如何正确运用*t*检验——两算术均值比较非劣效性*t*检验及SAS实现

陈 阳[1],刘媛媛[1],李长平[1,2],胡良平[2,3*]

(1. 天津医科大学公共卫生学院,天津 300070;
2. 世界中医药学会联合会临床科研统计学专业委员会,北京 100029;
3. 军事科学院研究生院,北京 100850
*通信作者:胡良平,E-mail:lphu927@163.com)

【摘要】 本文主要介绍临床试验中非劣效性检验的概念、原理和作用以及成组设计一元定量资料非劣效性检验的SAS实现。通过实例展示SAS在非劣效性检验中的应用,分别基于原始的定量数据或者基于给定样本含量、均数、标准差两种数据结构下进行操作,对结果进行解释并作出结论。

【关键词】 非劣效检验;非劣效性界值;*t*检验;算数均数比较
中图分类号:R195.1 文献标识码:A doi:10.11886/scjsws20200526006

How to use *t* test correctly——the noninferiority *t* test and the SAS implementation

Chen Yang[1], *Liu Yuanyuan*[1], *Li Changping*[1,2], *Hu Liangping*[2,3*]

(1. *Department of Epidemiology and Health Statistics, School of Public Health, Tianjin Medical University, Tianjin* 300070, *China*;
2. *Specialty Committee of Clinical Scientific Research Statistics of World Federation of Chinese Medicine Societies, Beijing* 100029, *China*;
3. *Graduate School, Academy of Military Sciences PLA China, Beijing* 100850, *China*
**Corresponding author: Hu Liangping, E-mail: lphu927@163.com*)

【Abstract】 This article mainly introduced the concept, principle and function of noninferiority testing in clinical trials and the implementation of SAS for noninferiority testing of unary quantitative data in group design. The application of SAS software in the noninferiority testing was demonstrated through examples. The operation was based on the original quantitative data or the data structure of the given sample size, mean, and standard deviation. And it was also to explain the results and make conclusions.

【Keywords】 Noninferiority testing; Noninferiority limit value; *t* test; Comparison of arithmetic mean

当事先知道某两种药物或疗法的初步信息(如A药可能比B药疗效差),但其差值可能不会超出专业上允许的一个界值时,可以考虑采用非劣效性检验[1]。本文主要介绍临床试验中非劣效性检验的相关内容,包括非劣效性检验的概念、意义以及假设检验的原理、重要参数的解释和确定。再结合临床实例,展示SAS中两算术均值比较非劣效性*t*检验的应用,并对程序语句和运行结果进行解释。

1 概 述

1.1 非劣效性检验简介

以安慰剂作为对照的随机双盲临床试验一直被视为药物开发中的金标准,然而随着可应用的有效药物的不断出现,具有突破性疗效的新药却越来越少,所以临床研究的目的也逐渐改变。尽管某些疗法能够提供更高的功效,但其他新疗法可能具有更高的安全性或便利性,或更少的经济花费,同时提供相似的功效。为了寻求良好替代疗法,提出了非劣效性检验。在真实的非劣效临床研究中,当以阳性药物作为对照时,实际上是默认了阳性对照的疗效是客观存在且稳定的,且已知试验药的疗效不可能等于或优于对照药的疗效,当能证明两者疗效之差未超过临床上认可的界值时,即可称为非劣效。如果非劣效性检验成立,试验药物虽然相比于阳性对照药物在疗效上没有优势,但考虑到其他方面的优点,如给药方便、原材料丰富、价格便宜、不良反应少等,那么新药也是值得投入的。此外,在临床试验的设计和分析阶段,评估非劣性比优效性

基金项目:国家自然科学基金项目(项目名称:贝叶斯生存分析方法在肝细胞癌肝移植患者预后预测中的应用研究,项目编号:81803333)

更为复杂，但在某些情况下，非劣效检验的原假设可能比通常的优效性分析更合理，因为它基于先验的效应以及安全性信息[2]。随着临床试验中设计和统计分析的不断改进与规范，标准阳性对照试验的非劣效性设计的应用更加广泛，成为了评价药物、器械、生物制品和其他医学治疗的主要工具，近十年内，评估非劣效性的随机试验数量增加了6倍[3]。目前科研需要发展具有相当功效但兼具其他优点的新方法替代标准方法，在这种前提下，非劣效性试验的应用就越来越普遍，但与此同时，为了保证结果和结论的可靠性，非劣效试验的报告内容和相关的规定也需要不断完善[4]。

1.2 非劣效性设计下两均数比较的假设检验及参数解释

检验假设：μ_1表示试验措施的干预效果（试验组主要指标的效应值，均数）；μ_2表示对照措施的干预效果（对照组主要指标的效应值，均数）。

同时还假定：主要指标为高优指标，即均值取值大为好（若主要指标为低优指标，下面零假设和备择假设中的不等号需要改变方向，且非劣效界值取正值）。δ_L表示非劣效界值（通常取负值）；α表示检验水准（在通常的情况下取α=0.05，单侧检验）；原假设和备择假设分别用H_0和H_1表示。基于以上定义，非劣效性检验的检验假设可表述如下，H_0:$\mu_1-\mu_2\leqslant\delta_L$；$H_1$:$\mu_1-\mu_2>\delta_L$。

$$t=\frac{\left(\overline{X}_1-\overline{X}_2\right)-\delta_L}{s_{\overline{X}_1-\overline{X}_2}}=\frac{\left(\overline{X}_1-\overline{X}_2\right)-\delta_L}{\sqrt{S_C^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}},$$

$$df=\left(n_1+n_2-2\right) \tag{1}$$

$$s_c^2=\frac{\left(n_1-1\right)S_1^2+\left(n_2-1\right)S_2^2}{n_1+n_2-2}=\frac{\Sigma\left(X_1-\overline{X}_1\right)^2+\Sigma\left(X_2-\overline{X}_2\right)^2}{n_1+n_2-2} \tag{2}$$

在式（1）和式（2）中：$\overline{X}_1$和$\overline{X}_2$分别为试验组和对照组的样本均数；δ_L为非劣效界值（通常取负值，代表“差”或“劣”）；n_1和n_2分别为试验组和对照组的例数；df为t分布的自由度；s_1^2和s_2^2分别为试验组和对照组的样本方差；$s_{\overline{X}_1-\overline{X}_2}$为两样本均数之差的标准误；$s_c^2$为两组合并方差。

采用单侧检验。当$t>t_{(1-\alpha),n1+n2-2}$时，差异有统计学意义，拒绝H_0，能够得出试验组非劣于对照组的结论[5]。

在非劣效检验中使用单侧检验，当试验组疗效劣于对照组且其差值小于或等于非劣效界值δ_L时，不能拒绝H_0，尚不能得出试验组非劣效于对照组的结论；如果大于非劣效界值δ_L，则拒绝H_0，能够得出试验组非劣效于对照组的结论。

在置信区间法中，非劣效性试验仅关注试验方法相对于对照方法评价指标的效果差值的置信区间下限的大小。因此，当试验措施与对照措施的效果差异的单侧95.0%置信区间完全落在非劣效性界值右侧时，即其单侧95.0%置信区间的下限大于设定的判断界值δ_L时，则可判定非劣效性假设成立。对于两个均数比较的非劣效检验，按照单侧100(1−α)%的可信度，可以计算单侧置信区间的下限C_L，公式如下：

$$C_L=\left(\overline{X}_1-\overline{X}_2\right)-t_{\left[(1-\alpha),\left(n_1+n_2-2\right)\right]}S_{\left(\overline{X}_1-\overline{X}_2\right)} \tag{3}$$

若(C_L，∞)不包括δ_L，可以得出试验组疗效非劣效于对照组疗效的结论。

1.3 非劣效性界值δ_L的设定

美国食品药品监督管理局指出，在非劣效试验中，设定检验界值应基于“恒定假设”，即尽可能确保本次非劣效性试验中阳性对照药的疗效与既往临床试验保持一致。只有当前非劣效试验与其阳性对照药的历史试验在所有重要研究设计和实施方面均保持一致时，历史试验才可用于估计当前非劣效试验的阳性对照药疗效和非劣效界值[6]，否则会导致错误的非劣效结论。非劣效性界值δ_L设定一般分为两个步骤，可采用综合分析法（常用Meta分析法）估计阳性对照的绝对疗效M_1，计算阳性对照与安慰剂效应之差的95%双侧置信区间下限（必须大于0），鉴于疗效的一致性和既往的临床数据质量，一般M_1取值小于计算所得下限值。非劣效性界值$|\delta_L|=M_2=f^*M_1$，f一般建议取值为0.5。当无法借鉴历史资料时，临床试验的反应率（有效率）高于80%时，$|\delta_L|$一般可以取阳性对照疗效的10%~15%。非劣效性界值应该结合文献，由临床专家确定。根据经验，血压可取为0.67 kPa（5 mmHg），胆固醇可取为0.52 mmol/L（20 mg/dL），白细胞可取为0.5×10^9/L（500个/mm^3）[7]（注意：前面举例给出的是非劣效界值的绝对值，代入公式计算时，应取负号，代表“方向”）。若确实没有公认的文献资料作为参考依据，需要由在所研究问题方面具有权威性的多位临床专家共同商定。

2 实例分析

2.1 基于“样本含量、均值和标准差”进行非劣效检验

【例 1】为了比较拉西地平与苯磺酸氨氯地平治疗中老年原发性轻中度高血压的效果及安全性[8]，入选 8 个中心年龄在 50~80 岁的轻中度高血压患者共 263 例，随机分为拉西地平组和苯磺酸氨氯地平组，于治疗 20 周后比较两组患者 24 h 平均收缩压，评价拉西地平控制中老年原发性轻中度高血压的效果是否不劣于苯磺酸氨氯地平。见表 1。

表 1 拉西地平组与苯磺酸氨氯地平组 24 h 平均收缩压变化情况(mmHg)

组 别	n	$\bar{x}$	s
拉西地平组	132	15.2	16.3
苯磺酸氨氯地平组	131	15.5	13.1

基于该成组设计一元定量资料，能够计算得到样本量、均数及标准差，为了比较两组的 24 h 平均收缩压控制效果，可以采用非劣效性检验(假定：经临床专家商定，非劣效界值为-5 mmHg)，结合两组的均数、标准差判断该药物是否有推广价值。一般认为，降血压的数值越多，表明降压药的疗效越好，故本例的评价指标(血压下降值)为“高优指标”。

SAS 程序如下：

```
data example1;
n1=132; n2=131;
mean1=15.2; mean2=15.5;
s1=16.3; s2=13.1;
/*第 1 步*/
L=-5.00;
ss1=s1**2*(n1-1);
ss2=s2**2*(n2-1);
sc=(ss1+ss2)/(n1+n2-2);
se=sqrt(sc*(1/n1+1/n2));
/*第 2 步*/
t=((mean1-mean2)-L)/se;
p=1-probt(t, n1+n2-2);
/*下面的 utc 为 t 分布曲线下 95% 的上单侧临界值*/
/*下面的 df 为 t 分布的自由度*/
/*下面的 up 为 t 分布曲线下分位数 utc 左边的概率*/
df=n1+n2-2;
up=1-0.05;
utc=TINV(up, df);
cl=(mean1-mean2)-utc*se;
run;
ods html;
proc print;
var t p utc cl;
run;
ods html close;
```

【程序说明】第 1 步设定非劣效性界值 L(注意：程序中不便采用 δ_L 表示)；然后，根据式(2)计算中间结果。第 2 步进行非劣性假设检验，根据式(1)和式(3)计算 t 值、P 值和 C_L[9]。

【SAS 主要输出结果及解释】

Obs	t	p	utc	cl
1	2.57626	0.005269056	1.65071	-3.31148

以上主要输出结果中的 utc=1.65071 是基于 df=132+131-2=261、上尾概率为 0.05 条件下计算出 t 分布曲线下的分位数，若采用正态分布近似取代 t 分布，则此数值应该为 1.645。

统计与专业结论：t=2.57626，P=0.005269056，按照 α=0.05(单侧检验)，拒绝 H_0，接受 H_1，可以认为拉西地平控制中老年原发性轻中度高血压的效果不劣于苯磺酸氨氯地平。从 95% 置信区间下限来看，C_L=-3.31148>-5.00，可以认为拉西地平不劣于苯磺酸氨氯地平，该结论与假设检验结果一致。

2.2 基于原始定量数据

【例 2】为了观察地赐康的降血脂疗效[10]，将 79 例原发性高血脂症患者随机分为治疗组(地赐康冲剂 1 袋，tid)和对照组(血脂康胶囊 2 粒，bid)，治疗 8 周后，观察两组总胆固醇下降的情况，评价地赐康的血脂调节效果是否不劣于血脂康胶囊。见表 2。

表 2 测定 79 例原发性高血脂症患者 8 周后总胆固醇下降情况(mmol/L)

序 号	治疗组	对照组
1	0.33	3.05
2	1.21	1.77
3	3.26	0.98
…	…	…
25	-0.23	1.32
26	1.13	1.40
…	…	-
52	1.80	-
53	3.20	-

注：详细数据见后面的 SAS 程序

表2所示属于成组设计一元定量资料，可以采用非劣效性检验比较治疗组与对照组的效果。根据临床经验，设定非劣效性界值为-0.52 mmol/L，以进行后续的非劣效性检验，试判断地赐康冲剂是否具有推广和应用价值。

SAS程序如下：

```
/*第1步创建数据集*/
data example2;
input group$ n;
input x @@;
output;
end;
cards;
Group1 53
0.33 1.21 3.26 0.48 1.46 1.92 0.74 1.00 1.11
-0.29 0.72 -0.10 2.71 2.13 1.76 1.18 1.07 1.69
2.59 0.18 2.48 2.04 2.85 2.35 -0.23 1.13 1.53
-0.22 1.75 2.32 1.84 1.67 1.46 0.67 1.62 1.19
2.34 2.87 2.55 2.00 1.27 0.83 1.25 0.87 3.29
1.92 0.38 2.11 1.74 2.51 0.45 1.80 3.20
Group2 26
3.05 1.77 0.98 1.50 1.22 0.85 3.29 1.54 1.19
1.00 1.36 0.27 0.79 1.86 1.53 1.69 2.09 1.50
1.67 2.33 1.95 2.51 1.24 1.98 1.32 1.40
;
run;
/*第2步非劣效性检验*/
/*求均值之差的双侧90.0%置信区间，相当于单侧95.0%置信区间*/
ods html;
proc ttest data=example2 alpha=0.10 sides=u h0=-0.52;
class group;
var x;
run;
ods html close;
```

【程序说明】第一步，根据原始定量数据建立临时数据集“example2”，程序对试验组和对照组有明确的要求，第一组为试验组，第二组为对照组。第二步，在TTEST过程中，sides=u表示采用上单侧检验，h0=-0.52为设定的非劣效性界值；“alpha=0.10”代表求均值之差的双侧90.0%置信区间，相当于单侧95.0%置信区间。

【SAS主要输出结果及解释】

The TTEST Procedure

Variable：x

group	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
Group1		53	1.5279	0.9291	0.1276	-0.2900	3.2900
Group2		26	1.6108	0.6727	0.1319	0.2700	3.2900
Diff(1-2)	Pooled		-0.0828	0.8544	0.2046		
Diff(1-2)	Satterthwaite		-0.0828		0.1836		

group	Method	Mean	90% CL Mean		Std Dev	95% CL Std Dev	
Group1		1.5279	1.3142	1.7417	0.9291	0.8018	1.1100
Group2		1.6108	1.3854	1.8361	0.6727	0.5481	0.8799
Diff(1-2)	Pooled	-0.0828	-0.3473	Infty	0.8544	0.7554	0.9862
Diff(1-2)	Satterthwaite	-0.0828	-0.3205	Infty			

以上结果分别为两组总胆固醇变化量的均值以及它们差值的均值、标准差、标准误、双侧90.0%（相当于单侧95.0%）置信区间等信息。

Method	Variances	DF	*t* Value	Pr>t
Pooled	Equal	77	2.14	0.0179
Satterthwaite	Unequal	65.932	2.38	0.0101

Equality of Variances

Method	Folded F	Den DF	F Value	Pr > F
Folded F	52	25	1.91	0.0811

根据方差齐性检验的结果（Equality of Variances），F=1.91，P>0.05，认为两总体方差相等。对应的t检验结果中，应该参照汇总方法（Pooled），对应方差相等时的计算结果。

若采用置信区间法，需要参考两组效应差值的95% C_L Mean的下限，再与δ_L（取负值）进行比较，得出统计结论。

统计与专业结论：t=2.14，P=0.0179，按照α=0.05，拒绝H_0，接受H_1，可以认为地赐康的血脂调节效果不劣于血脂康胶囊，结合地赐康冲剂价格以及患者的服药依从性，可以考虑在临床推广应用。此外，按照单侧95%的置信度，两组效应差值的置信区间下限C_L为-0.3473>-0.52（非劣效界值），可以下非劣效性的结论，与假设检验的结果一致，即可以认为地赐康的血脂调节效果非劣效于血脂康胶囊的血脂调节效果。

3 讨论与小结

3.1 讨论

在非劣效性分析中，先验确定了两种治疗方法之间的可接受的差异δ_L，但非劣效性界值的设定是研究设计时最复杂的问题。由于确定的非劣效界值一般很小，导致非劣效设计中阳性对照的样本量需求大于安慰剂对照，而样本量的确定高度依赖于界值和试验方法的效应，这些都必须明确而具有现实性[11]。关于非劣效或者等效性研究的文献有很多混乱及缺陷，如界值的正确选取、样本量的合理计算等。造成这一问题的重要原因是术语缺乏统一性和通透性，这是应用新方法时不可避免的。但可以预料的是，随着这些方法的应用和研究指南的不断改善，科学研究将会向着正确的方向更进一步[12]。

3.2 小结

少数情况下当安慰剂对照不被允许或违反伦理，或者想要与已上市的有效药物或标准治疗方案进行比较以求能获得一个新的治疗选择时，可以考虑进行非劣效性分析。这种类型的分析存在一些缺点，如依赖于之前对照的结果以及确定非劣效界值较复杂。但是这些缺点可能会被其优点所抵消，如能够包含多种终点类型（二分类、有序变量、连续变量等），基于治疗方法和适应症，能够比安慰剂对照更具有伦理学意义。

随着非劣效设计的普遍应用，需注意在研究过程中遵守和维持此类有效性研究的基本原则和报告标准。通过合理的设计和执行，非劣效性试验能够提供具有临床价值的创新治疗方案。

参考文献

[1] 谷恒明，胡良平. 新药临床试验设计中的比较类型[J]. 四川精神卫生，2017，30(4)：317-322.

[2] Bermingham EC，del Castillo JRE，Radecki SV. The use of the noninferiority analysis in clinical studies [J]. Equine Vet J，2014，46(4)：399-401.

[3] Mauri L，D'Agostino RB. Challenges in the design and interpretation of noninferiority trials [J]. N Engl J Med，2017，377(14)：1357-1367.

[4] Piaggio G，Elbourne DR，Pocock SJ，et al. Reporting of noninferiority and equivalence randomized trials：extension of the CONSORT 2010 statement [J]. JAMA，2012，308 (24)：2594-2604.

[5] 方积乾. 卫生统计学[M]. 7版. 北京：人民卫生出版社，2012：321-323.

[6] 李新旭，唐智敏，辛晓娜，等. 对药物临床试验中置信区间法预设检验界值的考虑[J]. 中国临床药理学杂志，2019，35(19)：2487-2493.

[7] 陈卫，徐利娜，迭敏，等. 差异性、等效性、非劣效性和优效性设计中的t检验[J]. 成都医学院学报，2009，4(3)：211-213.

[8] 霍勇，张慧敏，葛均波，等. 拉西地平与苯磺酸氨氯地平治疗中老年轻中度原发性高血压的对比分析[J]. 中国新药杂志，2019，28(8)：967-972.

[9] 胡良平. SAS常用统计分析教程[M]. 2版. 北京：电子工业出版社，2015：274-275.

[10] 唐巍，崔洪滨. 地赐康治疗高脂血症的临床观察[J]. 中国新药杂志，2002，11(4)：306-308.

[11] Kaji AH，Lewis RJ. Noninferiority trials：is a new treatment almost as effective as another? [J]. JAMA，2015，313 (23)：2371-2372.

[12] Walker E，Nowacki AS. Understanding equivalence and noninferiority testing [J]. J Gen Intern Med，2011，26 (2)：192-196.

（收稿日期：2020-05-26）

（本文编辑：陈 霞）