

医疗器械临床试验设计中样本含量的计算

帅万钧¹, 吕晓宁², 晁 勇¹

(1. 解放军总医院第一附属医院医学工程科, 北京 100048;

2. 海军总医院全军航海航空医学专科中心, 北京 100048)

[摘要] 目的: 为科学设计医疗器械的临床试验, 合理选择试验样本含量。方法: 利用统计方法对医疗器械临床试验的样本含量计算进行探讨, 并结合一些案例进行分析。结果: 提供了常用统计参数的样本含量查询表, 所提出的计算方法经验证方便可行。结论: 医疗器械临床试验样本含量可通过查询表方式快速获取。

[关键词] 样本量; 临床试验; 医疗器械

[中国图书资料分类号] R318.6; R446 [文献标识码] A [文章编号] 1003-8868(2012)07-0018-04

Sample Size Calculation in Clinical Trial Plan of Medical Instruments

SHUAI Wan-jun¹, LV Xiao-ning², CHAO Yong¹

(1. Department of Medical Engineering, the First Affiliated Hospital of General Hospital of the PLA, Beijing 100048, China;

2. Aviation and Diving Medical Center, Navy General Hospital of the PLA, Beijing 100048, China)

Abstract **Objective** To plan clinical trials of medical instruments reasonably and select suitable sample size of the trial.

Methods The calculation methods of sample size were studied with statistical methods. Moreover, some cases were enumerated and analyzed.

Results Some kinds of consulted tables were designed and offered with common statistical parameters. The proposed methods proved to be feasible.

Conclusion The sample size of medical instruments in clinical trials can be rapidly gotten by the consultation tables. [Chinese Medical Equipment Journal, 2012, 33(7): 18-20, 34]

Key words sample size; clinical trial; medical instruments

1 引言

随着我国科学技术和医疗卫生事业的发展, 医疗器械临床试验开展得越来越多, 基于临床药理基地的试验研究也需要进一步科学和规范。在临床试验研究中, 一个重要的问题就是如何科学合理地确定试验样本含量^[1-3]。样本量的计算涉及诸多参数的确定^[4-5], 在临床医学研究中需要根据研究目的、研究要求和研究资料来具体决定。目前, 一些医疗器械临床试验的样本量参照国家主管部门提供的参考值, 未充分考虑各自具体情况和要求, 这在一定程度上给研究结果的可靠性和可信度带来影响。本文将对医疗器械临床试验的样本含量计算方法进行探讨, 并结合一些案例进行分析, 以期对越来越多的医疗器械临床试验设计和产品注册有所帮助。

2 样本量计算的影响因素

2.1 资料性质

临床试验统计分析方法的选用以及样本含量的计算, 是与试验数据资料类型密切相关的。医疗器械临床试验的数据资料可分为计量资料和计数资料, 分别为定量观察和定性观察。一般而言, 在其他参数要求相同条件下, 计数资料较计量资料样本含量大。

量资料样本含量大。

2.2 设计类型

主要有 3 种类型的设计^[6-7]: 一种为

优效性设计, 另一种为非劣效性设计, 还有一种为等效性设计。优效性/非劣效性试验关心的问题是单侧的, 对试验组可能优出的程度未加限制。而等效性试验关心的问题是双侧的, 对试验组和对照组的优劣并不关心, 只在乎其组间的差别。一般而言, 采用统计学检验时, 当研究结果高于和低于效应指标的界限均有意义时, 应该选择双侧检验, 所需样本量就大; 当研究结果仅高于或低于效应指标的界限有意义时, 应该选择单侧检验, 所需样本量就小。

2.3 显著性水平

即假设检验 Ⅰ类错误出现的概率(α), 为假阳性错误出现的概率。 α 越小, 所需的样本量越大, 反之越小。 α 水平由研究者根据研究资料具体决定, 通常取 0.1、0.05 或 0.01。

2.4 检验效能

检验效能又称把握度(power), 把握度的概念与 Ⅱ型错误水平 β 相对应, $\text{power}=1-\beta$ 。因此指定了 β 水平也就等于指定了把握度水平。 β 水平由研究者根据研究资料具体决定, 目前, 在医疗器械临床试验研究中, α 取 0.05 或 0.1, 相应的 β 取 0.2 或 0.1 均可被接受。

2.5 界值

界值即容许误差, 指可从临床意义角度判定差异的最小值。界值的确定需根据已有的文献数据, 设计类型及数据的分布类型, 从临床认识水平及成本效益来综合考虑, 将统计学推理和临床判断相结合。界值确定必须在试验设计阶段完成并在试验方案中阐明。当界值难以确定时, 对均值比较可酌取 1/5~1/2 个标准差或对照组均数的 1/10~1/5 等; 对 2 组率而言, 建议取 15% 以下的值, 通常最大不超过对照组样本率的 1/5。



作者简介: 帅万钧(1979—), 男, 博士, 工程师, 主要从事生物医学工程方面的研究工作, E-mail: small-shuai2002@hotmail.com; 吕晓宁(1981—), 女, 主治医师, 主要从事航海航空医学方面的研究工作, E-mail: lvxiaoning1@sina.com。

2.6 总体标准差或总体率

分别反映计量数据和计数数据的变异程度。一般根据前人经验或文献报道进行估计。如果没有前人经验或文献报道作为依据,可通过预实验取得样本的标准差 s 或样本率 p 分别作为总体标准差 σ 或总体率 π 的估计值。

2.7 样本分配

经常使用的是试验组和对照组等样本含量设计。相等的样本含量可使总的样本含量最少,而且在同等的总样本含量下可达到最高的统计效能。

2.8 其他因素

常见影响临床试验样本量的因素还有不依从和失访等样本脱落的问题。因此,在实际应用中应根据具体试验资料情况,在试验设计阶段考虑脱落因素,适当增加样本含量。

3 样本量的计算方法

为方便表述,以下所用到的符号作统一规定。 T 和 S 分别代表试验组和对照组,也泛指相应组的参数; δ 代表界值; s 代表标准差; N 代表样本量。并假定定量指标服从正态分布,二分类率指标服从二项分布^[7-9]。

考虑通用性,单侧检验非劣效/优效设计的样本量计算如下:

$$N = \left(\frac{1}{w} + \frac{1}{1-w} \right) \cdot (U_{\alpha} + U_{\beta})^2 \cdot \frac{\sigma^2}{\delta^2} \quad (1)$$

双侧检验等效设计的样本量计算如下:

$$N = \left(\frac{1}{w} + \frac{1}{1-w} \right) \cdot (U_{\alpha} + U_{\beta/2})^2 \cdot \frac{\sigma^2}{\delta^2} \quad (2)$$

表1 标准正态百分位数

x	U_x
0.001	3.090
0.002	2.878
0.005	2.576
0.010	2.326
0.020	2.054
0.025	1.960
0.050	1.645
0.075	1.440
0.100	1.282
0.150	1.036
0.200	0.842

其中, w 为试验组样本量占样本总量的比例, α 与 β 分别为可允许的最大的 I、II 类错误水平, U_{α} 和 U_{β} 表示对应于 α 、 β 的标准正态分布百分位数(常用的标准正态百分位数如表1所示), σ^2 表示总体方差, N 为试验组和对照组总的样本含量。一般而言,试验组和对照组采用等样本含量,因此常取 $w=0.5$ 。另外, σ 一般难以获得,对于计量资料,常用两样本合并标准差 s 代替;

对于计数资料,有 $\sigma^2 = \pi(1-\pi)$, π 为二分类指标的总体率,也常用样本率 p 来代替。

由于优效设计样本量计算与非劣效设计的方法基本相同,以下仅讨论非劣效设计的样本量计算,对于优效设计的样本量计算不再赘述。

3.1 非劣效设计的计量资料

对于非劣效设计的计量资料,按单侧检验水平 α ,把握度 $(1-\beta)$,在样本量 $T=S$ 条件下(即 $w=0.5$),试验所需样本总量可按公式(3)计算,式中 s 为两样本合并标准差, δ 为界值。表2为根据一般显著水平参数计算得到的样本含量。

$$N = 4 \cdot (U_{\alpha} + U_{\beta})^2 \cdot \left(\frac{s}{\delta} \right)^2 = G \cdot \left(\frac{s}{\delta} \right)^2 \quad (3)$$

3.2 非劣效设计的计数资料

对于非劣效设计的计数资料,按单侧检验水平 α ,把握度

$(1-\beta)$,在样本量 $T=S$ 条件下(即 $w=0.5$),试验所需样本总量可按公式(4)计算,式中 p 为两样本共同有效率, δ 为界值。表3为根据一般显著水平参数计算得到的样本含量。

$$N = 4 \cdot (U_{\alpha} + U_{\beta})^2 \cdot \frac{p \cdot (1-p)}{\delta^2} = G \cdot \frac{p \cdot (1-p)}{\delta^2} \quad (4)$$

3.3 等效设计的计量资料

对于等效设计的计量资料,按单侧检验水平 α ,把握度 $(1-\beta)$,在样本量 $T=S$ 条件下(即 $w=0.5$),试验所需样本总量可按公式(5)计算,式中 s 为两样本合并标准差, δ 为界值。表4为根据一般显著水平参数计算得到的样本含量。

$$N = 4 \cdot (U_{\alpha} + U_{\beta/2})^2 \cdot \left(\frac{s}{\delta} \right)^2 = G \cdot \left(\frac{s}{\delta} \right)^2 \quad (5)$$

3.4 等效设计的计数资料

对于等效设计的计数资料,按单侧检验水平 α ,把握度 $(1-\beta)$,在样本量 $T=S$ 条件下(即 $w=0.5$),试验所需样本总量

表2 非劣效计量资料在一定参数(α 、 β 、 δ 、 s)条件下的样本含量 N 计算值

δ/s	N			
	$\alpha=0.05, \beta=0.2$	$\alpha=0.05, \beta=0.15$	$\alpha=0.05, \beta=0.1$	$\alpha=0.05, \beta=0.05$
	$G=24.741$	$G=28.751$	$G=34.269$	$G=43.296$
10%	2 474.1	2 875.1	3 426.9	4 329.6
15%	1 099.6	1 277.8	1 523.1	1 924.3
20%	618.5	718.8	856.7	1 082.4
25%	395.9	460.0	548.3	692.7
30%	274.9	319.5	380.8	481.1
35%	202.0	234.7	279.7	353.4
40%	154.6	179.7	214.2	270.6
45%	122.2	142.0	169.2	213.8
50%	99.0	115.0	137.1	173.2
55%	81.8	95.0	113.3	143.1
60%	68.7	79.9	95.2	120.3
65%	58.6	68.0	81.1	102.5
70%	50.5	58.7	69.9	88.4
75%	44.0	51.1	60.9	77.0
80%	38.7	44.9	53.5	67.7
85%	34.2	39.8	47.4	59.9
90%	30.5	35.5	42.3	53.5
95%	27.4	31.9	38.0	48.0

表3 非劣效计数资料在一定参数(α 、 β 、 δ 、 p)条件下的样本含量 N 计算值

p	N							
	$\alpha=0.05, \beta=0.2$		$\alpha=0.05, \beta=0.15$		$\alpha=0.05, \beta=0.1$		$\alpha=0.05, \beta=0.05$	
	$\delta=0.1$	$\delta=0.15$	$\delta=0.1$	$\delta=0.15$	$\delta=0.1$	$\delta=0.15$	$\delta=0.1$	$\delta=0.15$
10%	222.7	99.0	55.7	258.8	115.0	64.7	308.4	137.1
15%	315.4	140.2	78.9	366.6	162.9	91.6	436.9	194.2
20%	395.9	175.9	99.0	460.0	204.5	115.0	548.3	243.7
25%	463.9	206.2	116.0	539.1	239.6	134.8	642.5	285.6
30%	519.6	230.9	129.9	603.8	268.3	150.9	719.7	319.8
35%	562.9	250.2	140.7	654.1	290.7	163.5	779.6	346.5
40%	593.8	263.9	148.4	690.0	306.7	172.5	822.5	365.5
45%	612.3	272.1	153.1	711.6	316.3	177.9	848.2	377.0
50%	618.5	274.9	154.6	718.8	319.5	179.7	856.7	380.8
55%	612.3	272.1	153.1	711.6	316.3	177.9	848.2	377.0
60%	593.8	263.9	148.4	690.0	306.7	172.5	822.5	365.5
65%	562.9	250.2	140.7	654.1	290.7	163.5	779.6	346.5
70%	519.6	230.9	129.9	603.8	268.3	150.9	719.7	319.8
75%	463.9	206.2	116.0	539.1	239.6	134.8	642.5	285.6
80%	395.9	175.9	99.0	460.0	204.5	115.0	548.3	243.7
85%	315.4	140.2	78.9	366.6	162.9	91.6	436.9	194.2
90%	222.7	99.0	55.7	258.8	115.0	64.7	308.4	137.1
95%	117.5	52.2	29.4	136.6	60.7	34.1	162.8	72.3

表4 等效计量资料在一定参数(α 、 β 、 δ 、 s)条件下的样本含量 N 计算值

δ/s	N			
	$\alpha=0.05, \beta=0.2$	$\alpha=0.05, \beta=0.15$	$\alpha=0.05, \beta=0.1$	$\alpha=0.05, \beta=0.05$
	$G=34.269$	$G=38.057$	$G=43.296$	$G=51.984$
10%	3 426.9	3 805.7	4 329.6	5 198.4
15%	1 523.1	1 691.4	1 924.3	2 310.4
20%	856.7	951.4	1 082.4	1 299.6
25%	548.3	608.9	692.7	831.7
30%	380.8	422.9	481.1	577.6
35%	279.7	310.7	353.4	424.4
40%	214.2	237.9	270.6	324.9
45%	169.2	187.9	213.8	256.7
50%	137.1	152.2	173.2	207.9
55%	113.3	125.8	143.1	171.8
60%	95.2	105.7	120.3	144.4
65%	81.1	90.1	102.5	123.0
70%	69.9	77.7	88.4	106.1
75%	60.9	67.7	77.0	92.4
80%	53.5	59.5	67.7	81.2
85%	47.4	52.7	59.9	72.0
90%	42.3	47.0	53.5	64.2
95%	38.0	42.2	48.0	57.6

可按公式(6)计算,式中 p 为两样本共同有效率, δ 为界值。表5为根据一般显著水平参数计算得到的样本含量。

$$N = 4 \cdot (U_{\alpha} + U_{\beta/2})^2 \cdot \frac{p \cdot (1-p)}{\delta^2} = G \cdot \frac{p \cdot (1-p)}{\delta^2} \quad (6)$$

表5 等效计数资料在一定参数(α 、 β 、 δ 、 p)条件下的样本含量 N 计算值

p	N								
	$\alpha=0.05, \beta=0.2$			$\alpha=0.05, \beta=0.15$			$\alpha=0.05, \beta=0.1$		
	$\delta=0.1$	$\delta=0.15$	$\delta=0.2$	$\delta=0.1$	$\delta=0.15$	$\delta=0.2$	$\delta=0.1$	$\delta=0.15$	$\delta=0.2$
10%	308.4	137.1	77.1	342.5	152.2	85.6	389.7	173.2	97.4
15%	436.9	194.2	109.2	485.2	215.7	121.3	552.0	245.3	138.0
20%	138.0	243.7	137.1	608.9	270.6	152.2	692.7	307.9	173.2
25%	642.5	285.6	160.6	713.6	317.1	178.4	811.8	360.8	203.0
30%	719.7	319.8	179.9	799.2	355.2	199.8	909.2	404.1	227.3
35%	779.6	346.5	194.9	865.8	384.8	216.5	985.0	437.8	246.2
40%	822.5	365.5	205.6	913.4	405.9	228.3	1 039.1	461.8	259.8
45%	848.2	377.0	212.0	941.9	418.6	235.5	1 071.6	476.3	267.9
50%	856.7	380.8	214.2	951.4	422.9	237.9	1 082.4	481.1	270.6
55%	848.2	377.0	212.0	941.9	418.6	235.5	1 071.6	476.3	267.9
60%	822.5	365.5	205.6	913.4	405.9	228.3	1 039.1	461.8	259.8
65%	779.6	346.5	194.9	865.8	384.8	216.5	985.0	437.8	246.2
70%	719.7	319.8	179.9	799.2	355.2	199.8	909.2	404.1	227.3
75%	642.5	285.6	160.6	713.6	317.1	178.4	811.8	360.8	203.0
80%	548.3	243.7	137.1	608.9	270.6	152.2	692.7	307.9	173.2
85%	436.9	194.2	109.2	485.2	215.7	121.3	552.0	245.3	138.0
90%	308.4	137.1	77.1	342.5	152.2	85.6	389.7	173.2	97.4
95%	162.8	72.3	40.7	180.8	80.3	45.2	205.7	91.4	51.4

4 应用举例

例1:某电子血压计进行临床验证,与传统汞柱式血压计进行配对对照试验,按非劣效设计,基于临床和统计学的综合考虑,取 $\alpha=0.05$, $\beta=0.2$, $\delta=0.67$ kPa, 试验组和对照组样本量 n 取多大合适?

由预实验可以获得样本标准差 $s=1.54$ kPa。依据公式(3)可得:

$$n = \frac{N}{2} = 12.37 \times \left(\frac{1.54}{0.67} \right)^2 = 66$$

即表示:每组用 66 例试验对象,可有 80% 把握度,在 $\alpha=0.05$ 的显著性水平及非劣效界值为 0.67 kPa 的情况下得到

试验组不差于对照组的结论。

若采用等效性设计,则依据公式(5)可得:

$$n = \frac{N}{2} = 17.14 \times \left(\frac{1.54}{0.67} \right)^2 = 91$$

即表示:每组用 91 例试验对象,可有 80% 把握度,在 $\alpha=0.05$ 的显著性水平及等效界值为 0.67 kPa 的情况下得到试验组等效于对照组的结论。

例2:某新型碎石机进行临床试验,与旧款机型按 1:1 比例随机分配试验对象,观察新型碎石机的碎石率是否不差于老款机型。根据临床资料及一般统计学要求,取 $\alpha=0.05$, $\beta=0.1$, $\delta=15\%$, 平均碎石率 $p=80\%$, 试验组和对照组样本量 n 取多大合适?

直接查阅表3,可得 $N=243.7$, 故 $n=N/2=122$ 。即:每组用 122 例试验对象,可有 90% 把握度,在 $\alpha=0.05$ 的显著性水平及非劣效界值为 15% 的情况下得到试验组不差于对照组的结论。

若按等效性设计,查阅表5有 $n=154$, 即每组用 154 例试验对象,可有 90% 把握度,在 $\alpha=0.05$ 的显著性水平及等效界值为 15% 的情况下得到试验组与对照组等效的结论。

需说明的是,以上例题中样本量均未考虑脱落因素。

5 讨论

本文根据试验设计资料的性质对医疗器械临床试验的样本含量进行了分析和计算,并针对常用的参数列出了样本含量计算式和查询表。从表2~5的计算结果可以看出, β 越小(即把握度 $(1-\beta)$ 越大),试验所需样本量越大。对于一定的 α 和 β ,只要试验设计资料的类型确定,试验样本量计算的系数 G 就可以确定下来,其余决定样本量大小的参数主要是界值 δ 与标准差(对于计量资料)或样本率 p (对于计数资料)。因此,统计检验类型的设计是样本量计算的先决条件。本文未对优效性试验的样本量单独进行计算,因为优效性试验设计与非劣效设计均为单侧检验,其计算公式相同,区别仅仅在于界值 δ 会有差异。若其他参数相同,对于优效性试验,其界值往往小于非劣效设计的界值,即可评估试验组有效。从公式(3)或(4)可见, δ 越小,样本量越大。因此,一般而言,优效性设计样本量大于非劣效设计的样本量。

对于非劣效和等效设计,在条件相同时,等效设计的样本量大于非劣效。图1给出了计量资料的某特定条件下非劣效和等效设计的样本含量变化曲线。图中,非劣效 $\alpha=0.05$, $\beta=0.1$ 的样本量曲线与等效 $\alpha=0.05$, $\beta=0.2$ 的曲线重合,因为非劣效是单侧检验,等效是双侧检验,其最终的样本量计算系数 G 相同。从图1可见,

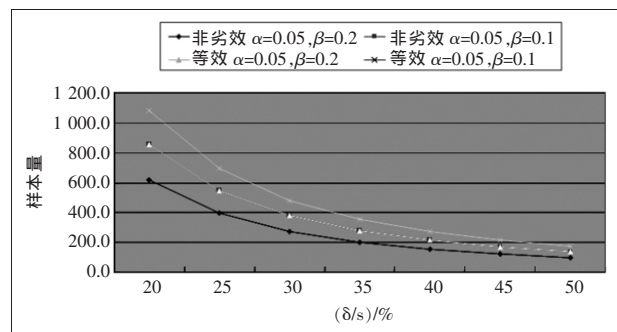


图1 计量资料不同设计参数的样本量

(►►下转第34页►►)

作不灵活;对溃疡灶深、位于角膜边缘处病灶常须多次取样,取样后多为坏死组织,真菌检出率低(70%),假阴性率较高;对操作者的手法要求较高;取材后对角膜病灶周围正常组织有一定损伤,有使病灶扩大的潜在危险性;且患者对使用10号圆刀片取样存在畏惧心理,依从性差,增加了取样的难度。我们研制的一次性角膜标本采集器,通过临床使用观察具有以下优点:(1)头颈部、颈体部的夹角设计,能够保证角膜任意位置的取材操作均可便利进行,无论是溃疡较深的病灶或位于角膜边缘的病灶均可一次性取到足量的标本;(2)头部前后缘及顶部的钝性刀刃可多角度刮取角膜溃疡面病灶分泌物,且能够有效保护正常角膜组织;(3)头部标本储存槽可一次性取到足量标本,且可一次操作同时在溃疡灶多处取材,取材范围广,不易遗漏,避免了重复取样操作,取样后检出率高(85%);(4)小巧轻便,握持舒适,操作简单、灵活,适合在裂隙灯显微镜下操作,对操作者手法的要求不高;(5)一次性使用,安全、卫生,避免了因反复使用消毒不严格造成的二次感染;(6)患者易接受,依从性好,不会引起患者的畏惧心理,可积极配合取样;(7)采集器制作成本低,每个采集器制作成本仅0.15元人民币。

总之,一次性角膜标本采集器使用安全、卫生、快捷,操作简单,取样量足,检出率高,价格低廉,应用于临床后明显提高了真菌性角膜炎的早期诊断率,赢得了治疗的宝贵时机,同时也可避免盲目的抗感染治疗,值得大力推广使用。

[参考文献]

- [1] 王丽娅,张月琴,王印其,等.中国三地区真菌性角膜病致病菌种的调查[J].中华眼科杂志,2000,36(2):138.
- [2] 王丽娅,杨子建,张月琴,等.河南地区真菌性角膜炎病原学及流行病学分析[J].中国实用眼科杂志,2006,24(3):331.
- [3] Tanure MAG, Gohen E J, Sudesh S, et al. Spectrum of fungal Keratitis at Wills Eye Hospital[J]. Philadelphia Pennsylvania Cornea, 2000,19(3):3 072-3 121.
- [4] 邓新国,吕雪芳,庞广仁.312例化脓性角膜溃疡的病因和病原分析[J].眼科研究,1995,13(2):110.
- [5] 罗时运,邹洋,张文华,等.角膜溃疡刮片检查[J].实用眼科杂志,1993,11(1):34.
- [6] O'Day D M, Akrabawi P L, Head W S, et al. Laboratory isolation techniques in human and experimental fungal infection[J]. Am J Ophthalmol, 1979,87(10):688-693.

(收稿:2011-06-14 修回:2011-08-05)

(◀◀上接第20页◀◀)

相同条件下,非劣效样本量小于等效, δ 越大,样本量越小。

检验显著性水平和检验效能是决定样本量大小的关键因素,即对应于统计推断的I型错误 α 和II型错误 β 。 α 和 β 是相互关联的,在样本量一定条件下, α 越小, β 就越大;反之, α 越大, β 则越小。若想同时减小 α 和 β ,唯一的办法就是增加样本量。一般为了均衡, α 取0.1、0.05或0.01; β 取0.2、0.1或0.05。而且,无特殊情况, α 取0.05, β 取0.2能被认可,故本文主要针对常用的 $\alpha=0.05$, $\beta=0.2$ 进行计算。

需再次强调的是,进行非劣性/等效性检验时,需预先确定界值 δ ,这个界值应不超过临床上能接受的最大差别范围。界值的确定需要由主要研究者与生物统计人员共同商讨确定。严格讲,非劣性/等效性试验的标准,如样本含量以及估计样本含量用的 δ 、 α 、 β 等,都要在设计阶段确定。界值确定必须在试验设计阶段完成并在试验方案中阐明。

样本可能的脱落也是试验设计阶段应该考虑到的。试验对象入组的难易、依从性以及截尾因素等需要多方面权衡。而且,对数据要求的苛刻程度也需要把握,因为要求越苛刻,该试验对象试验数据被剔除的可能性越大。一般而言,试验设计中实际样本量需在计算得出的样本量基础上增加10%~20%,如果试验条件较苛刻,应该适当增加此比例。

需要注意的是,样本量 N 在不同环境下的意义。有时它是指试验组和对照组的总样本量,有时它特指某一组的样本量,而且对于配对试验设计,它还可能指配对样本量的对数。另外,目前医疗器械临床试验一般要求进行多中心的试验,那 N 是指多中心总的样本量还是某个分中心的样本量也需要特别留意。本文所提供计算公式中的 N 是指多中心试验组和对照组总的样本含量。

目前有许多软件支持样本量的计算,如SAS等^[10]。充分利用这些软件也是一种有效途径,但前提是对软件必须充分了解,特别是软件设置中各个参数的实际意义,否则容易犯错。

总之,医疗器械临床试验的样本量与诸多因素相关,其中主要包括检验设计、显著性水平、把握度和临床界值。本文主要就有关统计学方面的问题进行了探讨,提出了样本量的简单计算方法和一套快捷查询表,并对样本量计算的影响因素进行了分析,其他试验过程中的实际问题超出本文范围,需在实际操作中具体分析。

[参考文献]

- [1] 成琪,刘玉秀,陈林,等.单组临床试验目标值法的精确样本含量估计及统计推断[J].中国临床药理学与治疗学,2011,16(5):517-522.
- [2] Chow S C. Sample Size Calculations in Clinical Research[M]. 2nd Edition. New York: CRC Press, 2008: 117-120.
- [3] Hu L P, Bao X L, Wang Q. The repetition principle in scientific research[J]. Journal of Chinese Integrative Medicine, 2011, 9(9): 937-940.
- [4] Chen L M, Ibrahim J G, Chu H. Sample size and power determination in joint modeling of longitudinal and survival data[J]. Stat Med, 2011, 30(18): 2 295-2 309.
- [5] 郑青山,孙端云,陈志扬.新药临床非劣性及等效性试验中的例数估计和等效标准[J].中国新药杂志,2003,12(5):368-371.
- [6] 黄钦,赵明.对临床试验统计学假设检验中非劣效、等效和优效性设计的认识[J].中国临床药理学杂志,2007,23(1):63-67.
- [7] 袁小红,庄严,陈平雁.两均数比较的优效性/等效性和非劣性验证的样本量估计[J].中国新药杂志,2009,18(23):2 205-2 209.
- [8] 刘玉秀,姚晨,陈峰,等.非劣性/等效性试验的样本含量估计及把握度分析[J].中国卫生统计,2004,21(1):31-35.
- [9] 王杨,李卫,成小如,等.随机模拟法验证非劣效临床试验样本量计算公式[J].中国卫生统计,2008,25(1):26-28.
- [10] 张春茂,王素珍,李婵娟,等.二分类技术资料整群随机化样本量SAS宏实现[J].现代预防医学,2010,37(20):3 830-3 832.

(收稿:2011-12-21 修回:2012-02-22)