

• 论 著 •

实验设计中样本含量估算方法探讨 ——以 PI3K 抑制剂对小鼠气道炎症影响的实验研究为例*

喻宁芳¹, 喻宁芬²

(1. 广州市第一人民医院南沙医院检验科, 广东广州 511457; 2. 广州市第一人民医院儿科, 广东广州 510180)

摘要:目的 介绍和比较医学实验设计中不同的样本含量估算方法。方法 以磷酸酰肌醇-3 激酶(PI3K)抑制剂对小鼠气道炎症影响的实验研究为例, 运用不同方法计算样本含量。结果 公式法计算需 12 只, PASS 软件 Simple 法计算需 8 只, Stata 软件计算需 10 只, 验算其检验效能: $1-\beta > 0.9$ 。结论 3 种不同方法估算的样本含量都是合理有效的, 实验研究人员可以以多种计算结果为依据, 分析实验研究性质, 综合考虑研究成本、可行性与伦理学要求对样本含量的影响, 确定合适的样本数。

关键词: 样本估算; 实验设计; PASS 软件; Stata 软件

DOI: 10.3969/j.issn.1673-4130.2015.09.012

文献标识码: A

文章编号: 1673-4130(2015)09-1193-03

Discussion on different estimation methods of sample size in experiment design

——taking an experimental study on influence of PI3K inhibitor on mouse respiratory tract inflammation as an example

Yu Ningfang¹, Yu Ningfen²

(1. Department of Clinical Laboratory, Nansha Hospital, Guangzhou Municipal First People's Hospital, Guangzhou, Guangdong 511457, China; 2. Department of Pediatrics, Guangzhou Municipal First Municipal People's Hospital, Guangzhou, Guangdong 510180, China)

Abstract: **Objective** To introduce and compare the different estimation methods of sample size in the medical experiment design. **Methods** Taking the experimental study on the influence of PI3K inhibitor on respiratory tract inflammation in mouse. **Results** The formula method estimation needed 12 mice, while the PASS software Simple method needed 8 mice, and the stata software method needed 10 mice. The power of test was checked: $1-\beta > 0.9$. **Conclusion** The sample size estimated by 3 different kinds of method is rational and effective. The experimental researchers can take the multiple calculated results as the basis to analyze the property of experimental study, comprehensively consider the influence of the research costs, feasibility and ethics requirements on the sample size and to determine the most appropriate sample number.

Key words: sample estimation; experiment design; PASS software; Stata software

样本含量, 又称样本容量、样本大小, 是指在实验研究和调查研究中, 每个样本所包含的观察对象的数量。样本含量估算是医学实验设计中重要的部分, 在科学研究的过程中扮演着重要角色^[1]。样本过大固然造成资源的浪费, 而样本过少则实验统计效能不高, 失去统计学意义。实践中样本含量估计的常用方法有: 经验法(由查文献或咨询专家确定)、查表法和公式计算法^[2]。近 20 年来软件开发公司开发了种类繁多、简单易用的统计软件, 但由于缺乏经验交流, 医学工作者在医学实验研究中较少使用统计软件估算样本含量。本文以广东省科技计划项目课题一磷酸酰肌醇-3 激酶(PI3K)对小鼠气道炎症影响的实验研究为案例, 介绍公式法、PASS 软件 Simple 法和 Stata 软件计算法这 3 种样本含量估算方法的实际应用, 并通过实验所得数据验证其检验效能, 判断 3 种方法确定的样本含量是否合理有效, 为医学工作者实验设计时估算样本含量提供一定实践经验借鉴。

1 材料与与方法

1.1 材料

1.1.1 材料 Balb/c 小鼠: 由广州医学院动物中心提供, 4~6

周龄, 雄性, 清洁级, 体质量 20~25 g, 无卵清蛋白(OVA)饮食。致敏液: 由 0.08% OVA 0.1 mL(美国 Sigma 公司)和等体积液态铝(美国 Pierce 公司)混合。LY294002(PI3K 抑制剂): 美国 Sigma 公司提供。

1.1.2 动物分组 分组方法: 随机数字表法。组别: OVA 组、LY294002 组和对照组。

1.1.3 实验方法 OVA 组: 本组小鼠于第 0、7、14 天腹腔注射致敏液 0.2 mL, 第 15 天将小鼠置于透明密闭容器中以 1% OVA 溶液 10 mL 雾化吸入, 每次 20 min, 隔日一次, 连续 5 周。LY294002 组: 同期同时给予注射致敏液和吸入 OVA, 第 13 天开始经尾静脉给予 LY294002, 用量为 7.5 mg/kg, 连续 3 d 每天 1 次, 第 15 天于吸入 OVA 前 30 min 给药。对照组: 同期同时给予相同体积生理盐水腹腔注射及雾化吸入。观察指标: 支气管肺泡灌洗液(BALF)中细胞分类计数、骨髓悬液和肺组织学检查。通过实验检测 LY294002 对 OVA 致敏哮喘小鼠肺组织病理、BALF 和骨髓中细胞分类的影响, 计数细胞总数和嗜酸性粒细胞(Eos)并进行统计学分析^[3]。

1.1.4 实验目的 运用公式法、PASS 软件 Simple 法和 Stata

* 基金项目: 广东省科技计划项目(73092); 广州市医药卫生科技资助项目(2007-YB-026)。 作者简介: 喻宁芳, 女, 副主任技师, 主要从事临床生化检验工作。

软件计算法求得本实验的样本含量,并通过实验所得数据验证其检验效能,判断 3 种方法确定的样本含量的有效性。

1.2 方法

1.2.1 公式法介绍 $n = [(Z_{\alpha/2} + Z_{\beta}) * \sigma / \delta]^2$, n 为所需要的样本含量, δ 为总体差值, σ 为总体标准差, $Z_{\alpha/2}$ 为标准正态分布的双侧临界值; Z_{β} 为正态分布的单侧临界值。

1.2.2 PASS 软件介绍 PASS 是样本含量估计和效能分析中常用的一款优秀的统计分析软件^[4]。它界面良好、功能齐全,可以进行描述性统计、相关及回归分析、实验设计、生存及可靠性分析、统计图表绘制等操作,只需要输入相应的参数,即可实现对样本含量或检验效能的预测。常用方法有 Compound Symmetry 法、AR 法、Banded 法和 Simple 法。本实验研究选用 Simple 法。

1.2.3 Stata 软件介绍 Stata 是一个功能强大的统计分析软件,它可以进行 t 检验、参数估计、协方差分析、单因素和多因素的方差分析、方差齐性检验、缺项数据的处理、正态性检验等一般分析。它采用具亲和力的窗口接口,操作灵活简便、学习方便,使用者可以通过 Stata 官方网站学习使用方法。该软件用于样本含量和检验效能估计的主要命令是 sampsi (sample size and power),命令格式为: . sampsi #1 #2[,一般选择项][重复测量选择项]。#1 表示处理前测量的均数,#2 表示处理后测量的均数。一般选择项包括:检验效能、检验水平、 $n1$ 与 $n2$ 样本量的比值、 $sd1/sd2$ (sd 为标准差)、单侧/双侧检验、单样本(缺省时为两样本比较)。常用方法有 change 法、post 法和 ancova 法,本试验研究选用 ancova 法。

2 结果

2.1 参数设置 I 类错误概率大小 α 越小,所需要的样本含量越大,通常情况下 α 取 0.05,可取单侧或双侧用统计学检验。选择双侧检验的条件是研究结果高和低于效应指标的界限均有意义,所需样本量就大;选择单侧检验条件是研究结果仅高或低于效应指标的界限有意义,所需样本量就小。II 类错误概率大小 β 越小,检验效能 $1-\beta$ 越大,所需样本量也越大,一般要求检验效能不低于 0.80,一般只取单侧,本实验取 $\beta=0.1$ 。 $n1$ 与 $n2$ 样本量的比值常用的为 4:1、2:1、1:1,考虑成

本最低取 1:1。总体标准差 σ 或总体率 π ,常根据预实验及前人的研究结果或统计理论进行估计,根据何胜东等^[5]的研究结果和其他资料,取 $\pi=0.7, \sigma=7.87$ 。容许误差 δ 是指研究者要求的或客观实际存在的样本统计量与总体参数间或样本统计量间的差值,本实验认为 Eos 百分比降低 80% 以上为有效,取 $\delta=0.8\sigma$ 。

2.2 计算结果 公式计算值为 12 只, PASS 软件 Simple 法计算值为 8 只(见图 1), Stata 软件计算值为 10 只(由于软件界面图较复杂,本文不截取界面图)。

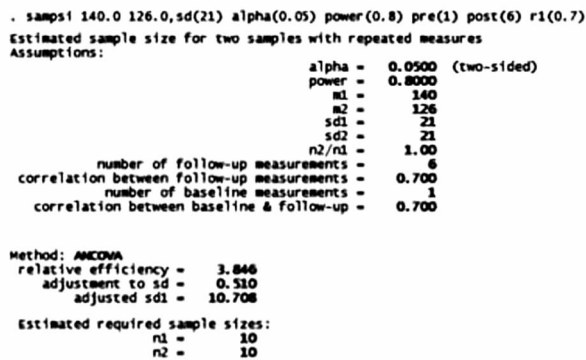


图 1 Stata 软件计算结果图

2.3 实验结果 本实验研究向广州医学院病理教研室咨询了实验材料的费用,根据以往的文献研究经验并考虑统计方便,最终确定样本含量为 10 只。但考虑样本的非正常失效(如取样失败、死亡等),笔者增加 2 只作为备份,将样本编号为 1~12 号。本实验在实验过程中未出现样本的非正常失效,12 只小鼠都取得了真实有效的实验数据。实验研究结束后,笔者验证了 LY294002 组对 OVA 组的检验效能,样本数为 8 只取样本编号为 1~8 号的数据,定为 A 组;样本数为 10 只,取样本编号为 1~10 号的数据,定为 B 组;样本数为 12 只,取样本编号为 1~12 号的数据,定为 C 组。其计算结果:所有组别 $1-\beta > 0.90$ (数据见表 1),说明这 3 种方法计算的样本数都是合适有效的。

表 1 OVA 组、LY294002 组细胞总数和 Eos 的检测结果

组别	样本含量(n)	OVA 组细胞总数 ($\times 10^4$ L)	LY294002 组细胞 总数($\times 10^4$ L)	OVA 组 Eos 总数($\times 10^4$ L)	LY294002 组 Eos 总数($\times 10^4$ L)	$1-\beta$
A 组	8	144.55 \pm 14.32	77.60 \pm 10.01	39.30 \pm 4.40	4.50 \pm 0.55	0.91
B 组	10	144.34 \pm 14.22	77.48 \pm 9.79	39.11 \pm 4.37	4.39 \pm 0.52	0.959
C 组	12	144.45 \pm 14.30	77.50 \pm 9.90	39.19 \pm 4.40	4.45 \pm 0.55	0.95

3 讨论

医学试验研究设计包含专业设计和统计设计两部分内容,统计设计对于保证研究结果的可靠性、科学性、重现性,具有重要的意义^[6]。样本含量估计与检验效能估算是统计设计中最重要的环节之一。只有科学地确定样本含量才能确保研究的可靠性和研究结果的可信性。样本含量的估计原则:在保证试验研究结果具有一定可信度($1-\alpha$),又具有一定检验效能($1-\beta$)的前提下,估算出能够达到主要研究目标所需要的研究

对象最小例数,以便通过样本研究结果来推断总体特征。其中 α 和 β 的取值大小是由试验设计人员希望达到的可信性和检验效能而确定。有时试验者会考虑到样本意外丢失(如动物的非正常生病或死亡,人员的失访等),会在估算出的最少样本例数上比例不能太大,增加 10%~20% 即可,因为盲目追求大样本量可能导致更多混杂因素的产生(如动物个体体质相差太大等),导致更大或更多的偏倚发生。

间重复序列中的两个重要技术^[8],利用细菌基因组中广泛存在的短重复序列,通过电泳条带比较分析,揭示基因组间的差异。在本研究中,这 3 种基因分型方法对鲍曼不动杆菌的基因分型结果显示:(1)每种方法 3 次扩增所得出的条带数量及相对分子质量大小较一致,说明在本试验条件下,3 种方法具有较好的可重复性。(2)试验中 REP-PCR 和 RAPD 虽能够满足分型需求,但扩增条带并不丰富,分析可能原因是技术水平的限制,以及试验中仅对扩增条件做了优化而未对扩增体系进行优化。(3)3 种基因分型方法对个别菌株的分型存在差异,可能是因为 3 种方法作用于基因组 DNA 锚定位点的不同所致^[9],但并不影响整体的分型结果。有文献报道,在鲍曼不动杆菌的基因分型中,ERIC-PCR 分辨率较强,结果较为可靠,并且与“金标准”PFGE 有一定的相关性,说明 ERIC-PCR 有其自身的优势,本研究亦证明其较另外两种方法有较好的分型效果。30 株菌株经聚类分析后,得到与电泳指纹图谱分型相一致的结果,说明在不进行统计分析的情况下,根据电泳条带的数量和相对分子质量大小进行分类亦可达到较好的分型结果。

研究结果表明,多重耐药的 I 型流行株已在该院 ICU 病区广泛存在,且 I 型流行株的多重耐药谱较为相似,分析原因可能为携带多种耐药基因的接合性质粒等遗传元件在不同菌株之间的传递从而导致多重耐药性的传播,对此,还需进一步试验证实。

综上所述,该院 ICU 病区存在鲍曼不动杆菌多重耐药现象,头孢哌酮/舒巴坦对其依然有较好治疗效果。初步试验证明,3 种 PCR 指纹图谱基因分型方法均可作为简便、经济、快捷的基因分型方法对鲍曼不动杆菌进行分型。其中,ERIC-PCR 扩增条带较多,分辨率较好,但本研究仅局限于 30 株菌株,还有待扩大试验菌株数量,同时结合“金标准”分型方法——PFGE 进一步比较分析,以得出更为准确的结论。

(上接第 1194 页)

文献法和专家咨询法等方法估算样本含量,这些方法主观因素对估算结果影响大,而统计软件由于操作简单,考虑的客观因素多,计算结果相对合理。统计软件除了本文介绍的 PASS 软件和 Stata 软件外,还有 nQuery Advisor 软件、SamplePower 软件(SPSS 公司研发)、SASA 软件和 SAS 软件等,使用方法都大同小异。本文采取的 3 种方法计算结果不相同但却都是合理有效的,因为医学试验研究中样本含量不是唯一的,不同的研究方法、研究目的,研究要求和研究资料决定了不同的样本含量,从表 1 中可看出样本含量也并不是越大越精确。这些样本含量估算方法参考了研究个体的变异度、研究结果的精确度(抽样误差),但未考虑研究成本、可行性与伦理学要求对样本含量的影响。

总之,估算样本含量的方法有许多,统计软件具有更客观、更全面和简单方便的优势,在参数选择正确的情况下其计算结果都是合理有效的,试验研究人员要熟练掌握和运用,并以其计算结果为依据,正确分析试验研究性质,再综合考虑研究成

参考文献

- [1] 王璐,任微,褚美玲,等.耐碳青霉烯的多重耐药鲍曼不动杆菌分子流行病学研究[J].现代检验医学杂志,2010,25(1):87-89.
- [2] 李永丽,应春妹,陈艺升.ERIC-PCR 技术在鲍曼不动杆菌基因分型中的应用评估[J].检验医学,2013,28(7):621-624.
- [3] Clinical Laboratory Standards Institute. M100-S23 Twenty-third Informational Supplement[S]. Wayne, DA, USA: CLSI, 2013.
- [4] Grundmann HJ, Towner KJ, Dijkshoorn L, et al. Multicenter study using standardized protocols and reagents for evaluation of reproducibility of PCR-based fingerprinting of *Acinetobacter* spp [J]. J Clin Microbiol, 1997, 35(12):3071-3077.
- [5] Maluping RP, Ravelo C, Lavilla-Pitogo CR, et al. Molecular typing of *Vibrio parahaemolyticus* strains isolated from the Philippines by PCR-based methods[J]. J Appl Microbiol, 2005, 99(2):383-391.
- [6] Zhong Q, Xu W, Wu Y, et al. Clonal spread of carbapenem non-susceptible *Acinetobacter baumannii* in an intensive care unit in a teaching hospital in China[J]. Ann Lab Med, 2012, 32(6):413-419.
- [7] Schleicher X, Higgins PG, Wisplinghoff H, et al. Molecular epidemiology of *Acinetobacter baumannii* and *Acinetobacter nosocomialis* in Germany over a 5-year period (2005-2009)[J]. Clin Microbiol Infect, 2013, 19(8):737-742.
- [8] Ishii S, Sadowsky MJ. Applications of the rep-PCR DNA fingerprinting technique to study microbial diversity, ecology and evolution[J]. Environ Microbiol, 2009, 11(4):733-740.
- [9] 张永,陈余清,唐英春,等. ARDRA 联合 RAPD 对不动杆菌基因型鉴定的研究[J].微生物学通报,2007(2):303-306.

(收稿日期:2015-01-21)

本、可行性与伦理学要求对样本含量的影响,最终确定合适的样本数。

参考文献

- [1] 吴圣贤,王成祥.临床研究样本含量估算[M].北京:人民卫生出版社,2008:1-2.
- [2] 常靖,常亮.单因素重复测量设计样本含量的估算及不同计算方法之间的比较[J].数理医药学杂志,2012,25(5):505-508.
- [3] 喻宁芬,喻宁芳,邓礼,等.磷酸酰肌醇-3 激酶抑制剂对小鼠气道炎症影响的实验研究[J].现代医院,2008,3(4):7-9.
- [4] Dattalo P. A review of software for sample size determination[J]. Eval Health Prof, 2009, 32(3):229-248.
- [5] 何胜东,赖克方,姚卫民,等.小鼠哮喘模型气道反应性检测方法的建立[J].广东医学,2006,27(11):1659-1661.
- [6] 李河,李卫,杨学宁,等.临床试验设计中样本含量的理解[J].循证医学,2012,12(6):374-376.

(收稿日期:2014-12-25)