
主要终点指标的选择与样本量计算

作者：褚红玲 赵一鸣 来源：临床流行病学和循证医学

本文原地址：<https://www.iikx.com/news/statistics/1919.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

前几天，一位临床的小伙伴带着他的标书来到办公室，让小编帮忙计算个样本量。小编看了下他的研究设计，一堆指标。小编一脸蒙圈，“您能将研究目的更精确一些，给出一个最关键的评价指标吗？”小伙伴一听，更是二脸蒙圈，“我都想评价啊，我觉得这几个都很重要”。几经讨论，终于达成共识，选择了一个相对最重要的评价指标计算了样本量。其实这种情况，搞临床研究或者统计的小伙伴们应该是常常碰到的了。

首先，不同的终点指标，将会需要不同的样本量来达到我们期望的Power;再次，在有限的预算、资源和精力条件下，只允许完成某个限定样本量的研究，那么如何选择一个终点指标，才能有效的回答研究问题就很重要。所以，合适的终点指标的选择可以增加研究目标的成功率。

那么，终点指标都有哪些类型呢

。一般来讲，主要有三种类型：连续变量，如血压、血脂、血糖等;二分类变量，如治愈率、死亡率、事件发生率等;还有一类，时依性变量

，比如在很多肿瘤相关的研究中，除了死亡、复发、进展或事件发生与否之外，还需要结合生存时间。

二分类变量和时依性变量相对比较固定，没有太多可变动的空间，就在于如何根据研究期限及疾病进展程度来选择合适的变量。而连续变量可能门道就多点了。下面我们通过一个例子来看看主要终点指标的类型对样本量的影响：

某研究者计划评价药物A(试验药)和药物B(阳性对照药)降压的效果。研究者假设药物A的降压效果不劣于单用药物B(对照药)，计划以6个月血压达标率(定义为治疗6个月，收缩压 $<140\text{mmHg}$)为主要终点。假设药物A的血压达标率为75%，药物B的达标率为80%，非劣效界值设为10%，那么样本量需要多少?我们使用PASS软件来计算一下：

Select a Procedure

Category + -

Favorites Recent Show All

- Normality
- Proportions**
- One Proportion
- Two Independent Proportions
 - Test (Inequality)
 - Test (Non-Zero Null)
 - Non-Inferiority**
 - Equivalence
 - Confidence Interval
 - Repeated Measures
 - Group-Sequential
- Two Correlated Proportions
- Two Proportions - Cluster Random
- Contingency Table (Chi-Square)
- Repeated Measures
- Stratified

Proportions > Two Independent > Non-Inferi

Non-Inferiority Tests for Two Proportions [Differences]

Non-Inferiority Tests for Two Proportions [Proportions]

临床流行病学和循证医学

File View Run Procedures Tools Window Help

Reset Open Save As Home Favorites Recent Loaded Output Gallery

Run

Data

Options

Reports

Plot Setup

Data

Solve For

Find (Solve For): N1

Error Rates

Power (1-Beta): .80

Alpha (Significance Level): .025

Sample Size

N1 (Sample Size Group 1, Treatment): 50 to 400 by 50

N2 (Sample Size Group 2, Reference): Use R

R (Sample Allocation Ratio): 1.0

Effect Size

Treatment (Group 1)

P1.0 (Non-Inferiority Proportion): 0.7

P1.1 (Actual Proportion): 0.75

$P1.0 = P1|H0, P1.1 = P1|H1$

Reference (Group 2)

P2 (Reference Group Proportion): 0.8

Test

Test Type: Test (Unpooled)

Higher Proportions Are: Better

Add This

File View Edit Window Help

Save As Print Copy Find Add Output to Gallery Auto Add Home Favorites Recent Loaded Gal

Navigation Pane

- Power Analysis of Non-Inferiority Tests of
 - Numeric Results
 - References
 - Report Definitions
 - Summary Statements
 - Chart Section

Power Analysis of Non-Inferiority Test
Numeric Results for Non-Inferiority Tests Based on the Diff
 H0: $P1-P2 \leq D0$. H1: $P1-P2 = D1 > D0$. Test Statistic: Z test (ur

	Sample Size Grp 1	Sample Size Grp 2	Grp 2 Prop P2	Non-Inf. Grp 1 Prop P1.0	Actual Grp 1 Prop P1.1
Power	N1	N2			
0.8000	1091	1091	0.8000	0.7000	0.7500

Note: exact results based on the binomial were only calculated

References
 Chow, S.C.; Shao, J.; Wang, H. 2003. Sample Size Calculation
 Farrington, C. P. and Manning, G. 1990. Test Statistics and Sample
 Trials with Null Hypothesis of Non-Zero Risk Difference or No
 Medicine, Vol. 9, pages 1447-1454.

样本量太大，木有那么多钱啊?怎么办呢?你给我少算点吧。行，那你换一个终点评价指标把，或者换一个形式。比如说直接使用血压下降值呢?假设使用药物A治疗6个月时，血压下降 $20\text{mmHg} \pm 20\text{mmHg}$ ，而使用药物B血压下降 $25\text{mmHg} \pm 20\text{mmHg}$ ，非劣效界值设为 5mmHg ，那么样本量需要多少呢?我们还是使用PASS软件来计算一下：

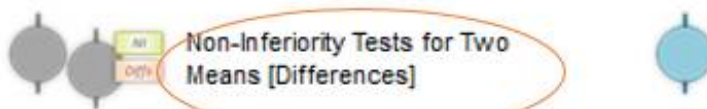
Select a Procedure

Category

Favorites Recent Show All

- Cluster-Randomized
- Confidence Intervals
- Correlation
- Design of Experiments
- Diagnostic Tests (ROC)
- Equivalence
- Group-Sequential
- Means**
- One Mean
- Paired Means
- Two Independent Means
 - Test (Inequality)
 - Test (Non-Zero Null)
 - Non-Inferiority**
 - Equivalence
 - Confidence Interval

Means > Two Independent Means > Non-Inf



临床流行病学和循证医学

Run

Data

Data

Iterations

Reports

Plot Setup

Solve For

Find (Solve For): N1

Error Rates

Power (1-Beta): .80

Alpha (Significance Level): .025

Sample Size

N1 (Sample Size Group 1, Treatment): 5 to 45 by 10

N2 (Sample Size Group 2, Reference): Use R

R (Sample Allocation Ratio): 1.0

Effect Size

Mean Difference

NIM (Non-Inferiority Margin): 5

D (True Difference, Trt Mean - Ref Mean): 5

Standard Deviations

S1 (Standard Deviation Group 1, Treatment): 20

S2 (Standard Deviation Group 2, Reference): 20

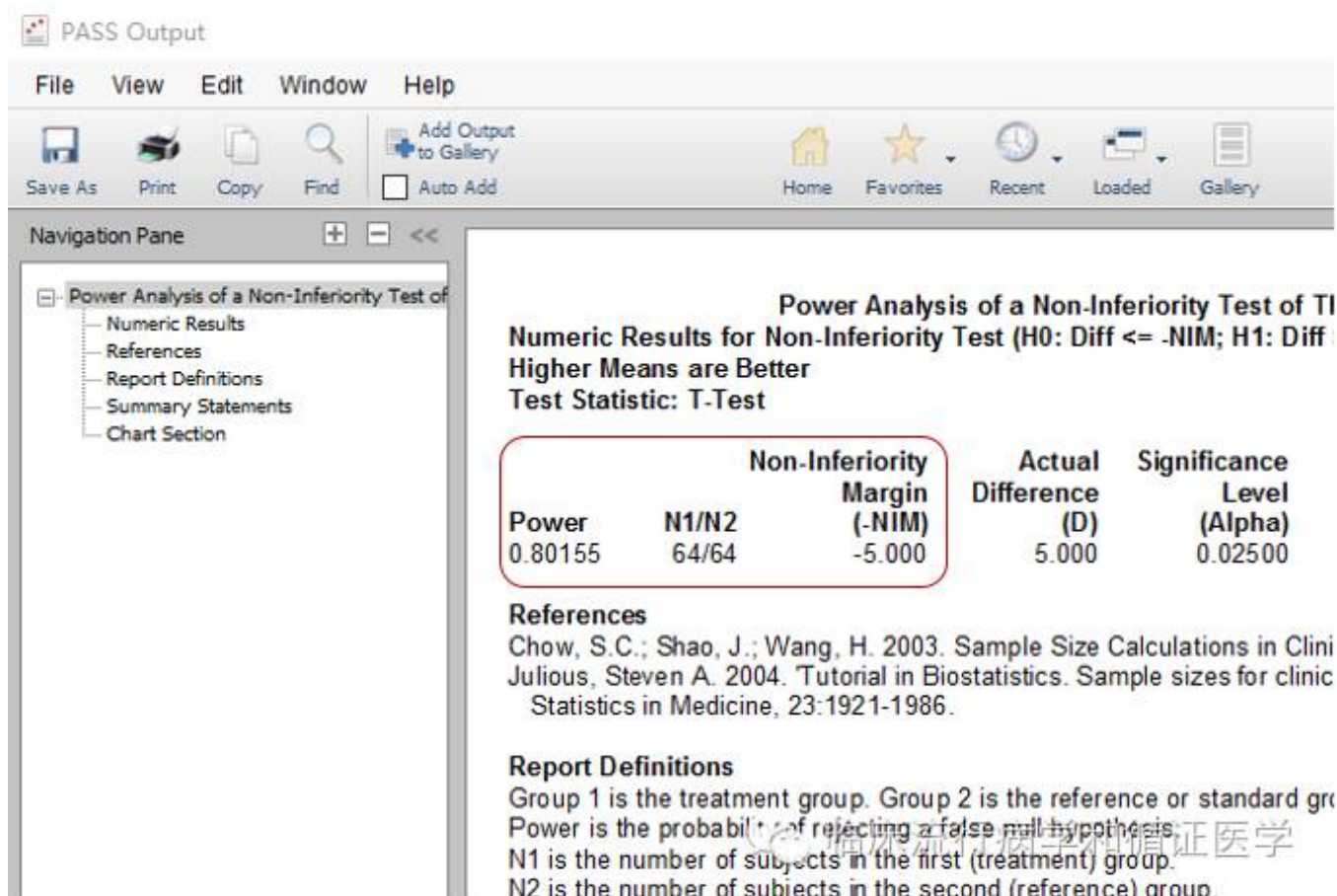
Test

Higher Means Are: Better

Nonparametric Adjustment: Ignore

Add This

临床流行病学和循证医学



看看这两组的样本量，相差巨大。当然，最终到底选择谁作为研究的主要终点指标，还得看研究者对自己所定义的非劣效界值，假设的各组的达标率或者血压下降值有确凿的证据和临床经验，才能在研究结果中有足够的Power检测出两组之间的统计学差异。

总之，主要终点指标的选择与样本量关系极其密切，除了科学估算，还要综合考虑医院病人的分布情况、研究实施的可行性及经费等多方面的因素，来决定主要终点指标。

更多 统计方法 请访问 <https://www.iikx.com/news/statistics/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](http://www.iikx.com)转发