

•学术探讨•

中药临床试验样本量合理优化的探讨

于旭华, 闫 蓉, 苏如玉, 杨 从, 章时杰, 许卫华, 王 奇(广州中医药大学临床药理研究所, 广东 广州 510405)

摘要: 中药临床试验的设计经常涉及到样本量的计算问题。样本量不足则检验效能过低, 不易检出已存在的阳性结果, 影响试验结果的准确性; 样本量过大会造成资源浪费和执行困难, 而且不适宜的样本量会有悖于伦理道德。本文比较了大样本试验和小样本试验的优缺点, 并从多个角度探讨临床试验样本量设计的优化思路, 为中药临床试验设计的样本量选择提供依据。

关键词: 中药临床试验; 样本量; 检验效能; meta 分析

中图分类号: R285.6 **文献标志码:** A **文章编号:** 1003-9783(2013)03-0326-03

doi: 10.3969/j.issn. 1003-9783.2013.03.031

Method of Sample Size Design Optimization

YU Xuhua, YAN Rong, SU Ruyu, YANG Cong, ZHANG Shijie, XU Weihua, WANG Qi (Institute of Clinical Pharmacology, Guangzhou University of Chinese Medicine, Guangzhou 510405 Guangdong, China)

Abstract: Sample size of Chinese medicine is a matter in design of clinical trails. The optimal sample size in a trial can assure the testing power to detect statistical significance. But using too many participants in a trial is expensive and is hard to carry out for researchers. Conversely, if the trial is underpowered for too few participants, it will be statistically inconclusive and may make the whole protocol a failure. So both too small and too large sample size are unethical. In this study we compare the advantages and disadvantages of different sample size, and present some method of design optimization, thus to provide a basis for the design of clinical trail of Chinese medicine.

Keywords: Clinical trails of Chinese medicine; Sample size; Testing power; Meta analysis

临床试验的样本量估算一直是倍受临床医生以及相关科研人员关注的问题。1978 年 Freiman、Chalmers 等人发表了关于讨论样本量和检验效能的文章^[1], 一时间产生了轰动性的效应。由于样本量不足(尤其对照组)所造成的检验效能过低, 影响阳性结果的检出率成为当时样本量问题的焦点。随着时间的推移, 有学者相继发表类似文章^[2-4], 质疑小样本研究的科学价值, 且至今仍被学者广泛认同。因此, 样本量估算成为临床试验中不可忽视的问题。同样, 在中药的临床试验中, 样本量的估算也是一个值得重视的环节, 并且因中药临床试验的特点而面临着新的挑战。

1 样本量估算的重要性及面临的问题

1.1 样本量与检验效能的关系 很多学者认为, 样本量不足主要导致的问题在于降低了检验效能, 即难以发现或不能发现客观存在的差别。那么, 样本量与检验效能是怎样的关系呢? 从统计学角度讲, 影响样本量估计的参数主要包括 I 型错误 α 、II 型错误 β 以及检验效能 $(1-\beta)$ 。以成组设计两样本均数比较样本含量估计为例, 样本量计算公式如下:

$$N = \left[\frac{(Z_{\alpha/2} + Z_{\beta}) \sigma}{\delta} \right]^2$$

其中 $Z_{\alpha/2}$ 为标准正态分布的双侧临界值, Z_{β} 为标

收稿日期: 2013-03-01

作者简介: 于旭华, 女, 博士研究生, 研究方向: 临床流行病学、脑病的临床与实验研究。Email: 87994822@qq.com。通讯作者: 王奇, 教授, 博士生导师, 研究方向: 临床流行病学、脑病的临床与实验研究。Email: wangqi@gzucm.edu.cn。

基金项目: 广东省重大科技专项(2012A080202017); 广东省科技计划项目(2010A30100009)。

准正态分布的单侧临界值, σ 为标准差, δ 表示两组计量资料有临床意义的最小差别。由公式可知, 当 α 取较小值时, Z 值越大, 所需样本量越大, 意味着假阳性错误的发生率越小, 样本量则越大, 反之亦然; 同理, 当 β 取较小值时, Z 值越大, 意味着假阴性错误发生概率越小, 试验把握度 $(1-\beta)$ 越大, 所需样本量也越大。所以我们很容易理解, 当样本量不能够满足需求时, 即使 $P < 0.05$, 也是以通过降低统计效能为代价的。因此, 样本量不足将导致统计上的不确定性或试验的失败^[5]。反之, 当所研究样本量极大, 为研究对象总体时, 得出的结论即符合总体的实际情况, 结果把握度最大, 且不存在 I 型错误和 II 型错误的发生。然而在实际临床研究中, 上述情况难以实现。

1.2 大规模临床试验面临的问题 近年来不少学者提出了大规模的临床试验的概念, 因其在样本含量上更接近总体, 试验的把握度更大。世界上一些较有实力的公司及组织进行了一些大规模的临床试验^[6], 为揭示临床实际情况及药品应用效果确实提供了有力的证据。然而, 大规模临床试验的实施面临着多重困难, 无论人力、物力, 或者受试者的募集方面耗费资源都比较大, 并且无法由一般的组织或个人单独完成。与此同时, 一些学者对大规模的临床试验在试验结果的准确性及应用性也提出了质疑^[8-9], 有些学者认为大规模临床试验在经济利益的驱动下, 将会存在严重的发表性偏倚。

虽然足够的样本量的研究可以保证阳性结果的检出率, 确保统计学意义, 拥有更加直接的科学价值。但样本量并非越大越好。从经济学的角度讲, 太大的样本量将耗费过多的人力和财力; 并且没有更大的科学意义和统计学需要; 从伦理学的角度讲, 让更多的受试者暴露于相对不确定的条件下, 也是悖于伦理道德的。

1.3 小样本临床试验的优势 与大规模临床试验相比, 小样本的临床研究更容易为基层的临床医师所实施。虽然很多人质疑其结果缺乏可靠性, 容易对临床决策带来误导, 但是小样本临床试验依然存在着很多自身的优势, 比如可节省试验资源, 可由普通单位或组织来完成。其次, 小样本临床试验内容覆盖病种广泛, 来源充足。第三, 我们可以通过小样本试验了解更多的还未被证实的临床治疗方法, 从而扩大视野, 为更深入的科学研究提供依据^[9]。

1.4 中药临床试验对样本量的要求 中医在独特生存环境下造就了大量“中药临床试验”, 这些试验以医

案的形式传承下来, 由于其缺乏规范化及足够的样本量而不能成为临床决策的有力依据, 但是却反映了临床实际的基本情况, 拥有重要的参考价值。目前中医药面临着重大的考验, 既需要一定规模的临床试验来确定其有效性和安全性, 又不能放弃对小规模临床报道的深入发掘。因此, 综合多方面考虑, 无论大样本或者小样本, 我们都应尽量使其在有应用意义的前提下选择较小的样本量进行研究。

2 中药临床试验中样本量的优化

2.1 根据试验目的不同适当调整样本量大小 临床试验目的不同, 样本量的选择是可以有差异的。对于大多数大规模的临床试验, 其目的是为了证明某种药物的有效性和安全性, 甚至证明其作用优于目前现有的药物, 试验结果可以作为药物应用于临床。在这种情况下, 试验的样本量要求比较大。目前, 国内外诸多指南的 I 类药物多以该类多中心的大规模临床试验作为基础。小样本的临床试验也有它存在的意义。很多病例报道, 甚至专家或者个人对于自身经验的总结, 实际上都应属于小样本范畴, 尽管其可能并不算完整的试验, 但可以看作是小样本试验极端的例子。个别病案或者一些小样本量的临床试验报道, 虽然在结果上具有不确定性, 但它间接地揭示了某种现象, 往往为临床医生或者科研工作者提供了新的思路。

2.2 不同的研究设计方案或者统计方法所需样本量的大小不同 众所周知, 根据临床研究性质的不同, 可以有不同的试验设计方案。中药临床试验中所涉及的设计类型通常包括前瞻性研究、横断面研究以及病例报告、病例分析等叙述性研究等。对于不同的研究类型, 样本量的估算方法存在着差异^[10-11]。同一种研究的不同试验设计也可以影响样本量的大小。如前后对照试验所需样本量比成组独立样本的试验设计少一半; 交叉设计的样本量仅为独立样本的样本量的 1/4。此外, 试验所期望结果的变量类型也可影响样本量的大小, 例如分类变量与连续变量相比较, 前者需要更大的样本量做支撑^[9]。中医临床试验中还常常涉及到等级相关变量的比较, 如疗效的评分定义为无效、有效、显效等。此时应用卡方检验, 样本量对其影响很大, 交叉列联表中如果有 20 % 以上的格子频数小于 5, 则不适合用卡方检验, 对较大样本, 卡方值应结合列联系数或 Phi 系数修正后, 才可拒绝原假设。

2.3 可以根据具体情况调整统计学参数的设定标准,

进而减少样本量 从统计学角度讲,小样本量试验存在的最大问题在于低检验效能,即试验结果可能没有发现潜在的阳性结果。检验效能越低,发现阳性结果的可能性越小。大多数试验者倾向于发现差别,所以总是认为检验效能越高越好。事实上,并不是所有的试验都需要高检验效能。例如青霉素对革兰氏阳性球菌的灭菌效果已被公认有效,且副作用较少。那么在一项研究新的抗生素的试验中,我们为了减少假阳性结论,可将 α 错误设定得更低($P < 0.01$),此时,因为有效而安全的治疗已经存在了,我们也并不在意错过有效的治疗,检验效能可以设定在80%以下,此时得出的阳性结果将更加有意义。

2.4 运用统计学方法弥补或校正因样本量不足所造成的结果的不确定性 一些统计方法也可以帮助解决小样本量问题,如序贯分析、层次模型等^[12], meta分析也被许多学者推荐。Freiman等^[1]人也认为可以通过 meta 分析将低检验效能的试验合并起来。值得注意的是,应用 meta 分析解决小样本量问题对小样本研究本身提出了更高要求。首先,应严格按照方法学要求在试验中实施随机、盲法、对照等。大多数人所担心的小样本量导致的偏倚,事实上是由拙劣的随机化、不充分的随机隐匿、失败的盲法以及过高的失访率导致的。其实小样本本身并不导致偏倚。偏倚控制得当的同类试验,可以应用 meta 分析进行合并,得出较为准确的结果。第二,小样本试验应报告相关详细信息,方便研究人员归类合并,并对试验进行评价。由于小样本试验来源广泛,试验受到各方面条件的制约,试验的质量参差不齐。为方便利用与评价,小样本试验的报告应包括更多方面的内容,如研究背景、经济基础、地域、研究人群等基本情况,特别对于我们较为关心的检验效能方面应加以描述,置信区间能准确地描述低把握度^[13]。第三,小样本试验应全面报告结果,避免发表性偏倚。meta 分析的前提必须克服发表性偏倚,而小样本试验容易因得出阴性结果而不被发表。存在严重发表偏倚的一类试验结果的综合是没有意义的。因此,无论大样本还是小样本,高质量无偏倚的试验设计及严格按照试验操作规程实施是试验本身实现价值的前提。

3 结语

控制样本量的因素还有很多,如预计失访率、预计治疗的依从性大小、多个结局指标的取舍等^[14]。并且在整个临床试验的进程中,根据试验情况的变化调整样本量也是一种合理可行的方法^[16]。但无论运

用何种策略对样本量进行设计,样本量的选择都应该在根据实际情况而定,盲目追求大样本或忽视样本量对试验结果影响的试验设计都是不合理的。因此,样本量的优化设计至关重要。只有理解了样本量对试验结果的影响,才能运用 meta 分析等统计方法有效地筛选和整合现有文献资源,为科研实验和临床实践提供更加可靠的依据。

参考文献:

- [1] Freiman, JA, Chalmers TC, Smith HJ, et al. The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial. Survey of 71 "negative" trials[J]. N Engl J Med, 1978, 299(13): 690-694.
- [2] Halpern SD, Karlawish JH, Berlin JA. The continuing unethical conduct of underpowered clinical trials[J]. JAMA, 2002, 288 (3): 358-362.
- [3] Ederer F, Church TR, Mandel JS. Sample sizes for prevention trials have been too small[J]. Am J Epidemiol, 1993, 137(7): 787-796.
- [4] Rosner B, Glynn RJ. Power and sample size estimation for the clustered wilcoxon test[J]. Biometrics, 2011, 67(2): 646-653.
- [5] Suresh KP, Chandrashekar S. Sample size estimation and power analysis for clinical research studies[J]. J Hum Reprod Sci, 2012, 5 (1): 7-13.
- [6] Camargo JR. Vitamin D and cardiovascular disease time for large randomized trials[J]. J Am Coll Cardiol, 2011, 58(14): 1442-1444.
- [7] 李光伟. 大规模临床试验未必就科学[J]. 中华外科杂志, 2006, 44 (7): 504
- [8] Tuma RS. Large trials, small gains: is change on the way[J]. J Natl Cancer Inst, 2010, 102(16): 1216-1217, 1223.
- [9] Sackett, DL, Cook DJ. Can we learn anything from small trials[J]. Ann NY Acad Sci, 1993, 703: 25-31.
- [10] Arya R, Antonisamy B, Kumar S. Sample size estimation in prevalence Studies[J]. Indian J Pediatr, 2012, 79: 1482-1488.
- [11] Moonesinghe R, Yang Q, Khoury MJ. Sample size requirements to detect the effect of a group of genetic variants in case-control studies [J]. Emerg Themes Epidemiol, 2008, 5: 24.
- [12] Hemmingsen B, Christensen LL, Wetterslev J, et al. Comparison of metformin and insulin versus insulin alone for type 2 diabetes: systematic review of randomised clinical trials with meta-analyses and trial sequential analyses[J]. BMJ, 2012, 344: e1771.
- [13] Donner A, Rotondi MA. Sample size requirements for interval estimation of the kappa statistic for interobserver agreement studies with a binary outcome and multiple raters[J]. Int J Biostat, 2010, 6(1): 31.
- [14] 田元祥, 翁维良, 陆芳. 中医临床研究样本量设计的优化[J]. 中华中医药杂志, 2010, 25(5): 710-714.
- [15] Dawson R, Lavori PW. Sample size calculations for evaluating treatment policies in multi-stage designs[J]. Clin Trials. 2010, 7 (6): 643-652.

(编辑: 邹元平, 梁进权)