

# Confidence Interval Construction of Risk Ratio of Chronic Diseases Based on Gradient Statistics under Poisson Distribution

Haimiao Meng<sup>1</sup>, Huiying Jia<sup>2</sup>, Zhen Yan<sup>1\*</sup>

<sup>1</sup>College of Mathematics and Statistics, Guangxi Normal University, Guilin Guangxi

<sup>2</sup>Department of Education, Guilin Normal College, Guilin Guangxi

Email: \*zhenyan@gxnu.edu.cn, 1328028408@qq.com

Received: Dec. 31<sup>st</sup>, 2019; accepted: Jan. 14<sup>th</sup>, 2020; published: Jan. 21<sup>st</sup>, 2020

## Abstract

Combining the advantages of Poisson sampling and the characteristics of long occurrence period and low incidence of chronic diseases, the confidence interval of risk ratio under Poisson distribution is constructed by using gradient statistics. The advantage of this method is that it does not need to calculate Fisher information matrix and its inverse matrix, which greatly simplifies the calculation process compared with the score statistics method. At the same time, an example and Monte Carlo simulation are used. Compared with the traditional confidence interval construction method, the results show that the method based on gradient statistics can obtain better coverage probability and shorter interval length.

## Keywords

Poisson Distribution, Gradient Statistics, Risk Ratio, Monte Carlo Simulation

# Poisson分布下基于梯度统计量的慢性病风险比的置信区间构造

蒙海苗<sup>1</sup>, 贾慧英<sup>2</sup>, 晏振<sup>1\*</sup>

<sup>1</sup>广西师范大学数学与统计学院, 广西 桂林

<sup>2</sup>桂林师范高等专科学校教育系, 广西 桂林

Email: \*zhenyan@gxnu.edu.cn, 1328028408@qq.com

收稿日期: 2019年12月31日; 录用日期: 2020年1月14日; 发布日期: 2020年1月21日

\*通讯作者。

## 摘要

结合Poisson抽样的优点以及慢性病发病周期长和发病率低的特点,利用梯度统计量的方法来构造Poisson分布下风险比的置信区间。该方法的优点在于不需要计算Fisher信息矩阵及其逆矩阵,与得分统计量方法相比,这大大简化了计算过程。同时,通过实例和蒙特卡洛模拟,与传统的置信区间构造方法进行比较。结果表明,基于梯度统计量的构建方法可以得到较好的覆盖概率和较短的区间长度。

## 关键词

Poisson分布, 梯度统计量, 风险比, 蒙特卡洛模拟

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

在一些罕见疾病的研究过程中,往往需要用很长的一段时间来对研究对象进行跟进,由于在长期的跟进过程中有些研究对象可能会退出研究,从而导致在跟进过程中,研究对象的数量会时常发生变化。Poisson分布是一种常用的离散分布,它常常与单位时间上的计数过程相联系,用来描述单位时间内事件发生特定次数的概率。因此,根据Poisson分布的优点可以较好地解决上述问题。

在流行病学中,风险比指的是暴露组的发病率与非暴露组的发病率的比值。在一定的显著水平上,当风险比小于1时,说明某药物或处理对疾病治疗有效;当风险比与1无显著差异时,说明该药物或处理对疾病治疗无影响;当风险比大于1时,说明该药物或处理可能会对人体产生副作用,会导致其他疾病的产生,进而需要更进一步研究探讨。

在流行病学的研究中,对风险比的研究很普遍,Liu [1]给出了基于Fieller定理的方法和基于F检验统计量的方法。Liu [2]给出了Delta方法,该方法运用泰勒展开原理,使得在已知单个随机变量的分布条件下,能够逼近随机变量函数的均值和方差。Tang [3]等人给出了基于得分统计量的方法、基于似然比统计量的方法,对于小样本量时,用鞍点逼近的方法等。由于鞍点逼近的方法能够在保证区间覆盖率的前提下使得区间长度在所有的的方法中是最短的。同时,得分统计量的方法也可以在覆盖率和区间长度之间取得很好的平衡。对于上述提出的几种方法,在计算过程中较为繁琐,例如,得分统计量方法需要计算Fisher信息矩阵及其逆矩阵,因此,对于参数检验问题,Terrell [4]提出了一种新的检验统计量,这一新的梯度统计量可视之为WALD以及得分统计量的一种巧妙结合,明显优于WALD以及得分统计量,不需要计算Fisher信息矩阵及其逆矩阵。Lemonte [5]等人认为梯度统计量、似然比统计量以及得分检验统计量在一阶渐近意义下是等价的,同时没有哪个统计量在二阶局部功效方面更优。此外,牛翠珍和范国良[6]基于梯度统计量得出负二项抽样下风险差的近似区间估计。

大多数的风险比置信区间的构造方法都是基于二项抽样和负二项抽样下的,相对而言,有关于Poisson分布和风险比研究的文献较少。白永昕和田茂再[7][8]研究了基于Poisson抽样下慢性病发病率的置信区间的构造,Poisson分布下基于鞍点逼近的慢性病风险差的置信区间的构造。因此,本文利用梯度统计量的优点,研究基于梯度统计量的Poisson分布下风险比的置信区间的构造,并通过例子和蒙特卡洛模拟,将基于梯度统计量的方法与对数转换方法、得分统计量的方法以及似然比统计量方法进行比较。

## 2. 风险比置信区间的已有方法

假设  $i=1,0$  分别表示暴露组和非暴露组, 对第  $i$  组中  $n_i$  个研究对象跟进并得到  $X_i$  例发病数。假设  $X_i$  服从参数为  $n_i^* \lambda_i$  的 Poisson 分布, 其中  $n_i^*$  表示单位时间内的平均跟进对象,  $\lambda_1$  和  $\lambda_0$  ( $0 < \lambda_i < 1$ ) 分别表示暴露组和非暴露组的发病率, 从而随机变量  $X_i$  的分布函数为:

$$P(X_i = x_i | \lambda_i, n_i^*) = \frac{(\lambda_i n_i^*)^{x_i}}{x_i!} \exp(-\lambda_i n_i^*), \quad x_i = 0, 1, \dots, n.$$

本文考虑暴露组和非暴露组发病率的比, 即风险比  $R = \lambda_1 / \lambda_0$ 。从而关于风险比  $R$  和  $\lambda_0$  的似然函数为

$$L(R, \lambda_0) = \frac{(R \lambda_0 n_1^*)^{x_1}}{x_1!} \exp(-R \lambda_0 n_1^*) \cdot \frac{(\lambda_0 n_0^*)^{x_0}}{x_0!} \exp(-\lambda_0 n_0^*) \quad (2.1)$$

则对(2.1)的似然函数极大化, 得到似然估计分别为

$$\hat{R} = \frac{x_1 n_0^*}{x_0 n_1^*}, \quad \hat{\lambda}_0 = \frac{x_0}{n_0^*}.$$

### 2.1. 对数转换方法

对数转换方法在一定程度上是 Delta 方法的改进。若用 Delta 方法估计风险比估计量  $\hat{R}$  的渐近分布, 只有在大样本量时  $\hat{R}$  的分布才能非常接近正态分布[9], Kata 等人提出的对数变换方法[10], 认为在样本量较小的情况下, 变量  $\log(\hat{R})$  会更加接近正态分布, 再对  $\log(\hat{R})$  运用 Delta 方法可以计算其均值为  $\log(R)$ , 方差为  $1/(\hat{R} n_1 \hat{\lambda}_0) + 1/(n_0 \hat{\lambda}_0)$ , 从而得到

$$\frac{\log(\hat{R}) - \log(R)}{1/(R n_1 \lambda_0) + 1/(n_0 \lambda_0)} \sim N(0, 1).$$

因此, 得到对数转换方法下风险比  $R$  在渐近正态分布下的一个水平为  $100(1-\alpha)\%$  的渐近置信区间为

$$\left[ \hat{R} \exp\left(-Z_{\alpha/2} \sqrt{\text{Var}(\log(\hat{R}))}\right), \hat{R} \exp\left(Z_{\alpha/2} \sqrt{\text{Var}(\log(\hat{R}))}\right) \right],$$

其中  $\text{Var}(\log(\hat{R})) = 1/(\hat{R} n_1 \hat{\lambda}_0) + 1/(n_0 \hat{\lambda}_0)$ ,  $Z_{\alpha/2}$  是标准正态分布  $1 - \frac{\alpha}{2}$  分位点。

### 2.2. 得分检验方法

在大样本的情况下, 得分统计量为

$$T_s = s(\theta)' I(\theta)^{-1} s(\theta),$$

其中,  $s(\theta) = \frac{\partial}{\partial \theta} \log f(X | \theta)$  是得分函数,  $I(\theta) = -E\left(\frac{\partial^2}{\partial \theta^2} \log f(X | \theta)\right)$  是 Fisher 信息矩阵。在本文中, 得分函数为

$$s(R, \lambda_0) = \begin{pmatrix} \frac{\partial \log(L(R, \lambda_0))}{\partial R} \\ \frac{\partial \log(L(R, \lambda_0))}{\partial \lambda_0} \end{pmatrix} = \begin{pmatrix} \frac{x_1}{R} - \lambda_0 n_1^* \\ \frac{x_1 + x_0}{\lambda_0} - R n_1^* - n_0^* \end{pmatrix};$$

Fisher 信息矩阵为

$$I(R, \lambda_0) = -E \begin{pmatrix} \frac{\partial^2 \log(L(R, \lambda_0))}{\partial R^2} & \frac{\partial \log(L(R, \lambda_0))}{\partial R \partial \lambda_0} \\ \frac{\partial \log(L(R, \lambda_0))}{\partial R \partial \lambda_0} & \frac{\partial^2 \log(L(R, \lambda_0))}{\partial \lambda_0^2} \end{pmatrix} = \begin{pmatrix} \frac{\lambda_0 n_1^*}{R} & n_1^* \\ n_1^* & \frac{n_0^* + R n_1^*}{\lambda_0} \end{pmatrix}.$$

令原假设下  $\lambda_0$  的限制极大似然估计为  $\hat{\lambda}_0(R_0)$ , 求解以下方程就可以得到  $\hat{\lambda}_0(R_0)$ :

$$\left. \frac{\partial \log L(R, \lambda_0)}{\partial \lambda_0} \right|_{R=R_0} = \frac{x_1 + x_0}{\lambda_0} - R n_1^* - n_0^* = 0;$$

解得  $\hat{\lambda}_0(R_0) = \frac{x_0 + x_1}{R_0 n_1^* + n_0^*}$ .

基于假设检验和置信区间的对偶关系, 考虑如下假设检验问题:

$$H_0 : R = R_0 \text{ 对 } H_1 : R \neq R_0$$

则关于  $R$  和  $\lambda_0$  的得分统计量渐近的服从自由度为 1 的卡方分布, 即

$$\begin{aligned} T_s &= s(R, \hat{\lambda}_0(R))' I(R, \hat{\lambda}_0(R))^{-1} s(R, \hat{\lambda}_0(R)) \\ &= \frac{x_1^2}{n_1 R \hat{\lambda}_0(R)} + \frac{x_0^2}{n_0 \hat{\lambda}_0(R)} + R n_1 \hat{\lambda}_0(R) + n_0 \hat{\lambda}_0(R) + 2x_1 + 2x_0 \sim \chi_\alpha^2(1). \end{aligned}$$

因此,  $R$  的  $100(1-\alpha)\%$  的置信区间为

$$\{R : T_s \leq \chi_\alpha^2(1)\}.$$

### 2.3. 似然比检验方法

与得分方法类似, 似然比检验是一种构造参数检验的方法, 基于假设检验和置信区间的对偶关系, 在假设检验问题:

$$H_0 : \theta \in \theta_0 \text{ 对 } H_1 : \theta \in \theta/\theta_0$$

下, 有似然比统计量为

$$LR = \frac{\sup_{\theta \in \theta_0} L(\theta)}{\sup_{\theta \in \theta/\theta_0} L(\theta)}.$$

考虑如下假设检验问题:

$$H_0 : R = R_0 \text{ 对 } H_1 : R \neq R_0$$

则有

$$T_{LR} = 2 \log \left( \frac{L(\hat{R}, \hat{\lambda}_0)}{L(R, \hat{\lambda}_0(R))} \right) \sim \chi_\alpha^2(1).$$

其中  $\hat{\lambda}_0(R_0)$  是  $\lambda_0$  的限制极大似然估计. 因此,  $R$  的一个置信水平为  $100(1-\alpha)\%$  的置信区间为

$$\{R : T_{LR} \leq \chi_\alpha^2(1)\}.$$

### 3. 基于梯度统计量的风险比的置信区间的构造

得分统计量方法，似然比统计量方法的计算过程比较复杂，特别是得分方法中需要计算 Fisher 信息矩阵及其逆矩阵，而本文提出的梯度统计量方法计算相对简便，不需计算 Fisher 信息矩阵及其逆矩阵。基于上述传统的构造置信区间的方法，下面对梯度统计量方法做具体阐述。基于假设检验和置信区间的对偶关系，在假设检验问题： $H_0: R = R_0$ ，引入梯度统计量方法。

Lemote [6]等人论述的梯度统计量具有以下形式：

$$\left( (\hat{R}, \hat{\lambda}) - (R_0, \hat{\lambda}_0(R_0)) \right) s(R_0, \hat{\lambda}_0(R_0)).$$

由上述式子可以知道，梯度统计量具有简单的表达形式，而且不涉及 Fisher 信息矩阵及其逆矩阵。根据上述给出的  $\hat{\lambda}_0(R)$  的定义可知得分函数为：

$$s(R_0, \hat{\lambda}_0(R_0)) = \begin{pmatrix} \frac{x_1}{R_0} - \hat{\lambda}_0(R_0) \cdot n_1^* \\ 0 \end{pmatrix}.$$

进而可以构造一个更加简单形式的梯度统计量为：

$$T = \left( \frac{x_1}{R_0} - \hat{\lambda}_0(R_0) \cdot n_1^* \right) \cdot (\hat{R} - R),$$

其中， $\hat{R} = \frac{x_1 n_0^*}{x_0 n_1^*}$ ， $\hat{\lambda}_0(R_0) = \frac{x_0 + x_1}{R_0 n_1^* + n_0^*}$ 。

根据上述的论述，可以构造出  $R$  的一个水平为  $100(1-\alpha)\%$  的置信区间为

$$\left\{ R : \left( \frac{x_1}{R_0} - \hat{\lambda}_0(R_0) n_1^* \right) \cdot (\hat{R} - R) \leq \chi_\alpha^2(1) \right\}.$$

### 4. 蒙特卡罗模拟

对于上述描述的四种置信区间方法，本文运用蒙特卡罗模拟从区间长度和覆盖率这两个方面来评价这四种方法构建的置信区间的表现。对于覆盖率，越接近于预先设定的水平，说明该方法越好；置信区间长度则通过平均区间长度来衡量，在覆盖率能控制的情况下，区间宽度越短越好。

设定样本量  $n = 20, 50, 100$ ；非暴露组的发病率为  $\lambda_0 = 0.1, 0.2, 0.3, 0.5$ ；风险比的真值为  $R = 0.1, 0.5, 1, 2, 3$ ；置信水平为  $\alpha = 0.05$ ；每种情形重复试验 10,000 次。如下表 1 到表 5 分别表示  $\lambda_0 = 0.1, 0.2, 0.3, 0.5$ ，置信水平为  $\alpha = 0.05$  下的模拟结果，表格中的数分别表示覆盖率和置信区间长度(括号中的数)。其中，用“FT”表示对数转换方法，“SC”得分检验方法，“LR”表示似然比检验方法，“GB”表示基于梯度统计量的方法。

从蒙特卡罗模拟的结果来看(见表 1~5)，当  $n$  较小( $n_1 = n_0 = 20$ )时， $\lambda_0$  从 0.1 增大到 0.5 的过程中，得分检验方法和梯度方法的区间长度与其他两种方法相比较短，覆盖率也越来越接近预先设定的置信水平，其中得分方法比梯度方法的表现稍好一些。对数转换方法和似然比方法，虽然得到较好的覆盖率，但在小样本量的情况下，这两种方法得到的区间长度相对较宽。当  $n$  较大时，这四种方法的区间长度都比小样本时短。此外，这四种方法的区间长度都随非暴露组的发病率的增大而减小。

当非暴露组的发病率一定时，随着样本量的增大，这四种方法的区间长度都逐渐减少，覆盖率也越来越好，这点在得分方法和梯度方法表现较为明显。

总之，似然比方法的覆盖率都非常接近预先设定的显著水平 95%，梯度方法一般比预先设定的水平小。对数转换方法的覆盖率较好，但其区间长度很宽；似然比方法、得分方法、梯度方法的表现都不错，但似然比方法的区间长度不及得分方法和梯度方法的好，得分方法的覆盖率比梯度方法的稍好，从而在这四种方法中，得分方法的表现最佳。此外，如果考虑到计算过程的问题，梯度方法不需要计算 Fisher 信息阵及其逆矩阵，比得分方法的计算简便很多，因此，若是综合覆盖率，区间宽度以及计算过程，本文提出的梯度方法也有较好的表现。

**Table 1.** Coverage probabilities and Interval width with  $R_0 = 0.1$  and  $\alpha = 0.05$

**表 1.** 在  $R_0 = 0.1$ ， $\alpha = 0.05$  的覆盖率和区间宽度

$n$	$\lambda_0$	FT	SC	LR	GB
20	0.1	0.9937 (7.5638)	0.6284 (1.9312)	0.9356 (11.0512)	0.6232 (2.2238)
	0.2	1.0000 (3.8914)	0.8148 (1.3301)	0.9424 (4.3636)	0.7878 (1.4663)
	0.3	1.0000 (2.9341)	0.8810 (0.9698)	0.9528 (1.9561)	0.8431 (1.0363)
	0.5	1.0000 (0.9352)	0.9314 (0.6222)	0.9648 (0.7861)	0.9085 (0.6465)
50	0.1	1.0000 (2.9389)	0.8537 (1.1276)	0.9489 (2.8565)	0.8199 (1.2221)
	0.2	1.0000 (1.1753)	0.9324 (0.6222)	0.9641 (0.7837)	0.9072 (0.6483)
	0.3	1.0000 (0.5662)	0.9497 (0.4419)	0.9699 (0.4831)	0.9373 (0.4539)
	0.5	1.0000 (0.3567)	0.9595 (0.2964)	0.9751 (0.3063)	0.95049 (0.3032)
100	0.1	1.0000 (0.9638)	0.9318 (0.6234)	0.9600 (0.7783)	0.9076 (0.6470)
	0.2	1.0000 (0.5078)	0.9561 (0.3502)	0.9700 (0.3621)	0.9464 (0.3588)
	0.3	1.0000 (0.2882)	0.9600 (0.2607)	0.9900 (0.2593)	0.9543 (0.2657)
	0.5	0.9292 (0.2872)	0.9607 (0.1877)	0.9500 (0.1919)	0.9588 (0.1909)

**Table 2.** Coverage probabilities and Interval width with  $R_0 = 0.5$  and  $\alpha = 0.05$

**表 2.** 在  $R_0 = 0.5$ ， $\alpha = 0.05$  的覆盖率和区间宽度

$n$	$\lambda_0$	FT	SC	LR	GB
20	0.1	1.0000 (7.5393)	0.9886 (2.5778)	0.9891 (14.2371)	0.9499 (2.8512)
	0.2	1.0000 (3.6517)	0.9729 (1.7414)	0.9722 (6.9377)	0.9296 (1.9520)
	0.3	1.0000 (2.1485)	0.9667 (1.3116)	0.9637 (3.5045)	0.9347 (1.4585)
	0.5	1.0000 (1.1235)	0.9586 (0.9293)	0.9509 (1.5901)	0.9389 (1.0167)

Continued

50	0.1	1.0000 (3.9179)	0.9693 (1.4894)	0.9680 (4.8487)	0.9320 (1.6677)
	0.2	0.9504 (2.1379)	0.9579 (0.9305)	0.9511 (1.6098)	0.9384 (1.0149)
	0.3	1.0000 (1.1217)	0.9531 (0.7289)	0.9479 (1.0896)	0.9407 (0.7865)
	0.5	1.0000 (0.7306)	0.9513 (0.5501)	0.9494 (0.7651)	0.9436 (0.5873)
100	0.1	1.0000 (1.3369)	0.9574 (0.9296)	0.9507 (1.6068)	0.9410 (1.0115)
	0.2	1.0000 (0.9606)	0.9542 (0.6212)	0.9485 (0.8846)	0.9436 (0.6651)
	0.3	1.0000 (0.6884)	0.9511 (0.4986)	0.9465 (0.6846)	0.9442 (0.5306)
	0.5	0.9978 (0.6388)	0.9513 (0.3811)	0.9482 (0.5082)	0.9480 (0.4042)

**Table 3.** Coverage probabilities and Interval width with  $R_0 = 1$  and  $\alpha = 0.05$ **表 3.** 在  $R_0 = 1$ ,  $\alpha = 0.05$  下的覆盖率和区间宽度

$n$	$\lambda_0$	FT	SC	LR	GB
20	0.1	1.0000 (7.6474)	0.9967 (3.1904)	0.9874 (19.2241)	0.9416 (3.5523)
	0.2	1.0000 (4.9493)	0.9760 (2.1308)	0.9471 (10.9694)	0.9097 (2.5070)
	0.3	1.0000 (2.7711)	0.9622 (1.6486)	0.9382 (5.8897)	0.9204 (1.9361)
	0.5	0.9998 (1.4068)	0.9519 (1.2233)	0.9456 (2.6570)	0.9348 (1.3900)
50	0.1	1.0000 (3.9837)	0.9669 (1.8476)	0.9401 (7.9420)	0.9126 (2.1838)
	0.2	1.0000 (2.9950)	0.9514 (1.2225)	0.9451 (2.6816)	0.9349 (1.3901)
	0.3	1.0000 (1.2705)	0.9534 (0.9786)	0.9467 (1.7984)	0.9406 (1.0936)
	0.5	1.0000 (0.9647)	0.9564 (0.7461)	0.9493 (1.2593)	0.9443 (0.8260)
100	0.1	1.0000 (1.8042)	0.9523 (1.2243)	0.9451 (2.6905)	0.9337 (1.3930)
	0.2	1.0000 (1.4619)	0.9550 (0.8390)	0.9474 (1.4639)	0.9427 (0.9331)
	0.3	1.0000 (0.9904)	0.9553 (0.6782)	0.9480 (1.1214)	0.9452 (0.7491)
	0.5	1.0000 (0.7358)	0.9550 (0.5213)	0.9496 (0.8319)	0.9469 (0.5734)

**Table 4.** Coverage probabilities and Interval width with  $R_0 = 2$  and  $\alpha = 0.05$

**表 4.** 在  $R_0 = 2$ ,  $\alpha = 0.05$  下的覆盖率和区间宽度

$n$	$\lambda_0$	FT	SC	LR	GB
20	0.1	1.0000 (8.6772)	0.9821 (4.0924)	0.9729 (31.0173)	0.9629 (4.6890)
	0.2	0.9999 (4.9531)	0.9754 (2.7632)	0.9552 (19.7313)	0.9154 (3.4738)
	0.3	0.9109 (2.8233)	0.9688 (2.1847)	0.9482 (10.5928)	0.9202 (2.7513)
	0.5	1.0000 (4.1774)	0.9677 (1.6490)	0.9499 (4.7101)	0.9293 (1.9934)
50	0.1	1.0000 (7.1879)	0.9714 (2.4246)	0.9800 (15.9228)	0.9173 (3.0654)
	0.2	1.0000 (4.3535)	0.9662 (1.6491)	0.9400 (5.0411)	0.9301 (1.9934)
	0.3	1.0000 (2.3616)	0.9651 (1.3326)	0.9600 (3.2566)	0.9368 (1.5762)
	0.5	1.0000 (1.9199)	0.9659 (1.0219)	0.9500 (2.1806)	0.9413 (1.1907)
100	0.1	1.0000 (4.1634)	0.9665 (1.6493)	0.960000 (4.7218)	0.9289 (1.9935)
	0.2	0.6478 (3.4604)	0.9658 (1.1481)	0.9300 (2.6471)	0.9392 (1.3428)
	0.3	0.7991 (1.2897)	0.9673 (0.9315)	0.9200 (1.8757)	0.9453 (1.0809)
	0.5	1.0000 (1.3826)	0.9659 (0.7189)	0.9300 (1.4624)	0.9462 (0.8287)

**Table 5.** Coverage probabilities and Interval width with  $R_0 = 3$  and  $\alpha = 0.05$

**表 5.** 在  $R_0 = 3$ ,  $\alpha = 0.05$  下的覆盖率和区间宽度

$n$	$\lambda_0$	FT	SC	LR	GB
20	0.1	1.0000 (18.8573)	0.9781 (4.8298)	0.97370 (43.9800)	0.9728 (5.6428)
	0.2	1.0000 (15.8279)	0.9807 (3.2668)	0.9553 (29.1713)	0.9142 (4.3351)
	0.3	1.0000 (5.8851)	0.9727 (2.6024)	0.9486 (15.4977)	0.9171 (3.4721)
	0.5	1.0000 (5.9329)	0.9745 (1.9816)	0.9493 (6.7729)	0.9295 (2.5075)
50	0.1	1.0000 (10.8503)	0.9759 (2.8784)	0.95115 (20.8995)	0.9137 (3.8553)
	0.2	1.0000 (4.6360)	0.9753 (1.9836)	0.9472 (6.7878)	0.9296 (2.5083)
	0.3	1.0000 (4.5666)	0.9720 (1.6065)	0.9492 (4.5019)	0.9378 (1.9749)
	0.5	1.0000 (2.1996)	0.9737 (1.2371)	0.94941 (3.1056)	0.9425 (1.4927)

Continued

100	0.1	0.9511 (8.0301)	0.9749 (1.9826)	0.9700 (6.4451)	0.9298 (2.5105)
	0.2	1.0000 (4.1763)	0.9721 (1.3857)	0.9800 (3.4785)	0.9408 (1.6846)
	0.3	1.0000 (2.7481)	0.9732 (1.1268)	0.9700 (2.6861)	0.9426 (1.3576)
	0.5	1.0000 (2.3677)	0.9733 (0.8695)	0.9700 (0.1990)	0.9457 (1.0404)

## 5. 实例分析

根据美国心脏病学会的新标准,对患高血脂的成年人,服用他汀类药物治疗和服用安慰剂两种情况下与未来得冠心病风险之间的关系进行了研究。根据研究数据可知,在 288 名服用他汀类药物的患者中有 101 人患有冠心病,在 289 名服用安慰剂的患者中有 23 人患有冠心病。即

$$x_0 = 23, x_1 = 101, n_0 = 289, n_1 = 288,$$

$$\hat{\lambda}_0 = 0.079, \hat{\lambda}_1 = 0.351, \hat{R} = 4.443.$$

根据这些数据,运用文中提出的四种风险比的置信区间构造方法,得到风险比  $R$  的 95%置信区间,具体如下表 6 (数据来源于 Ann Marie N B [11])。

**Table 6.** 95% confidence interval of risk ratio  $R$  of statins and placebo  
**表 6.** 他汀类药物和安慰剂风险比  $R$  的 95%置信区间

方法	置信区间	区间长度
对数转换方法	[2.8018, 6.9304]	4.1286
得分检验方法	[2.8124, 6.9042]	4.0918
似然比检验方法	[2.8567, 7.0973]	4.2406
梯度统计量方法	[2.9931, 7.7883]	4.7952

从表 6 可以发现,这四种方法构造的置信区间的区间长度相差不大,没有明显的差异,其中得分方法的置信区间最短,本文所提出的梯度统计量方法的表现也较好,与得分方法相差不大。此外,这四种方法的置信上限和置信下限都大于 1,说明在 5%的水平使用他汀药物治疗会使冠心病患者的比例增加。

## 6. 结论

本文主要介绍了在 Poisson 抽样下基于梯度统计量的风险比的置信区间的构造方法,通过蒙特卡罗模拟与三种传统的估计方法进行比较。在这四种方法中,得分方法的表现最佳,但梯度统计量方法在覆盖率和区间宽度上,与其他方法无显著差异。此外,梯度方法不需要计算 Fisher 信息阵及其逆矩阵,综合考虑这四种的方法的计算过程,梯度统计量方法简便很多。因此,若是综合覆盖率,区间宽度以及计算过程,本文提出的梯度方法也有较好的表现。

## 基金项目

国家自然科学基金项目(11901124, 11701109, 11861017); 广西自然科学基金项目(2018GXNSFAA138164)。

广西研究生教育创新计划项目(XYCSZ2017076); 广西师范大学重点科研项目(2016ZD002)。

## 参考文献

- [1] Lui, K.J. (1995) Confidence Intervals for the Risk Ratio in Cohort Studies under Inverse Sampling. *Biometrical Journal*, **8**, 965-971. <https://doi.org/10.1002/bimj.4710370808>
- [2] Lui, K.J. (1986) *Statistical Estimation of Epidemiological Risk*. Wiley, New York.
- [3] Tian, M., Tang, M.L., Ng, H.K., *et al.* (2010) Confidence Intervals for the Risk Ratio under Inverse Sampling. *Statistics in Medicine*, **27**, 3301-3324. <https://doi.org/10.1002/sim.3158>
- [4] Terrell, G.R. (2002) The Gradient Statistic. *Computing Science and Statistics*, **34**, 206-215.
- [5] Lemonte, A.J. and Ferrari, S.L.P. (2012) The Local Power of the Gradient Test. *Annals of the Institute of Statistical Mathematics*, **64**, 373-381. <https://doi.org/10.1007/s10463-010-0315-4>
- [6] 牛翠珍, 范国良. 基于梯度统计量的逆抽样下风险差的置信区间构建[J]. 统计与信息论坛, 2014(8): 9-13.
- [7] 白永昕, 田茂再. 慢性病发病率置信区间的构造[J]. 高校应用数学学报 A 辑, 2016, 31(2): 136-142.
- [8] 白永昕, 田茂再. Poisson 分布下基于鞍点逼近的慢性病风险差的置信区间[J]. 高校应用数学学报, 2017, 32(3): 253-266.
- [9] 钱政超, 张晨阳, 孟令宾, 田茂再. 二项抽样下基于鞍点逼近方法的流行病相对风险置信区间构造[J]. 数学的实践与认识, 2014, 44(21): 204-217.
- [10] Katz, D., Baptista, J., Azen, S.P. and Pike, M.C. (1978) Obtaining Confidence Intervals for the Risk Ratio in Cohort Studies. *Biometrics*, **34**, 469-474. <https://doi.org/10.2307/2530610>
- [11] Ann Marie, N.B., Peterson, E.D., D'Agostino, R.B., *et al.* (2015) Hyperlipidemia in Early Adulthood Increases Long-Term Risk of Coronary Heart Disease. *South China Journal of Cardiology*, **131**, 64-64.