

## · 基础理论与方法 ·

## 多临床终点的等效性检验方法探讨

闫宇翔 王洪源 王仁安

**【摘要】** 目的 探讨多临床终点的等效性检验方法。方法 应用改进的 *Hotelling T<sup>2</sup>* 检验对两组均数向量进行等效性检验,通过扣除变量间相关性的方法获得各变量是否等效的具体信息,并结合实例进行分析,以  $\pm 2\%$  为等效界值的上下界,对“ A、B 两药治疗绝经后骨质疏松症 ”的四个疗效指标分别进行单变量和多变量等效性检验,比较其结果。结果 当 *Hotelling T<sup>2</sup>* 检验的统计量对应的  $P > \alpha$  时,可认为所有变量不等效,若  $P \leq \alpha$  时,应进行不等效变量的筛选,经筛选无不等效变量可认为其等效,若经筛选有不等效变量,则认为该变量不等效,该方法克服了重复单变量等效性检验的缺点。结论 应用改进的 *Hotelling T<sup>2</sup>* 检验并结合不等效变量的筛选方法,可以对多个临床终点同时进行等效性检验,该方法合理且可操作。

**【关键词】** 临床等效性;多变量等效性检验;相关性

**Statistical test on multivariate in therapeutical equivalence analysis** YAN Yu-xiang\*, WANG Hong-yuan, WANG Ren-an. \*Department of Epidemiology and Biostatistics, Capital University of Medical Science, Beijing 100054, China

**【Abstract】 Objective** To explore the way for testing of multivariate in therapeutical equivalence. **Methods** By improved *Hotelling T<sup>2</sup>* test, the mean vector of two groups was tested. The information of each variable could then be obtained by deducting the their correlations. “Clinical Equivalence of Drug A and Drug B for Postmenopausal Bone Looseness” was used in this research. The primary measure of efficacy was regarded as the improvement rate of four variables with predefined limit of equivalence (2%) with the results of univariate test and multivariate test compared. **Results** Under  $P > \alpha$ , all the variables were nonequivalent. Under  $P \leq \alpha$  it was necessary to filtrate the variables of unequivalent until the residuals were equivalent. The method of multivariate test was a remedy to the shortage of univariate test. **Conclusion** With the combination of the filtration of nonequivalent variables and the improved *Hotelling T<sup>2</sup>* test, the test of multivariate could be done. The method is reasonable and practical.

**【Key words】** Therapeutic equivalence; Multivariate test of equivalence; Correlation

各个疾病的反应变量在大多数情况下不止一个,因而需要同时将其作为疗效观察和评价的指标,即临床终点。在等效性分析时,对各临床终点的组间差别分别进行单变量检验的处理方法至少有三个缺点:当指标个数较多时,重复进行 *t* 检验将大大增加假阳性错误; *t* 检验结果不一致时,统计结果难以表达;不能察觉多个反应变量之间的相互关系。克服上述缺点的方法是计算多元统计量<sup>[1]</sup>,并进行多变量等效性检验。我们根据等效性检验原理,提出了均数向量等效性检验方法,并进一步探讨了不等效变量筛选的原理和方法。

## 原理与方法

## 1. 等效性检验原理:等效性评价是以阳性对照

作者单位:100054 北京,首都医科大学流行病学与卫生统计学系(闫宇翔),北京大学医学部流行病学与卫生统计学系(王洪源、王仁安)

药取代安慰剂做临床试验的一种新药疗效评价方法,即判断新药与已有的有效药物疗效相比是否相当。等效性检验是事先根据临床终点给出等效临界值,若两组参数差值落在此范围内则认为等效<sup>[2,1]</sup>,即  $|\mu_T - \mu_R| < \Delta$  (其中  $\mu_T$  与  $\mu_R$  分别代表受试药 T 与参比药 R 的总体均数,  $\Delta$  为等效界值)。其检验假设为:

$$H_0: \mu_T - \mu_R \leq -\Delta \text{ 或 } \mu_T - \mu_R \geq \Delta$$

$$H_1: -\Delta < \mu_T - \mu_R < \Delta$$

若经两次单侧 *t* 检验均有统计学意义 ( $P < \alpha$ ) 就认为两药等效,两组均数和率的等效性检验公式分别为:

$$t_1 = \frac{\Delta - (\bar{X}_T - \bar{X}_R)}{S_{\bar{X}_T - \bar{X}_R}} \quad t_2 = \frac{(\bar{X}_T - \bar{X}_R) - (-\Delta)}{S_{\bar{X}_T - \bar{X}_R}}$$

$$u_1 = \frac{\Delta - (P_T - P_R)}{\sqrt{P(1-P)(1/n_T - 1/n_R)}}$$

$$u_2 = \frac{(P_T - P_R) - (-\Delta)}{\sqrt{P(1-P)(1/n_T - 1/n_R)}}$$

2. 两组均数向量的等效性检验 :根据双单侧  $t$  检验原理,对两组均数向量无差异性检验<sup>[1]</sup> (Hotelling  $T^2$  检验)进行了改进,改进后的  $T^2$  统计量为:

$$T^2 = \frac{n_T n_R}{n_T + n_R} (\Delta - |\bar{X}_T - \bar{X}_R|) S_c^{-1} (\Delta - |\bar{X}_T - \bar{X}_R|)$$

其中  $\bar{X}_T - \bar{X}_R$  代表试验药 T 与参比药 R 两组均数向量之差  $\Delta$  为变量等效界值向量,  $S_c$  为两样本协方差矩阵  $S_T$  和  $S_R$  的合并估计矩阵。

(1) 检验假设<sup>[3]</sup>:

$$H_0 : |\mu^{(i)}| \geq \Delta_i \text{ for all } i$$

$$H_A : |\mu^{(i)}| \geq \Delta_i \text{ for some } i$$

单侧  $\alpha = 0.05$

其中  $\mu_i$  为两总体均数差值向量 ( $\mu_T - \mu_R$ ) 的第  $i$  个元素,  $\Delta$  为变量的等效界值向量 ( $\Delta_1, \Delta_2, \dots, \Delta_p$ )

(2) 检验公式:

$$T_1^2 = \frac{n_T n_R}{n_T + n_R} [\Delta - (\bar{X}_1 - \bar{X}_2)] S_c^{-1} [\Delta - (\bar{X}_1 - \bar{X}_2)]$$

$$T_2^2 = \frac{n_T n_R}{n_T + n_R} [(\bar{X}_1 - \bar{X}_2) - (-\Delta)] S_c^{-1} [(\bar{X}_1 - \bar{X}_2) - (-\Delta)]$$

$$F = \frac{(n_T + n_R - p - 1)}{(n_T + n_R - 2)p} T^2, v_1 = p, v_2 = n_T + n_R - p - 1$$

当  $n_T + n_R$  较大时,  $F$  值近似地服从自由度为  $p$  的  $\chi^2$  分布。若  $P > \alpha$ , 不能拒绝  $H_0$ , 认为两药在所有指标上不等效; 若  $P \leq \alpha$ , 拒绝  $H_0$ , 接受  $H_1$ , 则说明  $p$  个指标中至少有一个是等效的, 还需要用下面的方法进行不等效变量的筛选。

3. 不等效变量的筛选: 上述方法的分析结果提供了两样本均数向量间是否等效的总的信息(即两组间只要有一个变量等效, 那么均数向量就是等效的), 但尚未提供各变量间是否等效的具体信息, 即尚未指出哪几个变量之间等效或不等效。为得到这方面的信息, 可进行不等效变量的筛选, 其原理是把变量间的相关性扣除<sup>[4]</sup>。

两样本混合后的总离差阵和组间离差阵分别为:

$$T = \sum_{r=1}^k \sum_{i=1}^{n_r} (X_i^{(r)} - \bar{X})(X_i^{(r)} - \bar{X})$$

$$B = \sum_{r=1}^k n_r (\bar{X}^{(r)} - \bar{X})(\bar{X}^{(r)} - \bar{X})$$

组内离差阵  $W = W_1 + W_2 + \dots + W_k$ , 每个母体  $r$  的离差阵为:

$$W_r = \sum_{i=1}^{n_r} (x_i^{(r)} - \bar{x}^{(r)})(x_i^{(r)} - \bar{x}^{(r)}), r = 1, 2, \dots, k$$

三个离差阵之间永远有:  $T = B + W$

为了扣除变量间的相关性, 将  $T, W, S_C$  分别求逆, 对应的逆矩阵对角线上第  $i$  个元素分别为

$T^{ii}, W^{ii}, S_C^{ii}$ 。扣除相关性后的组间离差阵:

$$SB = 1/T^{-1} - 1/W^{-1}$$

$$SB_{ii} = 1/T^{ii} - 1/W^{ii}$$

因组间自由度  $\nu = 1$ , 组间协方差阵与组间离均差阵相同, 称  $\sqrt{SB_{ii}}$  为条件组间均数差。

若  $E(\sqrt{SB_{ii}})$  落在等效范围内, 则认为几种处理是等效的, 反之, 认为它们是不等效的。构造检验统计量  $t$  (若两均数差值为负, 则在  $\sqrt{S_i}$  前加负号):

$$t_1 = \frac{\Delta - \sqrt{SB_{ii}}}{\sqrt{\left[\frac{1}{S_C^{ii}}\right] \left[\frac{1}{n_T} + \frac{1}{n_R}\right]}} \quad t_2 = \frac{\sqrt{SB_{ii}} - (-\Delta)}{\sqrt{\left[\frac{1}{S_C^{ii}}\right] \left[\frac{1}{n_T} + \frac{1}{n_R}\right]}}$$

如果经两个单侧  $t$  检验有统计学意义 ( $P \leq 0.05$ ), 就认为在第  $i$  个变量上等效。

### 实例分析

为了进一步比较单变量和多变量等效性检验方法的优缺点及检验结果的差异, 并评价后者的合理性与可操作性, 采用“口服 A (试验药), B (阳性对照药) 两药对治疗绝经后妇女骨质疏松症的临床等效性研究”的疗效指标, 4 个部位 [前后位 2~4 腰椎 ( $L_2 \sim L_4$ ), 股骨颈, Wards 三角, 大转子] 骨密度 (BMD,  $g/cm^2$ ) 的改善率: BMDL、BMDN、BMDW、BMDR 进行多变量等效性检验 (BMD 改变的  $\Delta$  值为 2%) 来评价每天口服 10 mg A 药和 B 药的临床等效性。

1. 四个指标单变量等效性检验结果: 首先对四个指标进行传统的单变量等效性检验, 取单侧  $t_{0.05, 247} = 1.653$ , 若  $t_1$  和  $t_2$  均  $> 1.653$ , 则认为两药等效。经检验 (表 1), A 药与 B 药在改善 Wards 三角骨密度值上不等效, 前者疗效不比后者差, 在其他三个指标上是等效的。

表1 试验药 (A) 与阳性对照药 (B) 的骨密度改善率 (%) 临床等效性检验

测量部位	等效 临界值	A (n = 162)		B (n = 86)		检验统计量	
		$\bar{x}$	s	$\bar{x}$	s	$t_1$	$t_2$
前后位腰椎 ( $L_2 \sim L_4$ )	2	3.86	4.52	4.41	2.83	2.53	4.46
股骨颈	2	2.53	4.42	2.36	5.18	3.47	2.92
Wards 三角	2	3.47	9.70	3.04	10.54	1.82	1.18*
大转子	2	2.62	4.48	2.30	5.50	3.36	2.44

\*  $P > 0.05$

2. 多变量等效性检验结果: 股骨颈、Wards 三角和大转子 3 个部位相邻, 因而其骨密度测量值之间存在相关性, 仅用单变量统计分析不能觉察变量间的相关关系, 因而采用两组均数向量等效性检验方

法,以对 A 药的疗效作出综合评价。

治疗前 BMDL、BMDN、BMDW、BMDR 四个测量指标的 Pearson 相关系数阵:

$$\begin{pmatrix} 1.000 & & & \\ 0.092 & 1.000 & & \\ -0.005 & 0.525^* & 1.000 & \\ 0.016 & 0.342^* & 0.373^* & 1.000 \end{pmatrix}$$

\* 在  $\alpha=0.01$  水平上有显著性

(1) 两组均数向量等效性检验:服用 A 药(试验组,  $T$ )和 B 药(对照组,  $R$ )两组样本(四个测量指标)的均数向量分别为:

$$\bar{x}^{(T)} = (3.860\ 5 \quad 2.527\ 6 \quad 3.466\ 0 \quad 2.615\ 5)$$

$$\bar{x}^{(R)} = (4.410\ 3 \quad 2.357\ 8 \quad 3.037\ 7 \quad 2.298\ 5)$$

从而可以求出两组的协方差阵  $S_T$ 、 $S_R$  与合并估计矩阵  $S_C$ :

$$S_C = \begin{pmatrix} 18.389\ 3 & 1.880\ 5 & -0.179\ 6 & 0.368 \\ 1.880\ 5 & 22.027\ 1 & 24.621\ 3 & 7.792\ 7 \\ -0.179\ 6 & 24.621\ 3 & 99.915\ 8 & 18.116 \\ 0.368\ 8 & 7.792\ 7 & 18.116\ 1 & 23.610 \end{pmatrix}$$

将两样本均数向量及合并估计矩阵代入检验统计量公式,求得  $T_1^2 = 30.769\ 7$ ,  $F_1 = 7.598\ 6$ ;  $T_2^2 = 24.700\ 7$ ,  $F_2 = 6.099\ 9$  ( $\nu_1 = 4$ ,  $\nu_2 = 243$ ),对应的  $P$  值均  $< 0.01$ ,按单侧  $\alpha = 0.05$  水准,拒绝  $H_0$ ,接受  $H_1$ 。说明两药对于治疗绝经后骨质疏松至少有一个指标是临床等效的。

(2) 不等效变量筛选:服用 A 药和 B 药两组混合后的样本均数向量为:

$$\bar{x} = (4.051\ 2 \quad 2.468\ 7 \quad 3.317\ 5 \quad 2.505\ 5)$$

对  $T$  和  $W$  分别取逆 ( $T^{-1}$  和  $W^{-1}$ ) 后求得扣除相关后的组间协方差阵  $SB$  的对角线元素向量为:

$$\begin{pmatrix} 17.394 \\ 0.369 \\ 0.180 \\ 2.833 \end{pmatrix}$$

组内合并协方差阵  $S_C$  的逆矩阵为:

$$S_C^{-1} = \begin{pmatrix} 0.055\ 1 & -0.006\ 6 & 0.001\ 7 & 2.445\ 2E-7 \\ -0.006\ 6 & 0.065\ 7 & -0.014\ 3 & -0.010\ 6 \\ 0.001\ 7 & -0.014\ 3 & 0.014\ 7 & -0.006\ 6 \\ 2.445\ 2E-7 & -0.010\ 6 & -0.006\ 6 & 0.050\ 9 \end{pmatrix}$$

分别取  $SB$  和  $S_C^{-1}$  的对角线元素  $SB_{ii}$  和  $S_C^{-1}$  代入检验统计量计算公式,等效界值向量  $\Delta$  为 (2 2 2 2), 求出  $t$  值:

$$t_1 = \begin{pmatrix} -3.818 \\ 5.008 \\ 2.205 \\ 6.230 \end{pmatrix} \quad t_2 = \begin{pmatrix} 10.854\ 0 \\ 2.674\ 3 \\ 1.433\ 3 \\ 0.535\ 5 \end{pmatrix}$$

BMDL、BMDN 两个指标双侧的  $P$  值均  $< 0.01$ ,按单侧  $\alpha = 0.05$  水准,拒绝  $H_0$ ,因而可以认为两药治疗绝经后骨质疏松在这两个指标上临床等效。BMDW、BMDR 两个指标一侧的  $P$  值均  $> 0.05$ ,不能拒绝  $H_0$ ,认为两药在这两个指标上不等效,A 药对于改善 Wards 三角和大转子两个部位骨密度值的效果不比 B 药差,该方法比单变量检验多筛出一个不等效变量。

### 讨 论

等效性试验源于美国 FDA 规定的对专利过期药物的申请要求,由于仿制药与专利药有相同的药理成分,仍采用安慰剂对照试验,会面临伦理上的困难。在这种情况下,只需做等效性试验就可以,即阳性对照的随机临床试验。

临床研究中,各个疾病的反应变量在大多数情况下不止一个,例如,血压的记录为收缩压和舒张压,甲状腺功能检查通常有  $T_3$ 、 $T_4$  和 TSH,肺功能等测量指标则多达十几个乃至几十个。如果从与试验主要目的有关的多个指标中难以确定单一的目标参数时,可同时将几个临床终点设为目标参数。目前已有学者提出了均数向量无差异性检验的方法,对于等效性检验,也应有多个临床终点的检验方法。鉴于统计量  $F$ ,拒绝域为超几何球体,采用 Monte Carlo 方法随机模拟,在大样本情况下,与本文结论一致。

均数向量的检验结果提供了两样本均数向量间是否等效的总的信息,但尚未提供各变量间是否等效的具体信息,为得到这方面的信息,需进行不等效变量筛选。当涉及的变量数  $> 2$  时,若仍用  $t$  检验对两组变量分别假设检验,则会增大第一类错误的概率,若每次比较的检验水准  $\alpha = 0.05$ ,则每次比较不犯第一类错误的概率为  $(1 - 0.05) = 0.95$ ,当这些检验独立进行时,正确接受全部 4 次无效假设的概率为  $0.95^4 = 0.814\ 5$ ,这时犯第一类错误的概率,也就是总的显著性水平变为  $(1 - 0.814\ 5) = 0.185\ 5$ ,这就增加了犯第一类错误的概率,即可能把本来不等效 ( $P > 0.05$ ) 的变量判为等效 ( $P \leq 0.05$ )。

对于多个单变量检验结果不一致的问题,典型的做法是在试验设计阶段就作出定义:对各指标分别进行单变量等效性检验,多个指标中至少有几个等效时就认为两药是等效的,但是这一做法并不能克服另外两个缺点。若均数向量的检验结果是  $P > \alpha$ ,则每个变量必然是不等效的,不必再进行进一步

的计算。若均数向量是等效的( $P \leq \alpha$ ),必然有等效变量的存在,但不一定所有变量都等效。这时可以进一步筛选两药在哪个或哪几个变量上等效(不等效)。其原理是对总的离差阵和组内离差阵取逆的方法扣除变量间的相关性。

通过以上计算及分析,可见单变量的等效性检验结果与多变量的等效性检验结果会出现不一致,而且变量筛选的方法更易出现不等效的结果。目前进行等效性评价的临床试验日益增多,这些试验通常采用多个临床终点,因而应该对其提出统一的评价方法,以规范药品市场的管理,加强统计学在医药

领域的正确应用。

### 参 考 文 献

- 1 方积乾,主编.医学统计学与电脑实验.上海:上海科学技术出版社 2001.361-368.
- 2 Huque MF, Fubey S, Fredd S. Establishing therapeutic equivalence with clinical endpoints. Proceedings of Biopharmaceutic Section of America Statistical Association, 1989. 46-52.
- 3 Wang WH, Hwang JT, Dasgupt Q. Statistical tests for multivariate bioequivalence. Biometrika, 1999, 86:395-402.
- 4 孙尚拱,编著.医学多变量统计与统计软件.北京:北京医科大学出版社 2000.171-182.

(收稿日期 2002-12-06)

(本文编辑:张林东)

## · 疾病控制 ·

### 北京市房山区剖宫产流行病学调查

宋文萍 刘盛辉 王丽静 李兰双 王文仪 李冬梅

近 20 年来随着剖宫产率不断上升,孕产妇及围产儿病死率也相应升高。为降低剖宫产率,对北京市房山区剖宫产情况进行了调查。查阅 1994 年 1 月至 1998 年 12 月区内 5 所医院的所有产科病历。剖宫产指征(第一指征)难产、胎儿窘迫、妊娠合并症/并发症、臀位、剖宫产再孕、社会因素、其他因素。剖宫产新生儿 1 min 及 5 min Apgar 评分  $\leq 7$  分为新生儿窒息。5 年分娩总数 22 372 例,剖宫产 7 835 例,平均剖宫产率 35.02%。孕妇平均年龄 25.50 岁  $\pm$  2.54 岁(20~38 岁),平均孕期 39.75 周  $\pm$  1.52 周(28~44 周),初产妇 7 033 例(88.76%)经产妇 802 例(10.24%)。各年度变化见表 1。各家医院分娩总数、剖宫产例数变化:房山区第一医院 6 044 例、2 259 例(37.38%);良乡医院 5 265 例、1 843 例(35.00%)燕山医院 4 826 例、1 531 例(31.72%)四〇一医院 3 003 例、1 051 例(35.00%)妇幼保健院 3 234 例、1 151 例(35.59%)。剖宫产指征变化:难产 2 396 例(30.58%)胎儿窘迫 2 179 例(27.81%),妊娠合并症/并发症 1 677 例(21.41%),臀位 673 例(8.59%),社会因素 341 例(4.35%),其他因素 355 例(4.53%),剖宫产再孕 214 例(2.73%)。7 835 例剖宫产产妇共娩出新生儿 7 919 例(双胎 82 例,三胎 1 例)。

讨论 20 世纪 90 年代以来我国大部分城市剖宫产率在 40% 左右,某些医院已超过 70%。本资料显示:房山区 5 所医院 1994~1998 年平均剖宫产率 35.02%,呈逐年上升趋势。剖宫产指征中难产所占比例最多;其次为胎儿窘迫;妊

表1 北京市房山区 1994~1998 年剖宫产及其新生儿窒息数变化

年份	分娩总例数	剖宫产例数 (%)	剖宫产新生儿例数	剖宫产新生儿窒息例数 (%)
1994	4 812	1 352(28.10)	1 366	20(1.46)
1995	4 510	1 351(29.96)	1 367	24(1.76)
1996	4 500	1 634(36.31)	1 653	45(2.72)
1997	4 723	1 781(37.71)	1 799	62(3.45)
1998	3 827	1 717(44.87)	1 734	82(4.73)
合计	22 372	7 835(35.02)	7 919	233(2.94)
P 值		0.000 00	0.000 00	0.000 00
$\chi^2$ 值		317.277	320.950	72.255

娠合并症/并发症第三,臀位第四,其他因素第五,社会因素第六,剖宫产再孕最少。由于手术安全性提高,生活水平提高,孕产妇年龄增大,初产妇增多,怕阴道分娩疼痛,怕受“两回罪”,常不愿试产,试产时间不足,催产素引产不规范;孕产妇及家属担心阴道分娩使新生儿损伤、窒息、死亡,使相对头盆不称、巨大儿、宝贵胎儿等非绝对指征因素增多,使剖宫产率升高。在剖宫产率上升初期,孕产妇及围生儿病死率下降。随着剖宫产率继续上升,术中发生麻醉意外、大出血、感染等机率增加,死亡率增加;剖宫产组新生儿病死率较阴道分娩明显增高。本资料显示:剖宫产率升高,剖宫产新生儿窒息率相应升高。阴道分娩减少,医务人员尤其年轻医师产程监测、处理的能力、阴道助产技术得不到提高,造成母婴产伤,医疗纠纷增加。

(收稿日期 2003-01-17)

(本文编辑:张林东)

作者单位:102401 北京市房山区良乡医院(宋文萍、刘盛辉);房山区第一医院(王丽静);房山区妇幼保健院(李兰双);燕山石化职工医院(王文仪);北京市四〇一医院(李冬梅)