- in mid to late Parkinson's disease[J]. Mov Disord, 2014, 29 (10); 1273–1280.
- [14] STOCCHI F, BORGOHAIN R, ONOFRJ M, et al. A randomized, doubleblind, placebo-controlled trial of safinamide as add-on therapy in early Parkinson's disease patients[J]. Mov Disord,

2012, 27(1):106-112.

[15] KRÖSSER S, MARQUET A, GALLEMANN D, et al. Effects of ketoconazole treatment on the pharmacokinetics of safinamide and its plasma metabolites in healthy adult subjects[J]. Biopharm Drug Dispos, 2012, 33(9): 550-559.

「文章编号]1007-7669(2016)04-0255-05

[DOI号]10.14109/j.cnki.xyylc.2016.04.007

临床试验中率差及其置信区间的估计方法

韩景静、曾 新、王 骏

(国家食品药品监督管理总局 药品审评中心, 北京 100038)

「关键词 临床试验;统计学;置信区间

[摘要] 目的 探讨临床试验二分类数据率差及其置信区间的估计方法。方法 分别介绍不存在协变量、只存在一个协变量、存在多个协变量三种情况时,率差及置信区间的估计方法,然后利用实例进行说明。结果 当不存在协变量时,可用两组反应率估计值进行计算;当只存在一个分层因素时,针对不同情况可选择 CMH、逆方差、最小风险加权方法进行估计;当存在多个协变量时,可结合 logistic 回归模型和 delta 方法、bootstrap 方法计算。结论 临床试验二分类数据率差及其置信区间的准确估计,需要结合试验设计,考虑是否需要校正协变量以及协变量的数量进而选择不同的统计方法。

[中图分类号] R195 [文献标志码] A

Methods of estimating rate difference and associated confidence interval in clinical trials

HAN Jing- jing, ZENG Xin, WANG Jun (Center for Drug Evaluation, China Food and Drug Administration, BEIJING 100038, China)

[KEY WORDS] clinical trial; statistics; confidence interval

[ABSTRACT] AIM To explore methods of estimating rate difference (RD) and associated confidence interval (CI) for binary data in clinical trials. METHODS The methods to estimate RD and its CI in three scenarios were introduced, with no covariate, one covariate, and more than one covariate in the statistical model, respectively. An example was provided for further illustration. RESULTS When there was no covariate, the estimated response rates of the two groups were used to calculate RD and its CI. When there was only one covariate, CMH, inverse of variance or minimum risk weighting methods was applied. The delta or bootstrap

[收稿日期] 2015-12-28 [接受日期] 2016-03-08

[作者简介] 韩景静,女,硕士研究生,主要研究统计学方法在临床试验中的应用,Phn: 86-10-6892-1487, E-mail: hanjj@cde.org.cn; 王 骏,男,博士,副研究员,主要从事药品的技术审评工作,Phn: 86-10-6892-1484, E-mail: wangj@cde.org.cn

[责任作者] 王 骏

method combined with the logistic regression was used in case of more than one covariate. CONCLUSION It is necessary to choose appropriate method to calculate RD and its CI for binary data considering whether to adjust for covariates and the possible number of covariates according to the trial design in clinical trials.

临床试验中经常可以看到成组设计结果为二 分类的数据,如有效或无效等。对于该类型资料, 通常需要估计组间率差及其置信区间,用以判断 试验结果是否达到等效、非劣效或优效等。在实 际应用分析时,率差及其置信区间的估计应结合 试验设计进行考虑, 如在多中心临床试验的首要 分析 (primary analysis) 中, 当不存在交互作用 时,可能需要考虑中心因素的校正,又如在随机 化时将年龄作为分层因素,在估计率差及其置信 区间时还需要校正年龄这一协变量。然而目前国 内注册临床试验 (含药物和器械) 普遍缺乏报告 考虑中心效应、基线等因素后的置信区间、这样 的结果并不准确,往往会影响统计学结论的判断。 因此,可以将率差及其置信区间的估计分为三类 情况:不存在协变量、只存在一个协变量因素、 存在多个协变量因素。

当不存在协变量时,估计率差置信区间的方 法很多。NEWCOMBE^[1]在 1998 年的一篇文章中就 对估计率差置信区间的 11 种方法进行了比较。 BROWN 等回也通过大量的模拟数据,总结了6种 率差置信区间在不同情况下的优劣。如果使用 SAS 中的 proc freq 过程,可以快速得到两组率差及 Agresti - Caffo, exact unconditional, Hauck -Anderson, Miettinen-Nurminen (score), Newcombe, Wald 6 种率差置信区间。目前, Newcombe 方法已 被美国临床实验室标准化协会 (CLSI) EP12-A2 文件以及 FDA 指南推荐使用, 作为有效率可能为 0%或 100%率差置信区间估计方法的首选[3,4]。当 只存在一个协变量因素时,需要对协变量进行校 正, 通常使用 Cochran - Mantel - Haenszel (CMH) 方法, 如果使用 SAS 频数过程的 CMH 选项只能直 接得到调整协变量后两组有效率比值(RR)及其 置信区间,但是无法得到校正的率差及其置信区 间。当存在多个协变量因素时,更是无法直接对 协变量进行调整得到校正率差及其置信区间,当 然 SAS 中也没有对应的过程步骤。因此,本研究 主要介绍存在协变量时两组率差及其置信区间的 常用的估计方法,以期为统计分析人员能够正确 分析该类资料提供清晰的思路和解决方案。

只校正一个因素 定义 p_{ij} 为第 i 组 (i = 1, 2) 第 j

层(j=1, …, s)的有效率,那么第j 层的率差 $d_j=p_{1j}-p_{2j}$ 。将第j 层看作如下 2×2 列联表,那么该层每组有效率的估计 $\hat{p}_{ij}=x_{ij}/n_{ij}$,有效率差值的估计值为 $\hat{d}_j=\hat{p}_{1j}-\hat{p}_{2j}=x_{1j}/n_{1j}-x_{2j}/n_{2j}$ 。而校正中心因素后率差的点估计则是所有层率差的加权平均值,即, $d_{aaj}=\sum_{j=1}^s w_j d_j$ [5],其中, w_j 为每层的权重。常见有 CMH、逆方差(inverse of variance)和最小风险(minimum risk)三种计算权重方法。

表 1 2×2 列联表 (第j层)

	V J 7	
层 <i>j</i>	第1组	第 2 组
有效	x_{1j}	x_{2j}
无效	n_{ij} - x_{ij}	n_{2j} - x_{2j}
合计	n_{1j}	n_{2j}

1 CMH 方法

$$w_{j} = \frac{\frac{n_{1j} n_{2j}}{n_{1j} + n_{2j}}}{\sum_{j=1}^{s} \frac{n_{1j} n_{2j}}{n_{1j} + n_{2j}}}$$
(1)

利用式 (1) 这种方法,人数越多的层分配的权重越大。但是,CMH 权重没有考虑每层率差估计值的变异程度。

2 逆方差法

$$w_{j} = \frac{\left(\frac{p_{1j}(1-p_{1j})}{n_{1j}} + \frac{p_{2j}(1-p_{2j})}{n_{2j}}\right)^{-1}}{\sum_{j=1}^{s} \left(\frac{p_{1j}(1-p_{1j})}{n_{1j}} + \frac{p_{2j}(1-p_{2j})}{n_{2j}}\right)^{-1}}$$
(2)

式 (2) 则考虑了层间的异质性,以率差的方差估计值倒数作为权重,使得 \hat{d}_j 越小的层分配的权重越大。

3 最小风险法

$$w_{j} = \frac{\beta_{j}}{\sum_{k=1}^{s} V(\hat{d}_{k})^{-1}} - \left[\frac{\alpha_{j} V(\hat{d}_{j})^{-1}}{\sum_{k=1}^{s} V(\hat{d}_{k})^{-1} + \sum_{k=1}^{s} \alpha_{k} d_{k} V(\hat{d}_{k})^{-1}} \right] \times \left[\frac{\sum_{k=1}^{s} d_{k} \beta_{k}}{\sum_{k=1}^{s} V(\hat{d}_{k})^{-1}} \right]$$
(3)

其中,
$$\alpha_j = d_j \sum_{k=1}^s V(\hat{d}_k)^{-1} - \sum_{k=1}^s d_k V(\hat{d}_k)^{-1}$$
 (4)

$$\beta_{j} = V(\hat{d}_{j})^{-1} \left(1 + \alpha_{j} \sum_{k=1}^{s} \frac{d_{k}(n_{1k} + n_{2k})}{\sum_{m=1}^{s} (n_{1m} + n_{2m})} \right)$$
 (5)

$$V(\hat{d}_{j}) = \frac{\hat{p}_{1j} (1 - \hat{p}_{1j})}{n_{1j}} + \frac{\hat{p}_{2j} (1 - \hat{p}_{2j})}{n_{2j}}$$
(6)

利用式(3)计算的权重由均方误差损失(the average squared error loss,or risk)函数 $Risk = V(\hat{d}_{adj}) + (E(\hat{d}_{adj}-d))^2$ 得到,能最小化平方误差损失(the squared error loss)[6]。值得一提的是,如果每层的率差 d_j 是常数,那么 $\alpha_j = 0$,最小风险权重等于逆方差权重。

RADHAKRISHNA^[7]指出,如果每层的优势比 p_{1j} (1- p_{2j}) $/p_{2j}$ (1- p_{1j}) 是一个常数,那么 CMH 权 重是最佳的选择;如果每层的率差是一个常数,则逆方差权重是最合适的。但是,如果我们不能 获得上述的先验信息,最好选择最小风险权重。

然后,相应地结合上述三种权重可利用 Wald 方法和 Newcombe 方法估计校正率差的置信区间。 Wald 方法:

$$\hat{d}_{\text{adj}} \pm z_{\alpha/2} \sqrt{\sum_{j=1}^{s} \hat{w}_{j} \left(\frac{\hat{p}_{1j} (1 - \hat{p}_{1j})}{n_{1j}} + \frac{\hat{p}_{2j} (1 - \hat{p}_{2j})}{n_{2j}} \right)}$$
(7)

Newcombe 方法:

$$\begin{split} \hat{d}_{\text{adj}} - & z_{\text{a/2}} \sqrt{\sum_{j=1}^{s} \frac{\hat{w}_{j}^{2}}{n_{\text{1}j}} L_{1} \left(1 - L_{1}\right) + \sum_{j=1}^{s} \frac{\hat{w}_{j}^{2}}{n_{\text{2}j}} U_{2} \left(1 - U_{2}\right)} \ , \\ \hat{d}_{\text{adj}} + & z_{\text{a/2}} \sqrt{\sum_{j=1}^{s} \frac{\hat{w}_{j}^{2}}{n_{\text{2}j}} L_{2} \left(1 - L_{2}\right) + \sum_{j=1}^{s} \frac{\hat{w}_{j}^{2}}{n_{\text{1}j}} U_{1} \left(1 - U_{1}\right)} \end{split} \tag{8}$$

式 (8) 中, $(L_i,\ U_i)$ 是 $\hat{p}_i=\sum_{j=1}^s w_j\hat{p}_{ij}$ 的 Wilson 置信区间,其计算公式为

$$\begin{split} &(L_{i} \,,\;\; U_{i} \,) \,=\; \sum_{j\,=\,1}^{s} \hat{w}_{j} \\ &\left[\frac{\hat{p}_{ij} + z_{\,\,ni}^{2} / 2n_{j}}{1 \,+\, z_{\,\,\gamma i}^{2} / n_{j}} \,\,\pm\, \frac{z_{\,\gamma i}}{1 \,+\, z_{\,\,\gamma i}^{2} / n_{j}} \,\,\,\,\,\,\,\, \sqrt{\frac{\hat{p}_{ij} \,\left(\,1 - \hat{p}_{ij}\,\right)}{n_{j}} \,+\, \frac{z_{\,\,\gamma i}^{2}}{4n_{\,\,j}^{2}}} \,\, \end{split} \right. \end{split}$$

$$z_{\gamma i} = z_{\alpha/2} \sqrt{\frac{\sum_{j=1}^{s} \hat{w}_{j}^{2} \hat{p}_{ij} (1 - \hat{p}_{ij}) / n_{j}}{\sum_{j=1}^{s} \hat{w}_{j} \sqrt{\hat{p}_{ij} (1 - \hat{p}_{ij}) / n_{j}}}}$$
(10)

校正多个协变量 当有多个协变量时,用 logistic 回归模型预测每个受试者的有效率、率差的点估

计即为两组有效率均值的差值^[8]。然后再利用 delta 方法^[9,10]或 bootstrap 方法^[11,12]计算校正率差的标准误,并进而得到置信区间。首先介绍 delta 方法和 bootstrap 方法。

1 delta 方法 如果一个试验有 n 个受试者,令二分类反应向量 $Y = (y_1, y_2, \dots, y_n)'$,协变量矩阵 $X = (x_1, x_2, \dots, x_n)$,那么 logistic 回归模型 logit [$P(y_i = 1 \mid x_i)$] = βx_i ,这里 logit (p) = ln [p/ (1-p)]。然后假设 b 是 β 的最大似然估计,它的协方差矩阵为 V。

我们要估计试验组和对照组的有效率以及两组率差。首先,将所有的受试者分配到试验组,建立协变量矩阵 X_{to} 然后,由 X_t 和 b [$\hat{p}_t = logit^{-1}$ (X_t b)] 计算得到有效率 \hat{p}_{to} 相似的,将所有受试者分配到对照组,然后重复上述步骤得到 X_c 和 \hat{p}_{co} 率差的估计值即为 $d = \sum_i (\hat{p}_u - \hat{p}_{ci}) /n$,其中 \hat{p}_{ti} 和 \hat{p}_{ci} 分别是 \hat{p}_t 和 \hat{p}_e 中第 i 个元素。定义矩阵 A_t 的元素 $A_{ti} = \hat{p}_{ti}$ ($1 - \hat{p}_{ti}$),类似地, A_c 的元素 $A_{ci} = \hat{p}_{ci}$ ($1 - \hat{p}_{ci}$)。那么利用 delta 方法估计率差标准误的公式为:

$$d_{t} = (A_{t}X_{t}) / n \tag{11}$$

$$d_{c} = (A_{c}X_{c})/n \tag{12}$$

$$SE(d) = \sqrt{d_t V d'_t + d_c V d'_c - 2d_c V d'_t}$$
(13)

2 bootstrap 方法 bootstrap 方法是EFRON^[13]在 20 世纪 70 年代后期建立的,是非参数统计中一种重要的估计统计量方差进而进行区间估计的统计方法,其核心思想是:在原始样本中(样本量为n)做有放回的抽样,原始数据中每个观察单位每次被抽到的概率是 1/n,样本含量仍为n。所得新的样本被称为 bootstrap 样本。于是,每个 bootstrap 样本可以得到参数的一个估计值,这样重复 B 次。参数的 bootstrap 估计就是基于这些 bootstrap 样本的统计量。例如,统计量的标准误的估计 SE (θ) =

$$\sqrt{\left[\frac{1}{B-1}\sum_{1}^{B}\left(\theta^{B}-\bar{\theta}^{B}\right)^{2}\right]} , \ \ \vec{\Xi} \ \dot{\Phi} \ \ \vec{\theta}^{B} = \frac{1}{B}\sum \theta^{B}_{\circ}$$

本研究中,首先从原始样本中有放回地抽取样本 (例如在 SAS 中可利用 surveyselect 过程实现),抽取的样本量等于原始样本量,然后利用 logistic 回归分析抽出的样本计算校正率差。重复上述过程 1 000 次,得到 1 000 个校正率差,最后利用 1 000 个校正率差计算其标准误。

实例分析 某多中心、随机开放、阳性药物平行

对照试验,采取以 2:1 的比例、考虑中心因素、性别分层以及基因型分层的随机方式,将 770 例受试者随机分配到试验组和对照组。通过两组病毒学应答率的率差及 95%置信区间评价试验药是否非劣效于对照药,非劣效界值为 12%。每层人数和应答人数在两组中的分布如表 2 所示。

表 2 每层人数及应答人数分布

中心	性别	基因型	试验组	对照组
1	男	A	$20 \ (n = 22)$	14 (n = 16)
	男	В	54 (n = 79)	27 $(n = 36)$
	女	A	19 $(n = 24)$	8 (n = 10)
	女	В	$50 \ (n = 73)$	22 $(n = 33)$
2	男	A	1 (n = 1)	2(n = 3)
	男	В	3 (n = 3)	0 (n = 0)
	女	В	1 (n = 1)	0 (n = 1)
3	男	A	$20 \ (n = 24)$	8 (n = 9)
	男	В	40 $(n = 56)$	26 (n = 33)
	女	A	15 $(n = 16)$	8 (n = 11)
	女	В	48 $(n = 68)$	24 $(n = 36)$
4	男	A	31 $(n = 36)$	12 (n = 16)
	男	В	21 $(n = 35)$	17 $(n = 20)$
	女	A	29 $(n = 35)$	14 $(n = 18)$
	女	В	28 $(n = 36)$	$16 \ (n = 19)$
合计			380 $(n = 509)$	198 (n = 261)

首先,不对性别、基因型、中心三个分层因素进行校正,利用 SAS 中频数过程直接估计,两组应答率率差(试验组-对照组)及 95%置信区间分别为-0.012 (-0.076, 0.052); 然后,分别将分层因素性别、基因型或中心单独作为协变量,计算每层的三种权重,利用三种权重求得的校正率差及其 Wald 置信区间和 Newcombe 置信区间见表3; 最后,将性别、基因型和中心同时作为协变量纳入分析,利用 logistic 回归模型得到校正率差,然后再利用 delta 方法或 bootstrap 方法计算校正率差的标准误,最后得到率差及其置信区间的准确估计,结果见表 4。为保证准确性,针对 delta 方法和 bootstrap 方法,参考了文献 [14-17],编写了SAS 程序。表 4 是运行 SAS 程序后的结果。其中,bootstrap 方法进行了 1 000 次重复抽样。

由表 3 和 4 可知,当只对性别分层因素进行校正时,CMH 权重校正的率差恰好等于未校正率差,而用其他两种权重校正的率差和未校正率差相比均变大,但是差异不大;当只对中心因素进行校正时,最小风险权重校正率差恰好等于未校正率差,其他两种权重校正的率差和未校正率差相比都变小,差异依然不大;相比之下,当只对基因型分层因素进行校正时,利用三种权重得到

表 3 校正率差及两种置信区间

校正因素	权重方法	率差	Wald 置信区间	Newcombe 置信区间
性别	CMH	-0.012	(-0.076, 0.052)	(-0.074, 0.054)
	逆方差	-0.015	(-0.079, 0.049)	(-0.077, 0.051)
	最小风险	-0.013	(-0.077, 0.051)	(-0.075, 0.053)
基因型	CMH	-0.011	(-0.075, 0.053)	(-0.073, 0.055)
	逆方差	-0.005	(-0.068, 0.059)	(-0.067, 0.062)
	最小风险	-0.009	(-0.073, 0.054)	(-0.071, 0.051)
中心	CMH	-0.011	(-0.075, 0.052)	(-0.074, 0.054)
	逆方差	-0.010	(-0.074, 0.053)	(-0.073, 0.056)
	最小风险	-0.012	(-0.076, 0.052)	(-0.074, 0.054)

表 4 同时校正性别、基因型、中心后的估计结果

方法	率差	标准误	Wald 置信区间
delta	-0.011	0.032	(-0.074, 0.052)
bootstrap	-0.011	0.033	(-0.076, 0.054)

的校正率差相对未校正率差均明显变小。当同时校正三个因素时,两种方法得到的校正率差均为-0.011,小于未校正率差。此外,将三个因素同时纳入 logistic 回归模型中进行分析时,只有基因型分层因素的回归系数有显著差异 (P < 0.05)。

由分析结果可知,各种情况下,两组率差的 95%置信区间下限都大于-12%,试验组非劣效于 对照组。

讨 论

综上,针对二分类数据作为临床试验主要疗 效指标,对两组率差及其置信区间进行估计时, 需要结合试验设计,考虑是否需要校正协变量以 及协变量的数量,并在试验方案和统计分析计划 (SAP) 中预先确定选用的统计方法。当不存在协 变量时、率差的估计即为两组反应率估计值的差 值, 其置信区间的估计有 Wald 方法、Newcombe 方法等; 当只校正一个分层因素时, 针对不同情 况可选择 CMH、逆方差、最小风险等加权方法, 进而计算加权率差及其相应置信区间; 当存在多 个协变量时,可利用 logistic 回归模型对协变量进 行校正后得到校正率差, 然后再利用 delta 方法或 bootstrap 方法计算校正率差的标准误,最后得到率 差及其置信区间的准确估计。目前尚未研究对 delta 和 bootstrap 两种方法的适用情况进行比较, 因此建议选择其中一种方法作为主要分析,另一 种方法作为敏感性分析。此外,估计置信区间的 Wald 方法和 Newcombe 方法的比较也值得进一步 研究。

[参考文献]

- NEWCOMBE RG. Interval estimation for the difference between independent proportions: comparison of eleven methods[J]. Stat Med, 1998, 17(8): 873–890.
- [2] BROWN L, LI XF. Confidence intervals for two sample binomial distribution[J]. J Stat Planning Inference, 2005, 130 (1-2): 359-375.
- [3] Clinical and Laboratory Standards Institute. User protocol for evaluation of qualitative test performance, approved guideline-second edition[EB/OL]. (2008-01-25)[2015-11-13]. http://shop.clsi.org/method-evaluation-documents/EP12.html.
- [4] FDA. Guidance for industry and FDA staff-statistical guidance on reporting results from studies evaluating diagnostic tests[EB/ OL] (2007-03-13)[2015-11-13]. http://www.fda.gov/medicaldevices/ deviceregulationandguidance/guidancedocuments/ucm071148.htm
- [5] KIM Y, WON S. Adjusted proportion difference and confidence interval in stratified randomized trials[R]. Chicago, 2013.
- [6] MEHROTRA D, RAILKAR R. Minimum risk weights for comparing treatments in stratified binomial trials[J]. Stat Med, 2000, 19(6): 811-825.
- [7] RADHAKRISHNA S. Combination of results from several 2 x 2 contingency tables[J]. Biometrics, 1965, 21: 86–98.
- [8] KLEINMAN LC, NORTON EC. What's the risk? A simple approach for estimating adjusted risk measures from no linear models including logistic regression[J]. Health Res Educational Trust, 2009, 1(44): 288–302.

- [9] GE MM, DURHAM LK, MEYER RD, et al. Covariate-adjusted difference in proportions from clinical trials using logistic regression and weighted risk differences[J]. Drug Inf J, 2011, 45 (4): 481-493.
- [10] PharmaSUG. An illustration of rate difference estimation with SAS in logistic regression[EB/OL]. (2012–06–07)[2015–06–05] http://www.lexjansen.com/pharmasug-cn/2012/papers/Pharma SUG_China_2012_d3.pdf.
- [11] 盛 骤,谢式千,潘承毅. 概率论与数理统计[M]. 北京:高等教育出版社,2008:270-281.
- [12] PhUSE. A practical introduction to the bootstrap using the SAS system[EB/OL]. (2005-10-12)[2015-10-28].http://www.lexjansen.com/phuse/2005/pk/pk02.pdf.
- [13] EFRON B. Bootstrap methods: another look at the jackknife[J]. Ann Stat, 1979, 7(I): 1-26.
- [14] GULICK RM, LALEZARI J, GOODRICH J, et al. Maraviroc for previously treated patients with R5 HIV - 1 infection [J]. N Engl J Med, 2008, 359(14): 1429–1441.
- [15] SAS Institute Inc. SAS/STAT ® 9.2 User's Guide [K].Second Edition. Cary, North Carolina, USA: SAS Institute Inc, 2009.
- [16] SAS Institute Inc. SAS/IML[®] 9.2 User's Guide [K]. Second Edition. Cary, North Carolina, USA: SAS Institute Inc, 2009.
- [17] UCLA. How can I estimate the standard error of transformed regression parameters in R using the delta method? [EB/OL]. (2014-08-01) [2015-05-25]. http://www.ats.ucla.edu/stat/r/faq/ deltamethod.html.

作者声明 本人喻明 (第一作者) 在《中国新药与临床杂志》2015 年第 8 期 617-621 页发表的 "中药止痛贴联合吗啡治疗中重度癌痛的多中心临床研究"一文中,署名单位顺序应为 "1.辽宁中医药大学,辽宁沈阳 110032; 2.辽宁中医药大学附属 医院,辽宁沈阳 110032"。