

# 四种率比置信区间计算方法的比较研究

李太顺<sup>1</sup> 蔡若雯<sup>1</sup> 陈炳为<sup>2</sup> 许碧云<sup>1Δ</sup>

**【摘要】** 目的 比较四种率比置信区间的计算方法,为临床应用提供参考。方法 本研究选取 Koopman 法、SCAS 法、MOVER 法衍生的 Jeffreys 和 Clopper-Pearson 法分别计算不同样本量下率比的置信区间,通过比较置信区间的覆盖率和精密度进行评价。结果 Koopman 法计算的率比置信区间精密度最高,Clopper-Pearson 方法的精密度最低。随着样本量的增大或率比的减小,四种方法的精密度均逐渐升高。无论样本量大小,Clopper-Pearson 法的覆盖率均高于其他方法;Jeffreys 法在小样本时、以及单组率取两端值时覆盖率均较差;SCAS 和 Koopman 法较为接近,覆盖率均在 0.95 左右变化。结论 在实际应用场景中优先推荐使用 Koopman 方法,其次为 SCAS 法,不推荐使用 Jeffreys 法和 Clopper-Pearson 法估计率比的置信区间。

**【关键词】** 置信区间 率比 覆盖率

**【中图分类号】** R195.1 **【文献标识码】** A

**DOI** 10.11783/j.issn.1002-3674.2024.06.013

临床试验中通常将两率的比简称为率比。率比作为常用的评价指标通常出现在两组比较非劣效临床试验和诊断试验评价当中。当对照组有效率较高时,非劣效临床试验评价中通常使用率比的置信区间进行有效性评价<sup>[1]</sup>。诊断试剂评价时,也通常需要计算阳性似然比、阴性似然比等率比的置信区间进行评价。当前关于率比置信区间的计算方法有很多,SAS 等统计软件计算率比的置信区间时默认选择 Katz log 法<sup>[2]</sup>,但该方法无法计算发生率为 0 的极端情况。本文选择常用的 Koopman 法、SCAS 法、MOVER 法三种率比置信区间计算方法进行比较评价,并介绍软件实现方法,以期为临床实际应用提供参考。

## 方法简介

本研究为两独立样本情景下二项资料率比置信区间估计研究,假设  $x_0 \sim \text{Bin}(n_0, P_0)$ ,  $x_1 \sim \text{Bin}(n_1, P_1)$ ,令  $\theta = P_1/P_0(P_1 = \theta P_0)$ 。本研究使用的率比置信区间估计方法介绍如下。

### 1. Koopman 法

Koopman 在 1984<sup>[3]</sup>年提出,基于卡方检验建立  $\theta$  的置信区间。置信区间的上下限为以下方程的解<sup>[4]</sup>:

$$X^2(\theta) = \frac{(x_0 - n_0 \tilde{P}_0)^2}{n_0 \tilde{P}_0 (1 - \tilde{P}_0)} \left\{ 1 + \frac{n_0 (1 - \theta \tilde{P}_0)}{\theta n_1 (1 - \tilde{P}_0)} \right\} = z_{\alpha/2}^2 \quad (1)$$

其中  $\tilde{p}_0$  是  $p_0$  的极大似然估计值,已知  $P_1 = \theta P_0$ ,由以下方程求出  $\tilde{P}_0$ :

$$(x_0 - n_0 \tilde{P}_0) / (1 - \tilde{P}_0) + (x_1 - n_1 \tilde{P}_1) / (1 - \tilde{P}_1) = 0 \quad (2)$$

则最终  $\theta$  置信区间的上下限为公式(2)中  $\tilde{P}_0$  最小两个解代入公式(1)中求解所得值。

### 2. Skewness-corrected asymptotic score (SCAS) 法

SCAS 法是一种基于偏度校正得分计算置信区间的方法,该方法可用于计算服从二项分布的率、率差和率比的可信区间。该法计算率比的置信区间上下限为以下方程的两个解<sup>[5]</sup>:

$$Z(\theta) = \frac{S(\theta)}{\tilde{V}^{1/2}} - \frac{(z_{1-\alpha/2}^2 - 1)\tilde{\mu}_3}{6\tilde{V}^{3/2}} = \pm z_{1-\alpha/2} \quad (3)$$

公式(3)中涉及到偏度的估计值:  $\tilde{\mu}_3/\tilde{V}^{3/2}$ ,其中  $\tilde{\mu}_3$  是基于  $\tilde{P}_0$  和  $\tilde{P}_1$  的  $S(\theta)$  的三阶中心矩, $\tilde{P}_0$  和  $\tilde{P}_1$  分别为  $P_0$  和  $P_1$  的极大似然估计值, $\tilde{V}$  是基于  $\tilde{P}_0$  和  $\tilde{P}_1$  的  $S(\theta)$  的方差。上式中  $S(\theta) = \hat{P}_1 - \theta \hat{P}_0$ ,其中  $\hat{P}_0 = \frac{X_0}{n_0}$ ,  $\hat{P}_1 = \frac{X_1}{n_1}$ ,分别为  $P_0$  和  $P_1$  的样本率。

### 3. Method of variance estimates recovery (MOVER) 法

假定  $\hat{P}_0 = \frac{X_0}{n_0}$ ,  $\hat{P}_1 = \frac{X_1}{n_1}$ ,根据 MOVER 法计算率比的置信区间原则,首先需获得  $P_1 - \theta P_0$  的置信区间,区间上下限值计算公式如下<sup>[6]</sup>:

$$\begin{cases} l = \hat{P}_1 - \theta \hat{P}_0 - \sqrt{(\hat{P}_1 - l_1)^2 + \theta_2 (u_0 - \hat{P}_0)^2} \\ u = \hat{P}_1 - \theta \hat{P}_0 + \sqrt{(u_1 - \hat{P}_1)^2 + \theta_2 (\hat{P}_0 - l_0)^2} \end{cases} \quad (4)$$

我们已知  $P_1 - \theta P_0 = 0$ ,推导出  $l = 0$  且  $u = 0$ ,则公式(4)就变成关于  $\theta$  的二次方程,上式中  $(l_0, u_0)$  为  $\hat{P}_0$  的置信区间,  $(l_1, u_1)$  为  $\hat{P}_1$  的置信区间。 $l = 0$  时方程的最小根和  $u = 0$  时方程的最大根即为  $\theta$  置信区间的上限和下限值。关于  $\theta$  置信区间公式转换如下:

$$\begin{cases} \theta_l = \frac{\hat{P}_1 \hat{P}_0 - \sqrt{(\hat{P}_1 \hat{P}_0)^2 - l_1 u_0 (2 \hat{P}_1 - l_1) (2 \hat{P}_0 - u_0)}}{u_0 (2 \hat{P}_0 - u_0)} \\ \theta_u = \frac{\hat{P}_1 \hat{P}_0 + \sqrt{(\hat{P}_1 \hat{P}_0)^2 - l_0 u_1 (2 \hat{P}_1 - u_1) (2 \hat{P}_0 - l_0)}}{l_0 (2 \hat{P}_0 - l_0)} \end{cases} \quad (5)$$

1. 南京大学医学院附属鼓楼医院 医学统计分析中心 (210009)

2. 东南大学公共卫生学院 流行病与卫生统计学系

Δ通信作者:许碧云, E-mail: biyunxu@163.com

由公式(5)可知,率的置信区间上下限( $l_0, u_0$ )和( $l_1, u_1$ )成为求解( $\theta_l, \theta_u$ )的关键参数。本研究中采用两种方法计算率的可信区间,分别是 Jeffreys 区间估计法和 Clopper-Pearson 区间估计法<sup>[7]</sup>。假定  $x \sim \text{Bin}(n, P)$ , 且  $P$  的先验分布为 Jeffrey 无信息先验:  $Beta(\frac{1}{2}, \frac{1}{2})$ , 则 Jeffreys 置信区间估计法表达如下:

$$\begin{cases} P_l = Beta(\alpha/2; x+1/2, n-x+1/2) \\ P_u = Beta(1-\alpha/2; x+1/2, n-x+1/2) \end{cases} \quad (6)$$

Clopper-Pearson 置信区间估计法表达如下:

$$\begin{cases} P_l = Beta(\alpha/2; x, n-x+1) \\ P_u = Beta(1-\alpha/2; x+1, n-x) \end{cases} \quad (7)$$

### 评价指标

#### 1. 覆盖率

评价置信区间表现的主要属性是覆盖率(coverage probability, CP), 为置信区间包含真值的概率, 我们倾向于让这一值接近  $1-\alpha$  (名义覆盖率, 通常设置为 95%)。覆盖率精准计算公式如下<sup>[8]</sup>:

$$CP = \sum_{x_1=0}^{n_1} \sum_{x_0=0}^{n_0} f(x | P_1, P_0) \cdot I(x | P_1, P_0) \quad (8)$$

其中  $x$  表示四格表:  $\{x_0, n_0-x_0, x_1, n_1-x_1\}$ ,  $f$  是在给定  $P_1, P_0$  和  $I$  (如果  $x$  的置信区间包含真实值则等于 1, 其他情况为 0 的指标函数) 的条件下, 观察到  $x$  的概率。简而言之, 一个给定的区间和点的覆盖率是所有置信区间包含真实值的情况下对应的四格表概率之和。上述公式中  $f$  函数表达如下:

$$f(x | P_1, P_0) = \binom{n_1}{x_1} P_1^{x_1} (1-P_1)^{n_1-x_1} \times \binom{n_0}{x_0} (P_0)^{x_0} (1-P_0)^{n_0-x_0} \quad (9)$$

表 1 四种方法计算率比可信区间及精密度

$n$	$\theta$	$P_1$	SCAS			Jeffreys			Clopper-Pearson			Koopman		
			$\theta_l$	$\theta_u$	$W$	$\theta_l$	$\theta_u$	$W$	$\theta_l$	$\theta_u$	$W$	$\theta_l$	$\theta_u$	$W$
20	0.5	0.1	0.08	2.35	2.27	0.09	2.19	2.10	0.06	2.75	2.69	0.11	2.09	1.98
		0.2	0.15	1.35	1.20	0.17	1.31	1.14	0.13	1.47	1.34	0.18	1.31	1.13
		0.3	0.21	1.03	0.82	0.22	1.02	0.80	0.19	1.10	0.91	0.23	1.02	0.79
		0.4	0.25	0.85	0.60	0.26	0.85	0.59	0.24	0.90	0.66	0.27	0.85	0.58
	1	0.2	0.27	3.75	3.48	0.29	3.46	3.17	0.23	4.29	4.06	0.31	3.25	2.94
		0.4	0.45	2.23	1.78	0.46	2.16	1.70	0.42	2.38	1.96	0.47	2.14	1.67
		0.6	0.58	1.73	1.15	0.59	1.68	1.09	0.56	1.80	1.24	0.59	1.70	1.11
		0.8	0.70	1.43	0.73	0.72	1.39	0.67	0.68	1.47	0.79	0.70	1.42	0.72
	2	0.2	0.43	13.07	12.64	0.46	10.64	10.18	0.36	17.89	17.53	0.48	8.70	8.22
		0.4	0.74	6.49	5.75	0.76	6.01	5.25	0.68	7.45	6.77	0.76	5.55	4.79
		0.6	0.97	4.80	3.83	0.98	4.57	3.59	0.91	5.22	4.31	0.98	4.40	3.42
		0.8	1.17	3.97	2.80	1.17	3.83	2.66	1.11	4.22	3.11	1.18	3.76	2.58

#### 2. 精密度

本研究采用绝对精密度来反映置信区间精密程度大小, 绝对精密度 = 置信区间上限值( $\theta_u$ ) - 置信区间下限值( $\theta_l$ )<sup>[9]</sup>。此值越小, 说明置信区间精密度越高。

### 结果与分析

#### 1. 四种方法计算率比可信区间及精密度比较

本模拟情景假定两组样本相同即  $n_1 = n_0$ , 表 1 中统一用  $n$  表示,  $n$  的取值选择 (20, 60, 100) 三种情况; 率比  $\theta$  的取值选择 (0.5, 1, 2) 三种情况; 其中一组率  $P_1$  的取值选择 (0.1, 0.2, 0.3, 0.4) 四种选择。  $\theta_l$  为置信区间下限,  $\theta_u$  为置信区间上限,  $W$  为置信区间绝对精度。研究结果发现, 任意组合下 Koopman 方法计算的率比置信区间精密度最高, Clopper-Pearson 方法的精密度最低。纵向比较发现: 随着  $n$  的增大, 四种方法置信区间的精密度均逐渐升高; 随着  $\theta$  的增大, 四种方法置信区间的精密度均逐渐降低; 随着  $P_1$  的增大, 四种方法置信区间的精密度均逐渐升高。具体结果详见表 1。

#### 2. 四种方法计算率比可信区间及覆盖率比较

本模拟情景假定两组样本相同即  $n_1 = n_0$ ,  $n_1$  与  $n_0$  的取值选择 (10, 60) 两种情况; 率比  $\theta$  的取值选择 (1, 2) 两种情况; 其中一组率  $P_1$  的取值为 (0, 1) 之间。根据公式(9)计算出 4 种方法的覆盖率, 以单组率  $P_1$  为横轴, 覆盖率为纵轴, 在直角坐标系中作图, 得到图 1 和图 2。由图 1 和图 2 的结果可知, 无论在小样本还是大样本下, Clopper-Pearson 法的覆盖率均高于其他方法; Jeffreys 法在小样本下覆盖率表现欠佳, 且当单组率  $P_1$  取两端值时覆盖率较差; SCAS 和 Koopman 方法较为接近, 且均在 0.95 左右变化。

续表 1

n	$\theta$	$P_1$	SCAS			Jeffreys			Clopper-Pearson			Koopman		
			$\theta_l$	$\theta_u$	W	$\theta_l$	$\theta_u$	W	$\theta_l$	$\theta_u$	W	$\theta_l$	$\theta_u$	W
60	0.5	0.1	0.19	1.21	1.02	0.19	1.20	1.01	0.17	1.30	1.13	0.20	1.20	1.00
		0.2	0.26	0.89	0.63	0.27	0.89	0.62	0.25	0.92	0.67	0.27	0.89	0.62
		0.3	0.31	0.76	0.45	0.32	0.76	0.44	0.30	0.78	0.48	0.32	0.76	0.44
		0.4	0.35	0.68	0.33	0.35	0.69	0.34	0.34	0.70	0.36	0.35	0.68	0.33
	1	0.2	0.48	2.08	1.60	0.49	2.05	1.56	0.46	2.18	1.72	0.49	2.02	1.53
		0.4	0.64	1.56	0.92	0.64	1.55	0.91	0.62	1.60	0.98	0.64	1.55	0.91
		0.6	0.74	1.35	0.61	0.74	1.34	0.60	0.73	1.37	0.64	0.74	1.35	0.61
		0.8	0.83	1.21	0.38	0.83	1.20	0.37	0.82	1.22	0.40	0.83	1.21	0.38
	2	0.2	0.82	5.37	4.55	0.83	5.17	4.34	0.77	5.87	5.10	0.83	4.88	4.05
		0.4	1.13	3.77	2.64	1.13	3.71	2.58	1.09	3.93	2.84	1.13	3.64	2.51
		0.6	1.31	3.21	1.90	1.31	3.17	1.86	1.28	3.29	2.01	1.31	3.14	1.83
		0.8	1.46	2.89	1.43	1.46	2.86	1.40	1.43	2.94	1.51	1.46	2.85	1.39
100	0.5	0.1	0.24	0.99	0.75	0.24	0.99	0.75	0.23	1.04	0.81	0.25	0.99	0.74
		0.2	0.31	0.78	0.47	0.31	0.78	0.47	0.30	0.80	0.50	0.31	0.78	0.47
		0.3	0.35	0.69	0.34	0.35	0.69	0.34	0.34	0.71	0.37	0.35	0.69	0.34
		0.4	0.38	0.64	0.26	0.38	0.64	0.26	0.37	0.65	0.28	0.38	0.64	0.26
	1	0.2	0.57	1.75	1.18	0.57	1.74	1.17	0.55	1.81	1.26	0.58	1.73	1.15
		0.4	0.71	1.41	0.70	0.71	1.41	0.70	0.70	1.43	0.73	0.71	1.40	0.69
		0.6	0.79	1.26	0.47	0.80	1.26	0.46	0.79	1.27	0.48	0.80	1.26	0.46
		0.8	0.87	1.15	0.28	0.87	1.15	0.28	0.86	1.16	0.30	0.87	1.15	0.28
	2	0.2	1.01	4.22	3.21	1.01	4.15	3.14	0.96	4.45	3.49	1.01	4.02	3.01
		0.4	1.28	3.24	1.96	1.28	3.21	1.93	1.25	3.32	2.07	1.28	3.18	1.90
		0.6	1.44	2.86	1.42	1.44	2.85	1.41	1.42	2.91	1.49	1.44	2.83	1.39
		0.8	1.56	2.64	1.08	1.56	2.63	1.07	1.54	2.67	1.13	1.56	2.63	1.07

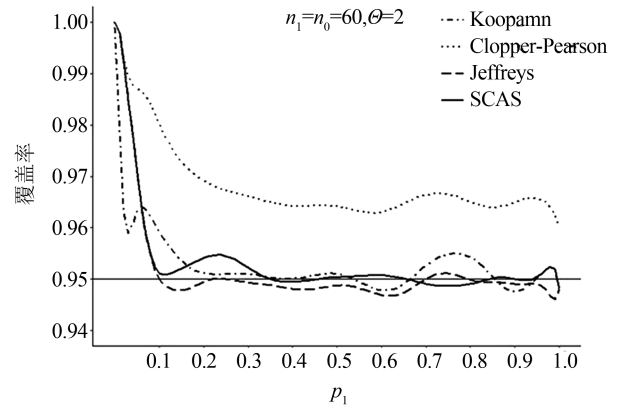
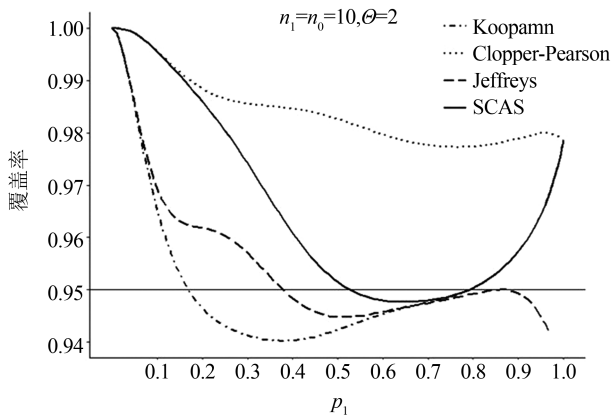
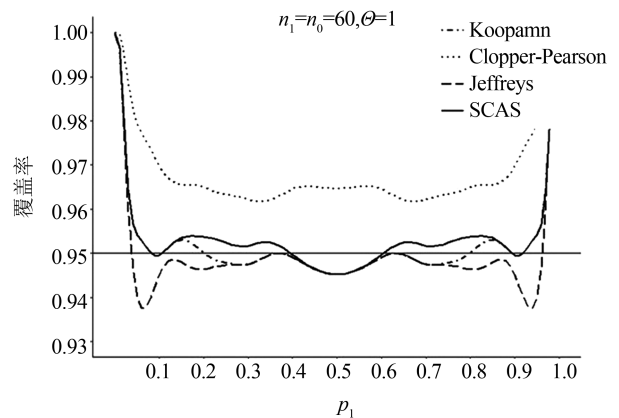
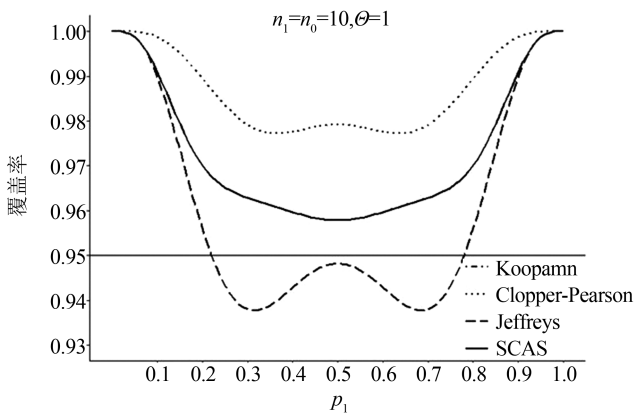


图 1 四种方法率比 95% 置信区间的覆盖率 ( $n_1 = n_0 = 10$ )

图 2 四种方法率比 95% 置信区间的覆盖率 ( $n_1 = n_0 = 60$ )

## 案例分析

在一项关于在急诊中快速诊断心衰的研究中<sup>[10]</sup>, 自1999年4月到2000年12月共纳入1586名来自美国、法国或挪威的有显著呼吸短促症状的患者。待研究的诊断方法为患者血液样本中的B-型利钠肽浓度(每毫升100毫克或以上为阳性,每毫升不足100毫克为阴性), 确诊标准为两名心脏病专家根据患者的Framingham充血性心力衰竭评分做出的诊断。试验结果如表2所示。评价指标包含灵敏度、特异度、阳性似然比及阴性似然比, 结果需汇报可信区间。本研究根据4种率比置信区间计算方法所得的该试验阳性似然比和阴性似然比的置信区间如表3所示。

表2 血液中B-型利钠肽浓度用于快速诊断心衰试验的结果

		充血性心力衰竭		
		有	无	总和
B-型利钠肽浓度 检测法	阳性	670	202	872
	阴性	74	640	714
	总和	744	842	1586

表3 四种方法计算B-型利钠肽浓度快速诊断心衰试验的阳性/阴性似然比的置信区间

	阳性似然比(95% CI)	阴性似然比(95% CI)
Koopman 法	3.75(3.33~4.25)	0.13(0.10~0.16)
SCAS 法	3.75(3.33~4.26)	0.13(0.10~0.16)
Jeffreys 法	3.75(3.33~4.26)	0.13(0.10~0.16)
Clopper-Pearson 法	3.75(3.32~4.27)	0.13(0.10~0.16)

## 讨论与建议

从模拟结果来看, Clopper-Pearson 法估计率比置信区间时最为保守, 虽然其覆盖率最高, 但其估计的精度最差。Jeffreys 法在小样本下覆盖率表现欠佳, 且当单组率处于两端时, 覆盖率并不稳定。因此并不推荐在实际场景中应用 Jeffreys 法和 Clopper-Pearson 法估计率比的置信区间。

SCAS 和 Koopman 方法覆盖率表现接近, 且均在0.95左右变化, 接近名义覆盖率。但从精度模拟结果来看, Koopman 方法精度最高。因此, 在实际应用场景中优先推荐使用 Koopman 方法, 其次为 SCAS 法。

从软件实现的角度来看, 四种方法均可通过 R 软件实现。SCAS 法、Jeffreys 法和 Clopper-Pearson 法可通过“ratesci”R 包计算; Koopman 法可通过“contingencytables”R 包计算。

## 参 考 文 献

- [1] James HH, Wang SJ, Tsong Y, et al. Some fundamental issues with non-inferiority testing in active controlled trials [J]. *Statistics in Medicine*, 2003, 22(2): 213-225.
- [2] Katz D, Baptista J, Azen SP, et al. Obtaining confidence intervals for the risk ratio in cohort studies [J]. *Biometrics*, 1978, 34: 469-474.
- [3] Koopman PAR. Confidence intervals for the ratio of two binomial proportions [J]. *Biometrics*, 1984, 40: 513-517.
- [4] Nam J. Confidence limits for the ratio of two binomial proportions based on likelihood scores: non-iterative method [J]. *Biometrical Journal*, 1995, 37: 375-379.
- [5] Laud PJ. Equal-tailed confidence intervals for comparison of rates [J]. *Pharmaceutical Statistics*, 2017, 16(5): 334-348.
- [6] Donner A, Zou GY. Closed-form confidence intervals for functions of the normal mean and standard deviation [J]. *Statistical Methods in Medical Research*, 2012, 21(4): 347-359.
- [7] Brown LD, Cai TT, DasGupta A, et al. Interval estimation for a binomial proportion [J]. *Statistical Science*, 2001, 16(2): 101-133.
- [8] Fagerland MW, Lydersen S, Laake P. Recommended confidence intervals for two independent binomial proportions [J]. *Statistical Methods in Medical Research*, 2015, 24(2): 224-254.
- [9] 刘沛. 四种方法计算总体率可信区间的比较研究 [J]. *中国卫生统计*, 2005, 22(6): 354-358.
- [10] Maisel AS, Krishnaswamy P, Nowak RM, et al. Rapid measurement of B-type natriuretic peptide in the emergency diagnosis of heart failure [J]. *N Engl J Med*, 2002, 347(3): 161-167.

(责任编辑:张悦)