

非参数模型在以等级资料为终点的非劣效性
临床试验中的应用及 SAS实现*

第二军医大学卫生统计学教研室 (200000) 贺江南 陆 健 张新佑 谢之辉 贺 佳[△]

近年来,随着临床试验研究的不断深入,越来越多的临床试验被设计成非劣效性试验,用来评价新药非劣效于标准有效药物。在非劣效性试验中,试验目的不再是为了证明新药优效于标准有效药,而是只要证明新药非劣效于标准有效药疗效即可。近年来非劣效性试验在我国新药临床试验中逐渐被广泛应用,国内越来越多的研究采用非劣效性评价 (non-inferiority evaluation),然而当今的非劣效性评价主要集中在对计量资料和计数资料 (两分类) 的非劣效性研究,对于等级资料 (ordered categorical data) 的非劣效性评价研究甚少。本文将重点讲述非参数模型 (non-parametric model) 在临床试验等级资料的非劣效性评价中的运用。

非参数模型

临床试验中常常设有一些等级定性指标,如临床疗效的评价、疾病的临床分期、病症严重程度的临床分级、中医诊断的一些临床症状等。对于两组等级资料样本中无差异的检验假设,通常采用 Wilcoxon Mann-Whitney 检验^[1-3]或 CMH 方法。比值比模型 (the proportional odds model) 同样也适用于两组等级资料样本的比较,Mehta^[4]曾经做过相关的比值比模型用来评价等级资料的非劣效性,但由于模型参数会随着等级层次的增加而增加,这就意味高精度结果要求大样本资料,不适合小样本等级资料的比较。

假设临床试验中有两个相互独立的等级资料样本, $X_{1i} \sim F_T(x)$, $i=1, \dots, n_1$ 和 $X_{2j} \sim F_S(x)$, $j=1, \dots, n_2$, 其中有 $N=n_1+n_2$ 个相互独立的随机变量, 1 为标准有效药治疗组, 2 代表新药治疗组。为了判断两个等级资料样本疗效的差异, 可以建立判别两组疗效差别的非参数模型, 两组之间的处理效应 p 为:

$$p=P(X_{11}<X_{21})+\frac{1}{2}P(X_{11}=X_{21}) \tag{1}$$

假定这两个分布函数相互独立不交叉, 当 $p>1/2$ (或 $p<1/2$) 时, 那么随机变量 X_{11} 趋向小于 (或大于)

随机变量 X_{21} 。当 $p=1/2$ 时, 两组随机变量趋向相等。显而易见, 在 $F_T=F_S$ 的条件下, 新药治疗组与标准有效药组之间没有疗效差异, 这也意味着 $p=1/2$ 。因此, p 与中心等效值 $1/2$ 的离散度为重点所关注的两组等级资料的差异^[6]:

$$p-\frac{1}{2}=P(X_{11}<X_{21})-\frac{1}{2}[1-P(X_{11}=X_{21})]=\frac{1}{2}[P(X_{11}<X_{21})-P(X_{11}>X_{21})] \tag{2}$$

依据上述非参数模型中的处理效应 p 值以及非劣效性界值 (non-inferiority margin 符号为 δ), 构建非劣效性评价的检验假设^[6]如下:

无效假设 $H_0: p \leq p = \frac{1}{2} - \delta$

备择假设 $H_1: p > p = \frac{1}{2} - \delta$

式中非劣效性界值 $\delta \in [0, \frac{1}{2})$ 表示, 的是最大不相关差异界值。当 $\delta=0$ 时, 检验结果转变为优效性检验结果。

基于 Wilcoxon Mann-Whitney 检验^[3]的基本原理, 可以推导处理效应 p 值的估计值 \hat{p} 与真实的 p 值相比倚度非常小, 其公式如下^[7]:

$$\hat{p}=\frac{1}{N}(R_2-R_1)+\frac{1}{2} \tag{3}$$

式中 $R_i=(1/n_i)\sum_{j=1}^{n_i}R_{ij}$, $i=1, 2$ 。 R_{1j} , R_{2j} 分别是随机变量 X_{1j} 和 X_{2j} 在 $N=n_1+n_2$ 个随机变量 $X_{11}, \dots, X_{1n_1}, X_{21}, \dots, X_{2n_2}$ 中的秩次。

为了检验无效假设 $H_0: p \leq p = \frac{1}{2} + \delta$ 可以设定

统计量为 $K_N=\sqrt{N}(\hat{p}-p)$ 。当 $N \rightarrow \infty$ 时, 那么 $N/n_i \leq N_0 < \infty$, $i=1, 2$ 依据中心极限定理, Bunner 和 Munzel^[8]曾证明了 $K_N=\sqrt{N}(\hat{p}-p)$ 渐近服从正态分布 $(0, \sigma^2)$, 正态分布中方差 $\sigma^2=N \cdot [\sigma_1^2/n_1 + \sigma_2^2/n_2]$ 。Bunner 和 Munzel^[8]对方差 σ^2 进行了估算, 其估计值为:

$$\sigma_i^2=S_i^2/(N-n_i)^2, i=1, 2 \tag{4}$$

其中

*: 十一五“重大新药创制科技重大专项” 新药临床研究技术平台建设; 数据管理和统计分析的平台建设 (2008ZX09312-007 和 2008ZX09312-025)
[△]通讯作者: 贺 佳, hejia@smmu.edu.cn

$$S_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (R_{ij} - R_{ij}^{(i)} - R_i + \frac{n_i + 1}{2})^2, i=1, 2$$

(5)

式中 $R_{ij}^{(1)}$ 是 X_{1j} 在样本 X_{11}, \dots, X_{1n_1} 中的秩次, $R_{ij}^{(2)}$ 是 X_{2j} 在样本 X_{21}, \dots, X_{2n_2} 中的秩次。结合公式 (3)、(4)、(5) 和统计量 K_N , 可以得到最终的检验统计量 T_N 为:

$$T_N = \frac{\sqrt{N}(\hat{p} - p_0)}{\hat{\sigma}} = \frac{R_2 - R_1}{\sqrt{N}\hat{\sigma}} - \delta \frac{\sqrt{N}}{\hat{\sigma}}$$

(6)

在 H_0 成立的条件下, T_N 渐近服从标准正态分布。但需注意的是 T_N 只是渐近服从标准正态分布, 不是完全服从标准正态分布。当样本量 $n_i \leq 50, i=1, 2$ 时, 正态分布需要渐近估算出 T_N 的自由度。根据 *Satterthwaite-Smith Welch* 渐近原理, 可得到 T_N 的自由度的渐近估计值为:

$$\hat{r} = \frac{(\sum_{i=1}^2 S_i^2 / (N - n_i))^2}{\sum_{i=1}^2 (S_i^2 / (N - n_i))^2 / (n_i - 1)}$$

(7)

由于 T_N 检验统计量渐近服从自由度为 \hat{r} 的标准正态分布, 因此可以得到接受备择假设 H_0 的概率为:

$$P(T_N \geq u_{\alpha} | p < p_0, \sigma^2) = \Phi(\sqrt{N} \frac{P - P_0}{V_N} - u_{\alpha})$$

(8)

式中 $\Phi(\cdot)$ 表示标准正态分布下的累积分布。

SAS程序实现

在评价等级资料的非劣效性的非参数模型中, 有多个秩次变量可以通过矩阵来描述, 这些变量之间的计算可通过 **ML** 过程来实现, 下面将建立一个宏程序来构建非参数模型, 以便计算出评价非劣效性的处理效应 p 值。在临床试验数据分析过程中, 统计分析人员只需整理好一个相应的数据集, 运行下面的宏程序就可以得到非劣效性结果。

ML 过程的主要语句如下:

```
%MACRO PERM_BF(DATA=, VAR=, GROUP=, MARGIN=, LOOP=10000); * 定义宏参数;
proc freq data=daten; tables & GROUP / out=ni noprint * 计算频数;
proc iml * 调用 ML过程;
use &data read all var{& VAR} into x; read all var {& GROUP} into group;
use ni read all var{count} into ni; n1=ni[1]; n2=ni[2]; N = n1+n2; * 初始化值;
perm = j(N, 1, 0); o_ranks = j(N, 1, 0); o_ranks_1 = j(n1, 1, 0); o_ranks_2 = j(n2, 1, 0);
w_ranks_1 = j(n1, 1, 0); w_ranks_2 = j(n2, 1, 0);
```

```
mean_o_ranks_1 = 0; mean_o_ranks_2 = 0;
sigma_1 = 0; sigma_2 = 0;
V_N = 0; T_0 = 0; T_perm = 0; pvalue_t = 0;
pvalue_perm = 0; * 计算 T_0值;
o_ranks = ranktie(x); * 总秩和;
o_ranks_1 = o_ranks[1:n1]; o_ranks_2 = o_ranks[(n1+1):N]; * 1, 2组中的总秩和;
w_ranks_1 = ranktie(x[1:n1]); w_ranks_2 = ranktie(x[(n1+1):N]); * 1, 2组内中的秩和;
mean_o_ranks_1 = sum(o_ranks_1)/n1;
mean_o_ranks_2 = sum(o_ranks_2)/n2; * 1, 2组中的总秩和的均数;
releff=1/N * (mean_o_ranks_2 - mean_o_ranks_1) + 1/2; * 处理效应 p值;
* Behrens-Fisher方差估计值;
sigma_1 = 1/(n1-1) * (o_ranks_1 - w_ranks_1 - mean_o_ranks_1 + (n1+1)/2) * * 2;
sigma_2 = 1/(n2-1) * (o_ranks_2 - w_ranks_2 - mean_o_ranks_2 + (n2+1)/2) * * 2;
V1=sigma_1/(n2* * 2); V2=sigma_2/(n1* * 2);
V_N =N* (V1/n1+V2/n2);
if V_N = 0 then V_N = N/(2* n1* n2);
T_0 = sqrt(N) * (releff-0.5 + & delta)/sqrt(V_N);
freheitsgrad = (sigma_1/n2 + sigma_2/n1)##2 / ((sigma_1/n2)##2/(n1-1) + (sigma_2/n1)##2/(n2-1));
pvalue_t = 1-probt(T_0, freheitsgrad);
quit %MEND; * 宏程序结束;
```

实 例

在一个止痛膏新药阳性对照临床试验中, 每个病人的疗效判定标准等级为临床痊愈、显效、有效、无效, 疗效等级的赋值分数分别为: 3、2、1、0。此次试验中 200 名病人的治疗结果见表 1。

表 1 止痛膏新药试验中 200 名病人的疗效等级数据

组别	临床痊愈	显效	有效	无效	样本量
新药组	24	37	26	13	100
对照组	20	42	24	14	100

通过使用上述的非参数模型宏程序, 可以得到以下结果: $n_1=100, n_2=100$, 处理效应 $=0.511$ 。在该临床试验中, 根据统计学和临床意义两方面因素的考虑, 确定其非劣效性界值 $\delta^{[10]}$ 为 0.06, 处理效应 $\hat{p} > p_0 = 0.44$ 。在 $\alpha=0.05$ 的水准下, 检验统计量 T_N 为 1.814, 单侧 P 值为 0.0356, 因此不能拒绝无效假设 H_0 , 由此可以判定该新药疗效非劣效于阳性对照药。

讨 论

本文介绍了评价等级资料的非劣效性的非参数模型,该模型通用性强,无论样本量是大还是小,无论是等级资料还是不服从正态分布的任意分布参数资料,都能够得到较准确的结果。在临床试验实践中,统计分析人员可以结合自身的数据集运行本文中的 SAS 宏程序,分析工作将高效便捷。当然,使用本文介绍的等级资料的非劣效性评价过程,也存在一些缺陷和不足,如等级资料的非劣效性界值确定仍然是个难题,这需要统计专家和临床专家共同合作确定。另外,本文中的非劣效性评价未考虑协变量的效应,临床试验中应尽可能根据试验的实际情况消除协变量的影响。

参 考 文 献

1 Welk S A new approach to equivalence assessment in standard comparative bioavailability trials by means of the mann-whitney statistic Biometrical journal 1996(38): 695-710.

2 Binbaum ZW, OrvalMK. Bounds for the variance of the ann-W hitney Statistic M athematical Statistics 1957 (28): 933-945.

(上接第 515页)

患者总费用的初步预测,可以为相关管理部门制定单病种付费补偿标准提供参考。

表 2 脑梗死患者住院次均费用影响因素的回归分析结果

影响因素	回归系数	标准误	标准化系数	P值
常数	3.295	0.03		0.000
住院天数 X_6	0.015	0.00	0.65	0.000
药费比例 X_8	0.675	0.04	0.26	0.000
是否死亡 X_{52}	0.235	0.04	0.09	0.000
是否院感 X_7	0.110	0.02	0.10	0.000
是否医保 X_{32}	0.028	0.01	0.04	0.004
是否急诊入院 X_{41}	0.041	0.01	0.04	0.005
是否治愈好转 X_{51}	-0.055	0.02	-0.05	0.008
是否危重 X_{42}	0.152	0.06	0.04	0.013

讨 论

从费用回归方程可以看出,住院时间是影响住院费用重要因素之一。缩短平均住院日可以有效地降低患者住院次均费用,能加快病床周转,提高床位使用率,有效降低患者住院费用。影响住院时间的因素有许多,包括可控和不可控的。降低患者住院时间,必须通过科学、合理的方法和方式。如可以通过引入虚拟病床、减少待术日等方法。

昂贵药物的使用是导致患者费用上升的另一重要

3 Subhabrata C, Mark A. van de W iel A nonparametric control chart based on theM ann-W hiney statistic M athematicalStatistics 2008(1): 156-172.

4 Cynus RM, N itin RP, Anastasios A. Tsiatis Exact significance testing to establish treatment equivalence with ordered categorical data International Biometric Society, 1984(40): 819-825.

5 Edgar B runner U llrich M unzel The nonparametric behrens-fisher problem; asymptotic theory and a small-sample approximation Biometrical Journal 2000(42): 17-25.

6 W ellek S H ampe l B. A distribution-free two-sample equivalence test allowing for tied observations Biometrical Journal 1999, (41): 171-186.

7 U llrich M unzel D ieter Hauschke A nonparametric test for proving non-inferiority in clinical trials with ordered categorical data Pharmaceutical Statistics 2003(2): 31-37.

8 B runner M unzel Nichtparametrische Datenanalyse: Unverbundene Stichproben 2002.

9 Neubert Karin B runner Edgar A studentized permutation test for the non-parametric Behrens-Fisher problem Computational Statistics 2007 (51): 5192-5204.

10 Huang HM, W ang SJ O `NEILL Robert A regulatory perspective on choice of margin and statistical inference issue in non-inferiority trials Biometrical journal 2005 (47): 28-36.

因素。随着国家医疗服务价格改革的推进,药品多次降价、定价方法逐步改进,但患者药费总负担却没有明显减轻。大量进口、贵重药品的使用,无形加重了患者的费用负担。因此,应加强和完善高档药品使用审批制度和告知制度,尽量使用目录内的药品,设法鼓励医生使用疗效确切的低价药品,同时加大患者合理用药和合理检查,避免盲目使用高档药品及滥用大型仪器检查治疗。

总之,由于脑梗死疾病本身的特殊性和易复发性,老龄患者的费用负担较重,可能成为家庭的主要经济负担,社会相关部门应加大患者的大病保障和补助力度,防止因病返贫、因病致贫。由于患者的医疗费用受多方面主客观因素的交织影响,控制住院费用必须以保证医疗质量和安全为前提,从可控因素入手,不能单纯依靠控制几个因素,而是要发挥整个控制体系的作用,全程有重点地控制显著影响因素,以最大程度满足患者医疗需求为条件达到降低住院费用的目的^[2]。

参 考 文 献

1 李志辉,罗平. SPSS forW indows统计分析教程.北京:电子工业出版社,2006.

2 刘析梅,温小霓.医保患者住院费用分布与影响因素分析.中国医院管理,2007,27(1): 26.