

数学

基于效应量服从偏正态分布的 PM 法

郑萍¹ 秦超英¹ 秦思达^{2*}

(西北工业大学理学院应用数学系¹, 西安 710072; 西安交通大学医学部第一附属医院胸外2科², 西安 710061)

摘要 探讨 Meta 分析中异质性方差和偏倚的估计以及对合并效应量的影响。在 Meta 分析随机效应模型中, 假设各研究的效应量服从偏正态分布的基础上, 基于 Q 统计量及其一、二阶矩, 得到 PM 估计方程组, 采用搜索法求得方程组的解, 即为异质性方差和偏倚系数的估计, 由此进一步得到合并效应量的估计。在实例计算中, 经典的 DL 法计算的合并 OR 值为 1.370 6, 而提出方法计算结果为 1.365 3, 计算结果十分相近。基于偏正态分布的 PM 法, 不仅估计了异质性方差, 而且估计了偏倚参数, 即同时考虑了异质性和偏倚对合并效应量的影响, 使 Meta 分析模型和方法更具一般性。

关键词 Meta 分析 偏正态分布 Q 统计量 异质性方差 偏倚

中图分类号 O211.6; **文献标志码** A

在 Meta 分析随机效应模型中, 为了获得合并效应量的估计, 解决异质性方差的估计是一个关键问题^[1,2]。目前其主要方法有 DL 法^[3]、PM 法^[4]、最大似然估计法以及 BT 法^[5]等。这些方法对合并效应量的估计方法相同, 主要区别就是对异质性方差的估计方法不同。采用的模型也是相同的, 即假设纳入模型的各研究的效应量均服从正态分布。实际上, 由于偏倚的存在^[6,7], 纳入 Meta 分析的各研究的效应量并不完全服从正态分布。为了考虑偏倚的影响, 特别是将偏倚量化后纳入模型中, 本文在正态分布的基础上, 引入形状参数得到偏正态分布, 并假设纳入 Meta 分析的各研究的效应量服从偏正态分布。从而, 建立基于偏正态分布的 Meta 分析随机效应模型。在此基础上, 类比于效应量基于正态分布的 PM 法, 得到效应量服从偏正态分布的 PM 法, 即采用 PM 法对异质性方差和偏倚同时进行估计, 进而得到合并效应量的估计。实例计算表明本文提出的方法与经典的 DL 法相比计算结果相近, 但模型和方法更具一般性。

1 效应量服从正态分布的 PM 法^[4]

设 Meta 分析随机效应模型为:

$$y_i = \mu + \tau\xi_i + \varepsilon_i。$$

式中 $y_i, i = 1, \dots, k$ 是 k 个相互独立的观察效应量, k

表示纳入的研究个数; μ 为效应量; $\xi_i \sim N(0, 1)$; $\varepsilon_i \sim N(0, \sigma)$, 且 ξ_i 和 ε_i 相互独立, σ_i^2 为研究内方差, τ^2 为异质性方差。

根据上述模型, 采用马尔科夫估计法, 可得效应量 μ 的合并估计量为:

$$\hat{\mu} = \frac{\sum_{i=1}^k \omega_i y_i}{\sum_{i=1}^k \omega_i}。$$

式中 $\omega_i = \frac{1}{\sigma_i^2 + \tau^2}$ 。

PM 法是对模型中异质性方差 τ^2 进行估计的一种方法。它以 Q 统计量为依据, 通过计算 $E(Q)$, 并利用矩估计思想得到 PM 方程, 采用搜索法求解该方程, 即可得到异质性方差 τ^2 的估计值。下面具体介绍这种方法:

构造 Q 统计量, 即

$$Q = \sum_{i=1}^k \omega_i \left[y_i - \frac{\sum_{i=1}^k \omega_i y_i}{\sum_{i=1}^k \omega_i} \right]^2。$$

式中 $\omega_i = \frac{1}{\sigma_i^2 + \tau^2}$ 。

依据上述模型, 可以求出 Q 的期望为:

$$E(Q) = k - 1。$$

按照矩估计思想, 用 Q 代替 $E(Q)$, 即可得到 PM 方程:

$$F(\tau^2) = Q - (k - 1) = 0。$$

该方程的解, 即为 PM 估计法所得到的异质性

2015 年 5 月 19 日收到 国家自然科学基金(81402506)资助

* 通信作者简介: 秦思达, 男, 博士, 助理研究员。E-mail: qinsida@mail.xjtu.edu.cn。

方差 τ^2 的估计值 $\hat{\tau}^2$ 。虽然解此方程需要进行一个简单的搜索,但是 PM 法对异质性方差的估计与 DL 法相比有较好的统计学特性^[4]。

2 效应量服从偏正态分布的 PM 法

若随机变量 X 的分布密度为^[9]:

$$f_X(x) = \frac{2}{\tau} \varphi\left(\frac{x-\mu}{\tau}\right) \Phi\left(\gamma \frac{x-\mu}{\tau}\right); x \in \mathbb{R}.$$

式中 $\varphi(x)$ 和 $\Phi(x)$ 分别为标准正态分布密度函数和分布函数,则称随机变量 X 服从偏正态分布,记为 $X \sim SN(\mu, \tau^2, \gamma)$ 。显然,当 $\gamma = 0$ 时, X 退化为正态分布。当 $\mu = 0, \tau = 1$ 时,即 $X \sim SN(0, 1; \gamma)$,称为标准偏正态分布。

在上述模型的基础上,建立基于偏正态分布的随机效应模型如下:

$$y_i = \mu + \tau \xi_i + \varepsilon_i; i = 1, \dots, k.$$

式中 y_i 是对效应量 μ 的第 i 个观察效应量, k 表示纳入的研究个数, $\xi_i \sim SN(0, 1; \gamma)$, $\varepsilon_i \sim N(0, \sigma_i^2)$,且 ξ_i 和 ε_i 相互独立, γ 为表示偏倚的参数, τ^2 为异质性方差, σ_i^2 为研究内方差。根据模型还可得观察效应量 y_i 的期望和方差分别为:

$$E(y_i) = \mu + \tau \gamma \sqrt{\frac{2}{\pi}} \sqrt{\frac{1}{1 + \gamma^2}};$$

$$\text{var}(y_i) = \tau^2 \left(1 - \frac{2}{\pi} \frac{\gamma^2}{\gamma^2 + 1}\right) + \sigma_i^2.$$

若记 $m = \frac{2}{\pi} \frac{\gamma^2}{\gamma^2 + 1}$, 可见 m 亦为表征偏倚的参数,此时有

$$\text{var}(y_i) = \sigma_i^2 + (1 - m) \tau^2.$$

根据基于偏正态分布的随机效应模型,采用马尔科夫估计法,可得效应量 μ 的合并估计量为:

$$\hat{\mu} = \frac{\sum_{i=1}^k \omega_i y_i}{\sum_{i=1}^k \omega_i}.$$

式中权系数 $\omega_i = \frac{1}{\sigma_i^2 + (1 - m) \tau^2}$ 。为了利用 PM 估计法的思想得到异质性方差 τ^2 和表征偏倚的参数 m 的估计,利用该权系数 ω_i 构造 Q 统计量,即

$$Q = \sum_{i=1}^k \omega_i \left[y_i - \frac{\sum_{i=1}^k \omega_i y_i}{\sum_{i=1}^k \omega_i} \right]^2.$$

根据上述模型可以计算出统计量 Q 和 Q^2 的期望分别为:

$$E(Q) = k - 1;$$

$$E(Q^2) = k^2 - 1 + (4\pi - 4) \tau^4 m^2 \left(S_2 - \frac{S_3}{S_1} \right) + 2 \tau^4 m^2 (S_1^2 - S_2) - 4 \tau^4 m \left(S_2 - \frac{S_3}{S_1} \right).$$

式中 $S_j = \sum_{i=1}^k \omega_i^j, j = 1, 2$ 。于是,利用 PM 估计法思想可得 PM 估计方程组:

$$F(\tau^2, m) \stackrel{\Delta}{=} Q - (k - 1) = 0;$$

$$G(\tau^2, m) \stackrel{\Delta}{=} Q^2 - \left[k^2 - 1 + (4\pi - 4) \tau^4 m^2 \left(S_2 - \frac{S_3}{S_1} \right) + 2 \tau^4 m^2 (S_1^2 - S_2) - 4 \tau^4 m \left(S_2 - \frac{S_3}{S_1} \right) \right] = 0$$

采用搜索法即可求出上述方程组的一组解,它们即为异质性方差 τ^2 和偏倚 m 的估计 $\hat{\tau}^2$ 和 $\hat{m}$ 。进而可以得到合并效应量的估计:

$$\hat{\mu} = \frac{\sum \hat{\omega}_i y_i}{\sum \hat{\omega}_i}.$$

式中 $\hat{\omega}_i = 1/(\sigma_i^2 + (1 - \hat{m}) \hat{\tau}^2)$ 。

3 实例应用

肺癌的发病率和死亡率均居我国恶性肿瘤之首,并且呈普遍增长趋势,其中非小细胞肺癌(non-small cell lung cancer, NSCLC)约占肺癌患者的75%,而常规的手术及放疗化疗治疗效果有限。随着对肿瘤发生及化疗分子机制的深入研究,人们开始关注某些癌基因或蛋白质的异常表达与化疗敏感性的关系。文献[10]收集了16个临床实验结果,见表1,实验组为p53蛋白阴性患者化疗有效人数比总人数,对照组为p53蛋白阳性患者化疗有效人数比总人数,并以比值比OR的对数作为效应量,采用Meta分析中常用的DL法,探讨了NSCLC中p53蛋白质表达与铂类化疗敏感性的关系。

根据表1的实验数据,我们利用本文提出的基于效应量服从偏正态分布模型的PM法,计算出异质性方差 τ^2 、偏倚 m 或 γ 、效应量 μ 以及OR的估计值,并与文献[10]的结果进行了比较,见表2。

由表2可见,利用偏正态分布PM法得到的异质性方差为0.9312,偏倚参数 m 为0.2495或 γ 为0.8028,表明存在异质性和一定的偏倚,这与文献[10]通过检验方法得到的结论是一致的;OR值为1.3653,表明p53蛋白阴性患者化疗有效率是阳性患者的1.3653倍,而文献[10]采用DL法计算的OR值为1.3706,两种方法的计算结果十分接近,这

表明本文提出的方法是可行和有效的,两者的差异可认为是偏倚对合并 OR 值造成的影响。

表 1 NSCLC 中 p53 蛋白质表达与铂类化疗的实验结果
Table 1 The results of the expression of p53 and cisplatin based chemotherapy in NSCLC

研究	实验组	对照组	效应量 γ_i	抽样方差 $\hat{\sigma}_i^2$
1	14/24	10/28	0.924 3	0.327 0
2	9/20	5/34	1.557 2	0.436 5
3	16/48	20/75	0.318 5	0.161 9
4	6/24	10/25	-0.693 1	0.388 9
5	10/39	17/50	-0.401 4	0.223 6
6	2/8	3/7	-0.810 9	1.250 0
7	5/29	12/28	-1.280 9	0.387 5
8	16/27	4/18	1.627 5	0.474 8
9	32/56	12/46	1.329 1	0.185 7
10	6/28	6/17	-0.693 1	0.469 7
11	9/15	11/17	-0.200 7	0.535 4
12	10/18	8/20	0.628 6	0.433 3
13	9/22	16/21	-1.530 9	0.450 5
14	39/69	18/72	1.361 0	0.133 0
15	22/51	5/16	0.512 2	0.370 8
16	31/53	21/65	1.082 6	0.148 1

表 2 NSCLC 中 p53 蛋白质表达与铂类化疗敏感性关系的 meta 分析结果

Table 2 The results of Meta-analysis for the relation between the expression of p53 and the chemosensitivity to cisplatin in NSCLC

参数	$\hat{\tau}^2$	$\hat{m}$	$\hat{\gamma}$	$\hat{\mu}$	OR
正态分布 DL 法	0.645 3	—	—	0.315 2	1.370 6
偏正态分布 PM 法	0.931 2	0.249 5	0.802 8	0.311 4	1.365 3

4 结论

传统的 Meta 分析模型,假设纳入 Meta 分析的各项研究的效应量服从正态分布,仅考虑了异质性的影响,而忽略了偏倚的影响。本文建立了基于效应量服从偏正态分布的 Meta 分析模型,不仅考虑了异质性,而且同时考虑了偏倚的影响,使得 Meta 分析模型更具一般性。借助于传统的 PM 法推导思路,通过计算 Q 统计量的一、二阶矩,得到了基于效应量服从偏正态分布的 PM 方程组,解此方程便可求出异质性方差和偏倚参数的估计值,进而可以求出

合并效应量。由合并效应量的计算公式可见,异质性方差和偏倚通过权系数对合并效应量产生影响。与传统的 DL 方法相比,该方法把偏倚纳入了模型,对偏倚进行了量化处理,并考虑了偏倚对合并效应量的影响,从而使得计算结果应更接近真实情况。

参 考 文 献

- Lee K J, Thompson S G. Flexible parametric models for random-effects distributions. *Statistics in Medicine*, 2008;27(3):418—434
- 魏丽娟,董慧娟. Meta 分析中异质性的识别与处理. *第二军医大学学报*, 2006;27(4):449—450
- Wei L J, Dong J H. The identification and solution of heterogeneity in Meta-analysis. *Academic Journal of Second Military Medical University*, 2006;27(4):449—450
- Brockwell S E, Gordon I R. A simple method for inference on an overall effect in meta-analysis. *Statistics in Medicine*, 2007;26(25):4531—4543
- Der Simonian R, Kacker R. Random-effects model for meta-analysis of clinical trials: An update. *Contemporary Clinical Trials*, 2007;28(2):105—114
- Biggerstaff B J, Tweedie R L. Incorporating variability in estimates of heterogeneity in the random effects model. *Statistics in Medicine*, 1997;16(7):753—768
- 周旭毓,方积乾. Meta 分析的常见偏倚. *中国循证医学杂志*, 2002;2(4):216—220
- Zhou X M, Fang J Q. Bias in meta-analysis. *The Journal of Evidence-Based Medicine*, 2002;2(4):216—220
- 董碧蓉,马春华. 偏倚对系统评价质量的影响. *中国临床康复*, 2003;7(3):368—369
- Dong B R, Ma C H. How does the bias influence the quality of systematic reviews. *Chinese Journal of Clinical Rehabilitation*, 2003;7(3):368—369
- Viechtbauer W. Confidence intervals for the amount of heterogeneity in meta-analysis. *Statistics in Medicine*, 2007;26(1):37—52
- Arellano-Valle R B, Gomez H W, Quintana F A. A new class of skew-normal distributions. *Communications in Statistics Theory and Methods*, 2004;33(7):1465—1480.
- 汪进良,焦顺昌,叶平,等. 非小细胞肺癌患者 p53 蛋白异常表达与铂类化疗敏感性的 Meta 分析. *南方医科大学学报*, 2008;28(5):770—773
- Wang J L, Jiao S H, Ye P, et al. p53 protein expression and chemosensitivity to cisplatin in patients with non-small cell lung cancer: a Meta-analysis. *Journal of Southern Medical University*, 2008;28(5):770—773

(下转第 125 页)

souri, 2011

Numerical Study on the Ballistic Resistance Property of Aerospace Aluminum Alloy Targets Against Bird Strike

JIA Bao-hui, LI Jian-feng^{*}, ZHANG Feng, DENG Yun-fei

(Civil Aviation University of China, College of Aeronautical Engineering, Tianjin 300300, P. R. China)

[**Abstract**] Projectiles with different size and targets with different thickness have been established with the help of ANSYS-AUTODYN, then the process of bird striking aircraft have been simulated and the validation of simulation models has been proved by experiment. Based on the simulation results, the damage of targets at each circumstances have been analyzed and then the influence of targets' material parameters, bird-projectiles' speed and quality on the damage property of targets have been obtained. The simulation results indicated that target's material parameters would affect it's damage property, if the yield stress and tensile stress become stronger and stronger, the deformation of target and height of bulge would corresponding become insignificantly, Furthermore, the speed and quality of bird-projectiles can also influence the damage of targets effectively, with the increase of bird-projectiles' speed and quality, the height of targets' bulge would become higher and higher, until the target is tore by projectiles.

[**Key words**] bird-strike damage property ability of ontistrike numerical simulation

(上接第 119 页)

The PM Method in Meta-analysis Based on the Effects Obeying Skew Normal Distribution

ZHENG Ping¹, QIN Chao-ying¹, QIN Si-da^{2*}(Applied Mathematics, College of Science, Northwestern Polytechnical University¹, Xi'an 710072, P. R. China;Thoracic Surgery Department Two, The First Affiliated Hospital, School of Medicine², Xi'an Jiaotong University, Xi'an 710061, P. R. China)

[**Abstract**] To discuss the estimations of heterogeneity variance and bias and the influence they have on the overall effect in meta-analysis. In the random-effect model in meta-analysis, based on the assumption it is assumed that the effects in various researches obey skew normal distribution. The PM estimated equations can be obtained with Q -statistic and its first and second order moments. Then the solutions of the estimated equations, which are the estimations of heterogeneity variance and bias coefficient, are also can be obtained by using iteration method. After that we can get the estimation of overall effect. In practical calculations, the combined OR calculated by the classic DL method is 1.370 6, while using the proposed method, the calculation result is 1.365 3, which is very similar to the result obtained by the classic DL method. Based on the skew normal distribution, the PM method estimates both the heterogeneity variance and the bias coefficient. Namely, due to taking into account the influence of heterogeneity and bias of the overall effect, the model and method of meta-analysis becomes more general.

[**Key words**] meta-analysis skewed normal distribution Q -statistic heterogeneity bias overall effect PM method