

多区域临床试验中样本量分配方法的探讨

董沾健 赵耐青 林燧恒[△]

(复旦大学公共卫生学院卫生统计和社会医学教研室 上海 200032)

【摘要】 目的 研究在多区域临床试验中样本量在各个地区比例分配的合理方法,计算为保证治疗效应在各个地区的一致性概率在80%或90%以上时最小的地区样本量的最低比例为多少。**方法** 通过多次设定各个地区的样本比例,根据治疗效应一致性概率的计算公式进行多次的计算得到一致性概率并作相关的线图分析。**结果** 在两个地区的多区域临床试验中,检验效能为80%时,为保证两个地区的治疗效果一致性概率 P 在80%(或90%)以上,则最小地区样本比例至少要9.3%(或22.2%)。检验效能为90%时,为保证两个地区的治疗效果一致性概率 P 在80%(或90%)以上,则最小地区样本比例至少要6.8%(或15.9%)。**结论** 在两地区的多区域临床试验中,随着最小地区样本比例的增加,两个地区的治疗效果一致性概率 P 也不断的增大。

【关键词】 样本量; 多区域临床试验; 一致性概率

【中图分类号】 R 195.1 **【文献标志码】** A

A method to allocate sample size into each region in multiregional clinical trial

DONG Zhan-jian, ZHAO Nai-qing, LIN Sui-heng[△]

(Department of Biostatistics and Social Medicines, School of Public Health, Fudan University, Shanghai 200032, China)

【Abstract】 Objective To find a way to allocate sample size into each region in a multi-regional clinical trial, and to find the minimal sample size for the smallest region so that there is a high probability of consistent trend in treatment effect across regions. **Methods** We derive the probability of observing a consistent treatment effect and provide some numerical calculations. **Results** For two regions, with the power of 0.8, the smaller region should contribute at least 9.3% (or 22.2%) of the total sample size to get a probability of around 80% (or 90%); for a consistent treatment effect and when the power is equal to 0.9, the smallest region should contribute at least 6.8% (or 15.9%) to get probability around 80% (or 90%). **Conclusions** For two regions, the bigger the proportion of patients from the smaller region, the more the probability that the treatment effect is consistent across the two regions.

【Key words】 sample size; multiregional clinical trial; consistency probability

多区域临床试验(multiregional trial)是由多国研究者在多个国家或地区的多个临床试验中心共同进行的临床试验,并且使用同一个方案^[1]。多区域临床试验组织严密、病例数通常较大,可以入选不同种族不同地区的病例,能够比较广泛的代表各种人群的资料。在1998年,ICH-E5指导原则,题目为国外临床试验数据可接受性的种族因素,在美国、欧洲和日本执行^[2]。当今FDA、欧盟、日本的药监部门

对新药的审批通常都要求有较大样本量的多区域临床试验获得的数据对新药的药效和安全性加以充分的支持。对于中国而言,进入多区域临床试验,将面临着机遇和挑战。随着我国新药临床试验管理力度的不断增强,我国新药临床试验研究活动也日益规范化,我国和国际新药临床试验研究的差距越来越小;我国人口众多,病源也多种多样,潜在的受试者人群大,国外制药企业看准了我国巨大的临床试验

市场,纷纷在我国建立研发中心,把新药带到我国进行临床研究,或者邀请国内研究单位参与国际多中心临床试验和多区域临床试验。多区域临床试验主要为了解决药物延迟问题和时间延迟的问题,使得药物的审批时间缩短,使病人能够获得及时的治疗,这也是我们所期望发生的。

多区域临床试验和其它的一般临床试验一样,要从临床、管理注册机构和统计三个方面去考虑。2007年9月,日本提出了关于多区域临床试验的基本原则,以问题和答案的形式说明了一些关于多区域临床试验的注意问题和意见,作为设计和实施多区域临床试验的参考。关于多区域临床试验主要是评价药物的总体效果,同时也期望在每个国家或地区都有效果(一致性)和能够保证这些结果对种族因素不敏感^[2]。从统计的角度出发,在设计和开展多区域临床试验时,我们要考虑的有样本量、统计的效能和统计分析的方法,以及在不同的国家或地区病例数如何的分配等问题;同时在ICH-E5中提出了要评价每个国家或地区的药物疗效的一致性,目前关于这些问题的研究国内很少有报道。

因此,本文拟通过分析不同的国家或地区病例数分配,了解什么样的比例能够保证在每个国家或地区的药物疗效一致性。比如中国加入了多区域临床试验后,中国的病例占所有病例的比例最低为多少时,才能保证每个国家或地区的药物疗效的一致。目前,关于多区域临床试验的样本量如何分配,国内外没有统一的标准和规则。本研究假设两个地区(中国加入欧洲)的国际多区域临床试验,样本量是通过主要研究目的和主要的疗效指标计算,同时采用安慰剂对照设计,主要疗效指标是连续性变量,并进行优效性检验,而且假定治疗的效果为阳性和两个地区的治疗效果都为阳性。在以上假设的前提下,对于中国加入欧洲的这个多区域临床试验,最低要多少比例的样本量才能保证两个地区的治疗效果一致(一致性的概率达到80%或90%,关于一致性的概率在下文中详细的介绍),其中中国的比例小于等于欧洲的比例。

原理和方法

首先介绍一下一致性概率,在一个多区域临床试验中,假定设置安慰剂为对照组进行优效性检验,并且假设地区和治疗之间不存在交互作用。我们设一个随机变量为 Y_{ij} ,表示临床试验的主要终点疗效指标, $Y_{1j}(i=1)$ 表示第 $j(j=1,2,3,k)$ 个地区安慰剂组终点疗效指标, $Y_{2j}(i=2)$ 表示第 $j(j=1,2,3,$

$k)$ 个地区新药治疗组终点疗效指标。假定 Y_{ij} 服从正态分布 $N(\mu_{ij}, \sigma^2)$,并且假定 μ_{ij} 越大,其疗效越好。假定整个临床试验的总样本量是 $2N$,试验组和安慰剂对照组的样本量相等分别为 N 。假定在多区域临床试验中,共有 k 个地区, N_{ij} 表示在第 $j(j=1,2,3,k)$ 个地区接受第 i 种治疗方式的病例数。通过上面的假设和定义,我们可知 $N = N_{i1} + N_{i2} + \dots + N_{ik}$ 。我们用 D_j 表示第 j 个地区的新药的治疗的效应大小,其中 D_j 表示观察的两组的均值差异(新药治疗组-安慰剂对照组),即 N_{2j} 个病人接受新药治疗组和 N_{1j} 个病人接受安慰剂治疗组的主要疗效指标的均数的差异。由上可知,我们认为 D_j 也服从正态分布:

$$D_j \sim N\left(\Delta, \frac{N_j}{N_{1j}N_{2j}}\sigma^2\right)$$

其中, $\mu_{2j} - \mu_{1j} = \Delta$, $N_j = N_{1j} + N_{2j}$ 表示第 j 个地区的总样本量。因此, D_j 是从每个地区计算得到的,相互之间是独立的。那么在所有的地区治疗效应一致性的概率($D_j > 0$)我们可以表示如下公式1,其中 Φ 表示标准正态分布的累积概率方程^[3]。

$$\begin{aligned} P &= \Pr(D_j > 0, 1 \leq j \leq k | \Delta, \sigma) \\ &= \prod_{j=1}^k \Pr(D_j > 0 | \Delta, \sigma) \\ &= \prod_{j=1}^k \Pr\left(\frac{D_j - \Delta}{\sqrt{\frac{N_j \sigma^2}{N_{1j}N_{2j}}}} > -\frac{\Delta}{\sqrt{\frac{N_j \sigma^2}{N_{1j}N_{2j}}}}\right) \\ &= \prod_{j=1}^k \left[1 - \Phi\left(-\frac{\Delta}{\sigma} \sqrt{\frac{N_{1j}N_{2j}}{N_j}}\right)\right] \\ &= \prod_{j=1}^k \Phi\left(\frac{\Delta}{\sigma} \sqrt{\frac{N_{1j}N_{2j}}{N_j}}\right) \end{aligned} \quad (1)$$

由公式1可以看出,在同一个地区的新药治疗组和安慰剂治疗组的病例相等时($N_{1j} = N_{2j}$),概率 P 最大。因此我们假定在多区域临床试验,每个地区的两组的样本量都相等,即 $N_{1j} = N_{2j}$ 。并且设 p_j 为第 j 个地区样本量 N_j 占总样本量($2N$)的比例,那么每个地区样本量为 $N_j = 2Np_j (j=1,2,\dots,k)$, $\sum p_j = 1$,由此上述公式1可以写成公式2。

$$P = \prod_{j=1}^k \Phi\left(\frac{\Delta}{\sigma} \sqrt{\frac{Np_j}{2}}\right) \quad (2)$$

在样本量的计算中,关于正态分布的试验终点指标,两组的样本量相等,每组的样本量 N 的计算公式^[4],其中 $Z_{1-\alpha/2}$ 和 $Z_{1-\beta/2}$ 分别是双侧概率为 α 和单侧概率为 β 的标准正态分布的分位数。本次研究假定总的样本量是以此公式计算的。

$$N = \frac{2(Z_{1-\alpha/2} + Z_{1-\beta/2})^2 \sigma^2}{\Delta^2} \quad (3)$$

把3式代到公式2中,可以得到式4。由此我们可以看出,如果样本量是按照3式计算,那么一致性概率 P 和各个地区样本量比例有关。并且从公式4可以看出当总的样本量固定不变时和每个地区的病例数一样时,一致性概率 P 随着地区(k)数目的增加而减小。

$$P = \prod_{j=1}^k \Phi [\sqrt{p_j} (Z_{1-\alpha/2} + Z_{1-\beta})] \quad (4)$$

结 果

本研究假设共有两个地区(中国加入欧洲)的国际多区域临床试验,其中中国的样本量占总样本量的比例为 p_1 ,那么欧洲的样本比例就为 $p_2 = 1 - p_1$ 。假设中国样本量占的比例 p_1 小于等于欧洲的比例 p_2 ($p_1 \leq p_2$)。在临床试验中,样本量一般是通过主要研究目的和主要的疗效指标计算的,本研究假定治疗的效果是阳性的和两个地区是一致的,同时采用安慰剂对照设计,主要疗效指标是连续性变量,进行优效性检验;我们设置检验水准为0.05,检验的效能分别取80%和90%进行分析,把 p_1 的取值定为(0.05, 0.051, 0.052, ..., 0.5),根据公式4来计算和分析两个地区的治疗效果一致性概率 P 。最后用概率 P 和中国的样本比例 p_1 作线图,见图1。由图1可以看出,随着中国的样本比例 p_1 的增加,一致性概率 P 逐渐增大。在相同的 p_1 下,效能为90%时,计算的一致性概率 P 都大于效能为80%计算的一致性概率。当两个地区的比例相等时,即 $p_1 = p_2 = 0.5$,一致性的概率达到最大值;效能为90%和效能为80%时,一致性概率 P 都大于95%。

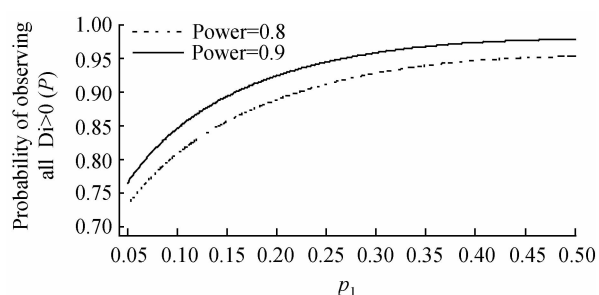


图1 一致性概率 P 和中国样本比例 p_1 的关系

Fig 1 The relation between P and p_1

当总的样本量固定时,检验效能为80%,为了保证两个地区的治疗效果一致性概率 P 在80%以上,那么中国的样本比例 p_1 至少要9.3%;为了保证一致性概率 P 在90%以上,那么 p_1 至少要22.2%,结果见表1。当总的样本量固定时,检验效能为90%,为了保证两个地区的治疗效果一致性概

率 P 在80%以上,那么中国的样本比例 p_1 至少要6.8%;为了保证一致性概率 P 在90%以上,那么 p_1 至少要15.9%,结果见表1。同时从表1的数据可以看出,随着中国的样本比例 p_1 的增加,一致性概率 P 也逐渐增大。

表1 一致性概率 P 和样本比例 p_1, p_2 的关系

Tab 1 The relation between the consistency probability and the sample proportion p_1, p_2 (%)

P	Power = 80%		P	Power = 90%	
	p_1	p_2		p_1	p_2
75	6.00	94.00	75	4.90	95.10
80	9.30	90.70	80	6.80	93.20
85	14.20	85.80	85	10.40	89.60
90	22.20	87.80	90	15.90	84.10
95	42.90	57.10	95	26.60	73.40

P is the probability of the consistent treatment effect

讨 论

本次研究假定有两个国家或地区参加的多区域临床试验,实际上也可以运用到3个国家或地区参加的临床试验。本次研究为两个国家或地区的临床试验,如果是3个国家的多区域临床试验(如欧洲,美国,中国),那么为了保证一致性概率 P 在80%以上,那么 p_1 至少要21.3%;并且在效能为80%时,一致性概率不可能超过90%。同样,如果是3个国家的多区域临床试验,那么为了保证一致性概率 P 在80%以上,那么 p_1 至少要15.1%。为了保证一致性概率 P 在90%以上,那么 p_1 至少要27.7%^[3]。关于多区域临床试验,如果有4个或更多的国家或地区参加,这也值得我们进一步研究。同时本研究假定样本量是通过主要研究目的和主要的疗效指标计算的,同时安慰剂作为对照组设计,主要疗效指标是连续性变量,并进行优效性检验。在以上假设的前提下,为多区域临床试验的样本量的分配提供参考。在实际的多区域临床试验中,可以采用阳性对照,反应变量是分类变量或是生存时间等,在这些条件下,样本量的分配和计算的问题还有值得进一步研究。

本次研究的前提是治疗的效果是阳性的(确实有效果)和两个地区是一致的,以此来研究中国的样本量所需要的最低比例,从而保证很高的治疗效应一致性概率。因为当治疗效果是阳性、且地区间一致时,我们同样可以观察到某个地区的治疗效果是相反的^[5],主要原因是由于交互作用的影响。本研究未考虑国家或地区和治疗之间的交互作用,因为

考虑交互作用比较复杂、交互作用很难解释,并且在管理机构如FDA的观点中表示交互作用不是同等重要和值得考虑^[6]。当两个地区的比例相等时,治疗效应一致性的概率要达到最大值,但是在实际过程中,很难保证每个地区的样本量相等,由于某个地区资金和时间的因素,病人入组的速度和疾病的患病率的不同等因素都会导致不同地区样本量的不同。

本次研究假设终点主要疗效指标是服从正态分布的,如果不服从正态分布,服从二项分布或者是其他特定的分布,其样本量的比例和一致性的概率又如何、本次方法是否适合,也值得我们进一步去研究。如果是二项分布,怎样去分析主要疗效,在结果分析的时候需要进一步注意,因为在不同的地区或国家,治疗作用对二项结局变量影响并不一致^[7]。

多区域临床试验和一般临床试验一样,研究的主要目的为两个方面:有效性和安全性。本次研究主要考虑对有效性的终点指标进行分析,并没有考虑对安全性指标进行分析。考虑到安全性,在多区域临床试验中,由于不同的国家或地区之间可能存在种族因素的差异,在实际工作中要考虑其影响。比如欧洲是多个国家和多个种族组成的,在ICH-E5中对种族的因素有详细的说明。本次研究是在一定的假设前提下进行的,实际的临床试验过程会更加复杂,如调查员的培训和认证不一致、医学操作和服务的不同及语言障碍等等,都会对多区域临床试验带来挑战,因此关于多区域临床试验有很多问题需要进一步研究和讨论。

参 考 文 献

- [1] ES Implementation Working Group. Q & A for the ICH-E5 guideline on ethnic factors in acceptability of foreign clinical data. [C/OL]// International conference on harmonization of technical requirements for registration of pharmaceuticals for human use, Geneva[2006-06-02]. <http://www.ich.org/LOB/media/MEDIA1194.pdf>
- [2] European Medicine Agency. ICH harmonized tripartite guideline E5 on ethnic factors in acceptability of foreign clinical data. [C/OL]// International conference on harmonization of technical requirements for registration of pharmaceuticals for human use, [1998-09-01]. <http://www.ich.org/LOB/media/MEDIA481.pdf>
- [3] Norisuke Kawai, Christy Chuang-Stein, Osamu Komiyama, An approach to rationalize partitioning sample size into individual regions in a multiregional trial[J]. *Drug Information J*, 2008; 42:139-147.
- [4] Chow SC, Shao J, Wang HS. Sample size calculations in clinical research[M]. NY: Marcel Dekker, 2003; 55-61.
- [5] Li Z, Chuang-Stein C, Hoseyni C. The probability of observing negative subgroup results when the treatment effect is positive and homogeneous across all subgroups[J]. *Drug Information J*, 2007, 41(1): 47-56.
- [6] Anello C, O'Neil RT, Dubey S. Multicenter trials: A US regulatory perspective[J]. *Stat Meth Med Res*, 2005, 14: 303-318.
- [7] Chuang-Stein C. A cautionary note on design implications when the primary analysis is a stratified analysis of a binary endpoint [J]. *Biometrical J*, 2006, 48: 978-983.

(收稿日期:2008-09-10;编辑:张秀峰)

(上接第 227 页)

过程,一旦出现临床症状,即进行对症治疗。总之,手术3年后的愈合方式多为骨性,有效率达到81%,牙根吸收在3年的随访中未见,证明黏接再植植骨术是有效保留自体牙的技术,不失为目前保存纵折牙并发牙周炎患牙的有效治疗方法。

参 考 文 献

- [1] Loe H, Waerhaug I. Experimental replantation of teeth in dogs and monkeys[J]. *Arch Oral Biol*, 1961, 3(2): 176-184.
- [2] 虞国君, 吴伟杰, 龚逸明, 等. 40例纵形折裂磨牙全冠修复的疗效观察[J]. *中国临床医学*, 2000, 7(3): 320-321.

- [3] 陈丹, 连克乾, 朱玮玮. 后牙牙折内外固定治疗的临床体会[J]. *广东牙病防治*, 2001, 9(2): 125.
- [4] Sugaya T, Kawanami M, Noguchi H, et al. Periodontal healing after bonding treatment of vertical root fracture [J]. *Dent Traumatol*, 2001, 17(4): 174-179.
- [5] Kawai K, Masaka N. Vertical root fracture treated by bonding fragments and rotational replantation [J]. *Dent Traumatol*, 2002, 18(1): 42-45.
- [6] 俞立英. 50例重症牙周炎进行牙周翻瓣植骨(HA)术后的临床疗效观察[J]. *口腔医学杂志*, 1998, 18(2): 93-95.

(收稿日期:2008-01-16;编辑:张秀峰)