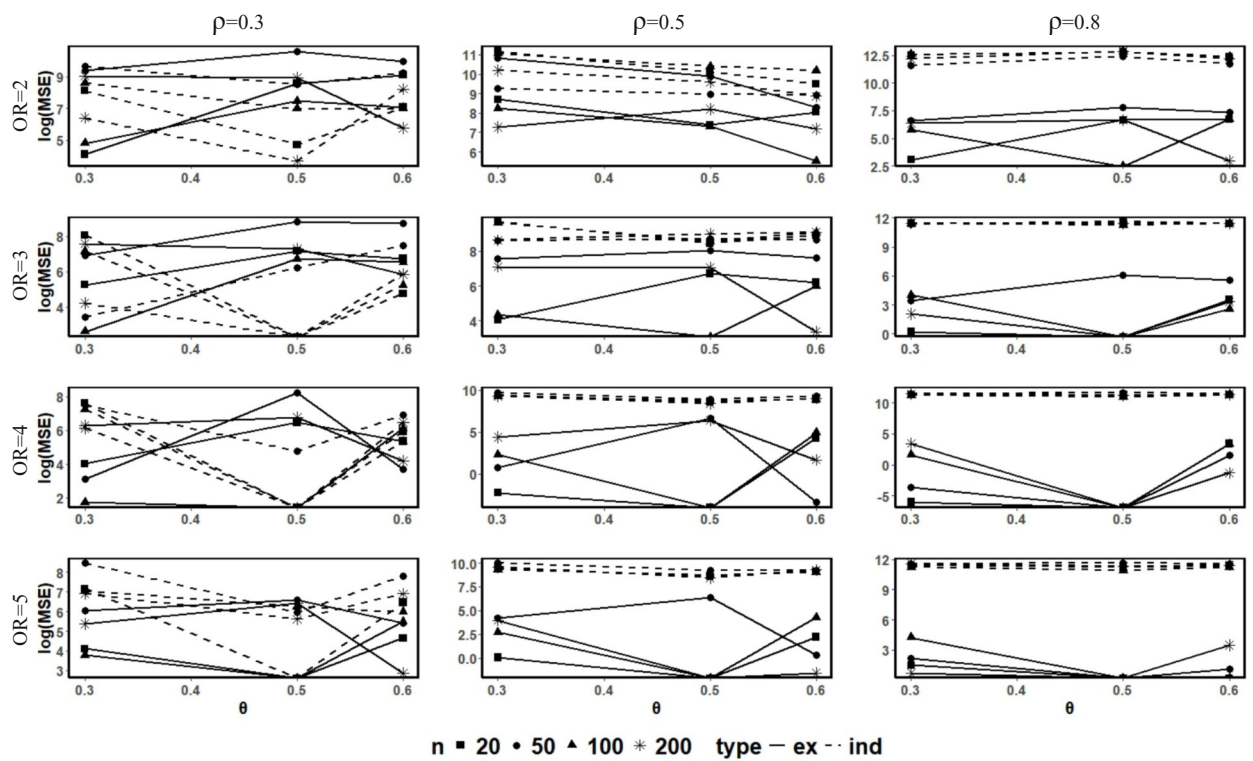


图4 关于分别固定 ρ 和 OR 后随 θ 的变化log(MSE)的变化趋势
 Fig.4 Changes of log(MSE) with θ after fixing ρ and OR .

到最大值。理论上, $\theta = 0.5$ 时样本量计算最理想,因为大多数研究者假设受试者平均分配到两个序列中,这可能导致 θ 对样本量的影响最小化,而其他参数的交互影

响最大化,从而导致bias和MSE出现极值。

从样本量估计效果来看,在大多数条件下,正确指定工作相关矩阵的估计效果优于错误指定。多数情况

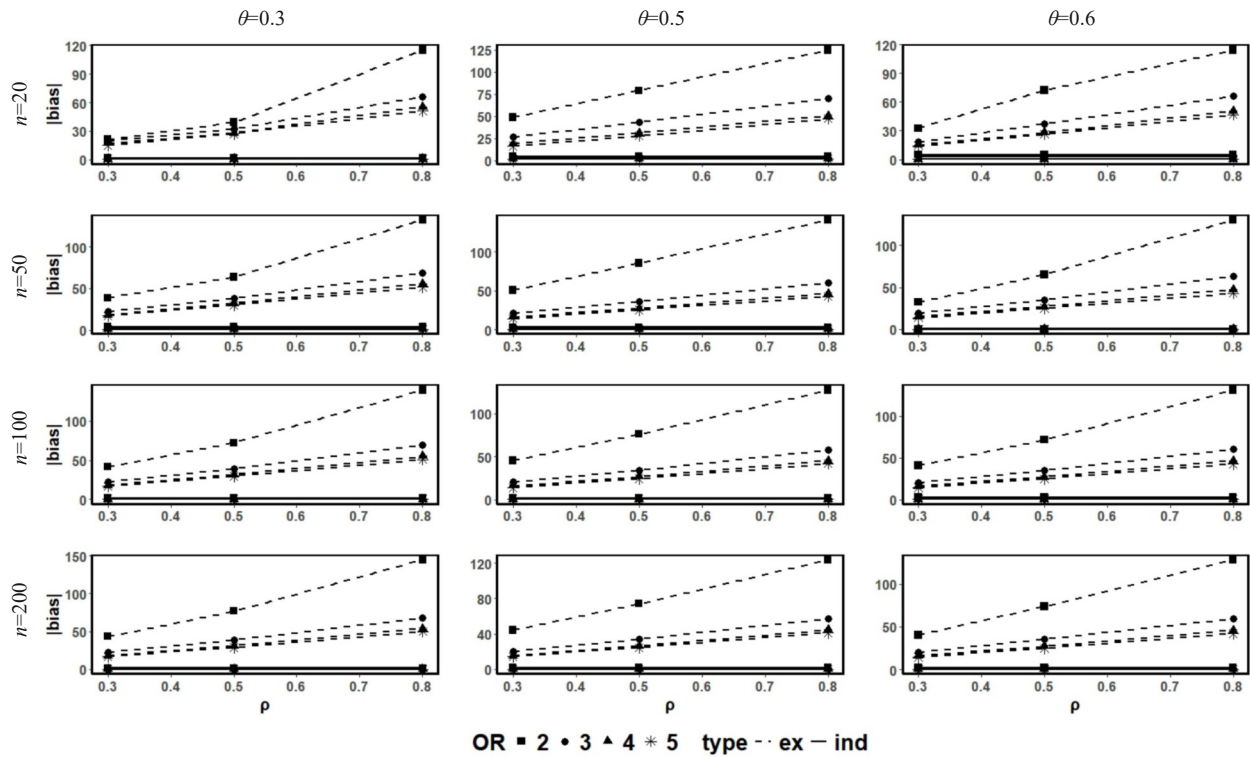


图5 关于分别固定n和θ后随着ρ的变化bias的变化趋势

Fig.5 Change of bias with ρ after fixing n and θ.

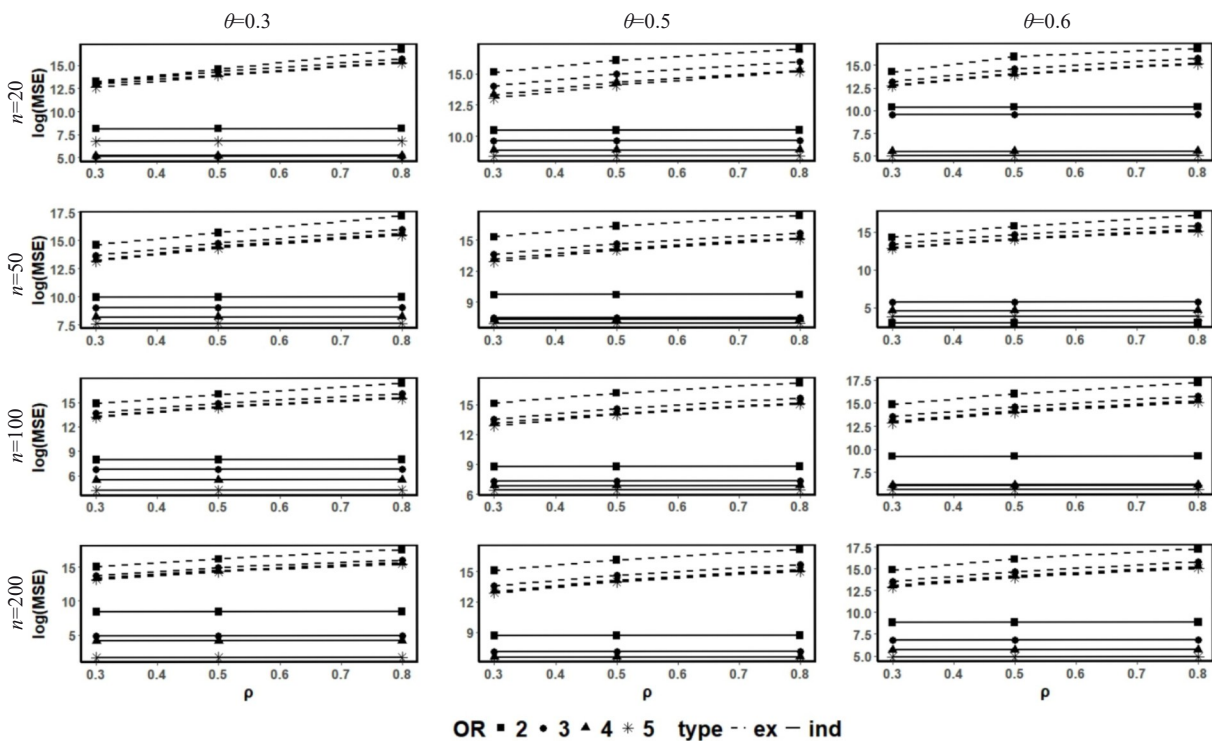


图6 关于分别固定n和θ后随着ρ的变化log(MSE)的变化趋势

Fig.6 Changes of log(MSE) with ρ after fixing n and θ.

下,正确指定后的bias在理想范围内。然而,模拟研究发现,当总样本量 $n \leq 50$ 、 $\theta \leq 0.5$ 、 $\rho \leq 0.5$ 且安慰剂效果 $OR = 2$ 时,错误指定工作相关矩阵的估计效果可能优于正确指定。由于错误指定为独立结构时 $\rho = 0$,当 ρ 较

小时,正确指定与错误指定的估计效果接近,特别是在 ρ 接近0且 $n \leq 50$ 、 $\theta \leq 0.5$ 、 $OR = 2$ 时,错误指定的估计效果可能更优。这可能是因为在小样本和低 θ 条件下,模型估计不够稳定,尤其是对于复杂的工作相关结构(如等

相关或自回归结构)。此时,错误指定为独立结构可能简化模型,减少估计方差,从而在特定条件下表现出较低的bias和MSE。然而,即使在这种情况下,正确指定与错误指定的bias差异较小,平均样本量差异通常在5以内。

本研究虽然探究了错误指定工作相关结构矩阵后对样本量的影响,但是仍存在一些不足之处,需要在未来的研究中进一步改进和拓展。首先,关于样本量修正方法的改进问题,从模拟结果得知错指工作相关矩阵会对样本量产生影响,但是目前通用的几个方法,如:MD^[28]、KC^[30]等,对样本的修正效果都不理想。未来可以参考已有的修正方法,尝试使用更灵活的相关结构或基于模型本身进行校正,以减少错误指定带来的影响^[13, 28, 29]。

其次,模拟研究的不充分,由于在实际研究中相关系数不仅是统计学中的基础指标,更是科学解释和模型构建的核心要素,在运用GEE进行分析时相关系数强弱误设可能会直接影响样本量计算的偏差,从而影响参数估计和推断可靠性,进而导致结论错误甚至资源浪费,所以相关关系的强弱误设模拟同样具有现实意义^[30-32]。但是本研究只考虑到了错误指定工作相关矩阵结构为等相关或独立时对样本量的影响,对于更加复杂的工作相关矩阵结构以及相关关系的强弱误设(如强相关误设为弱相关,或反之)的情况没有进行充分考虑。针对这一问题,未来可以从以下两个方面进行拓展:第一,关于工作相关结构矩阵的类型局限于等相关和独立的问题,由于在2×2交叉设计中只需要确定一个相关系数 ρ ,所以更复杂的工作相关结构矩阵在高阶的交叉设计(如3×3交叉设计)中才具有实际意义^[24]。因此,未来可以将样本量估计方法扩展到3×3交叉设计,并针对复杂相关矩阵结构(如一阶自回归和不确定性)的错误指定进行模拟研究。第二,针对误设相关性强弱对样本量影响的问题,研究下一步将扩展涵盖强相关、弱相关的误设模拟研究,例如模拟正确相关程度为强相关时错误指定为弱相关,或者正确相关程度为弱相关时错误指定为强相关的情况下,样本量出现的偏差。

最后,结局变量的类型考虑不全面,本研究的结局变量仅限于二分类数据,而未考虑连续型或有序型结局变量^[24]。这些类型的变量通常包含更多信息,对工作相关矩阵错误指定的敏感性可能较低,但仍可能导致样本量出现较大偏差。因此,未来研究将继续探讨不同结局变量下GEE模型的表现,尤其是其对相关结构错误指定的稳健性及其对样本量估计精度的影响。

总而言之,通过此次模拟研究,在2×2交叉设计中使用GEE计算样本量时,多数情况下错误指定工作相关结构矩阵会使样本量估计值远偏离理论值,在实际研究中要特别注意相关关系的确定以及参数的设置,以防

样本量计算出现偏差,进而导致最后试验分析结果出现错误。

Declaration of interests: The authors declare no competing interests.

参考文献:

- [1] Hills M, Armitage P. The two-period cross-over clinical trial[J]. Brit J Clinical Pharma, 1979, 8(1): 7-20.
- [2] Lui KJ, Chang KC. Test non-inferiority (and equivalence) based on the odds ratio under a simple crossover trial[J]. Stat Med, 2011, 30(11): 1230-42.
- [3] Lui KJ. Sample size determination for testing equality in Poisson frequency data under an AB/BA crossover trial[J]. Pharm Stat, 2013, 12(2): 74-81.
- [4] Chow SC, Wang H. On sample size calculation in bioequivalence trials[J]. J Pharmacokinet Pharmacodyn, 2001, 28(2): 155-69.
- [5] Chow SC, Cheng B, Cosmatos D. On power and sample size calculation for QT studies with recording replicates at given time point[J]. J Biopharm Stat, 2008, 18(3): 483-93.
- [6] Siyasinghe NM, Sooriyarachchi MR. Guidelines for calculating sample size in 2x2 crossover trials: a simulation study[J]. J Natl Sci Found Sri Lanka, 2011, 39(1): 77.
- [7] Liang KY, Zeger SL. Longitudinal data-analysis using generalized linear-models[J]. Biometrika, 1986, 73(1): 13-22.
- [8] 高燕宁, 蔡文玮, 周纪芩. 广义估计方程GEE1与纵向资料的回归分析[J]. 数理医药学杂志, 1994, 7(2): 118-23.
- [9] 高燕宁, 刘欣华, 蔡文玮, 等. 广义估计方程GEE2与GEE1的比较[J]. 数理医药学杂志, 1996, 9(4): 321-2.
- [10] 夏彦, 潘晓平, 刘元元, 等. 广义估计方程在临床试验重复测量资料中的应用[J]. 现代预防医学, 2005, 32(5): 444-5.
- [11] 宋秋月, 易东, 伍亚舟. 不同作业相关结构的纵向数据GEE分析[C]//2017年中国卫生统计学学术年会论文集. 武汉, 2017: 76.
- [12] Zou GY, Donner A. Extension of the modified Poisson regression model to prospective studies with correlated binary data[J]. Stat Methods Med Res, 2013, 22(6): 661-70.
- [13] Ford WP, Westgate PM. A comparison of bias-corrected empirical covariance estimators with generalized estimating equations in small-sample longitudinal study settings[J]. Stat Med, 2018, 37(28): 4318-29.
- [14] Li P, Redden DT. Small sample performance of bias-corrected sandwich estimators for cluster-randomized trials with binary outcomes[J]. Stat Med, 2015, 34(2): 281-96.
- [15] Pan W. Sample size and power calculations with correlated binary data[J]. Control Clin Trials, 2001, 22(3): 211-27.
- [16] Jung SH, Ahn CW. Sample size for a two-group comparison of repeated binary measurements using GEE[J]. Stat Med, 2005, 24(17): 2583-96.
- [17] Kim HY, Williamson JM, Lyles CM. Sample-size calculations for studies with correlated ordinal outcomes[J]. Stat Med, 2005, 24(19): 2977-87.
- [18] Shih WJ. Sample size and power calculations for periodontal and other studies with clustered samples using the method of generalized estimating equations[J]. Biom J, 1997, 39(8): 899-908.
- [19] Teerenstra S, Lu B, Preisser JS, et al. Sample size considerations for

- GEE analyses of three-level cluster randomized trials[J]. *Biometrics*, 2010, 66(4): 1230-7.
- [20] Li F, Forbes AB, Turner EL, et al. Power and sample size requirements for GEE analyses of cluster randomized crossover trials [J]. *Stat Med*, 2019, 38(4): 636-49.
- [21] Li F, Turner EL, Preisser JS. Sample size determination for GEE analyses of stepped wedge cluster randomized trials[J]. *Biometrics*, 2018, 74(4): 1450-8.
- [22] Li F. Design and analysis considerations for cohort stepped wedge cluster randomized trials with a decay correlation structure[J]. *Stat Med*, 2020, 39(4): 438-55.
- [23] Zhang S, Cao J, Ahn C. A GEE approach to determine sample size for pre- and post-intervention experiments with dropout[J]. *Comput Stat Data Anal*, 2014, 69: 10.1016/j.csda.2013.07.037.
- [24] Tang YQ. Power and sample size for GEE analysis of incomplete paired outcomes in 2×2 crossover trials[J]. *Pharm Stat*, 2021, 20(4): 820-39.
- [25] 刘 凯. 基于广义估计方程样本量的计算[D]. 大连: 大连理工大学, 2020.
- [26] Tang YQ. A noniterative sample size procedure for tests based on t distributions[J]. *Stat Med*, 2018, 37(22): 3197-213.
- [27] Spiegelman D, Hertzmark E. Easy SAS calculations for risk or prevalence ratios and differences[J]. *Am J Epidemiol*, 2005, 162(3): 199-200.
- [28] Mancl LA, DeRouen TA. A covariance estimator for GEE with improved small-sample properties[J]. *Biometrics*, 2001, 57(1): 126-34.
- [29] Kauermann G, Carroll RJ. A note on the efficiency of sandwich covariance matrix estimation[J]. *J Am Stat Assoc*, 2001, 96(456): 1387-96.
- [30] Huang YJ, Pan JX. Joint generalized estimating equations for longitudinal binary data[J]. *Comput Stat Data Anal*, 2021, 155: 107110.
- [31] Tsai MY. Concordance correlation coefficients estimated by generalized estimating equations and variance components for longitudinal repeated measurements[J]. *Stat Med*, 2017, 36(8): 1319-33.
- [32] Janse RJ, Hoekstra T, Jager KJ, et al. Conducting correlation analysis: important limitations and pitfalls[J]. *Clin Kidney J*, 2021, 14(11): 2332-7.

(编辑:吴锦雅)