

随机对照试验及统计学分析在临床医学应用中存在的问题

贺佳 (第二军医大学卫生勤务学系卫生统计学教研室, 上海 200433)

[摘要] 随机对照试验(RCT)研究被认为是临床治疗试验中论证强度最高的试验方法,可为临床实践提供真实的科学依据,引导临床医师在实践中做出正确的决策。本文介绍了 RCT 的方法及设计原则,分析了在临床实际应用中存在的问题,并对 RCT 中统计学方法应用的常见错误进行了分析,阐述了应用 RCT 及采用正确的统计学分析方法的重要性。

[关键词] 随机对照试验;统计学;医学;临床医学

[中图分类号] R 195.1 **[文献标识码]** A **[文章编号]** 0258-879X(2006)07-0697-04

Randomized controlled trial and statistical analysis methods for clinical medicine: the existing problems

HE Jia (Department of Health Statistics, Faculty of Health Services, Second Military Medical University, Shanghai 200433, China)

[ABSTRACT] Randomized controlled trial (RCT), considered as the strongest evidence-based method in clinical trial, can provide first-hand scientific evidence for clinical practice and instruct clinicians to make correct decisions in practice. This article introduces the principles for RCT designing, analyzes the common problems concerning RCT in clinical practice and the common mistakes made in RCT statistical methods. It is concluded that application of RCT method and correct statistical methods are of great importance in clinical medical research.

[KEY WORDS] randomized controlled trials; statistics, medical; clinical medicine

[Acad J Sec Mil Med Univ, 2006, 27(7): 697-700]

循证医学强调一切医学实践活动都必须基于现有的最好证据,随机对照试验(randomized controlled trial, RCT)是评价干预效果最科学的研究方法,其结果可以提供最好的决策证据。目前,国外关于药物临床试验的研究,绝大部分都采用 RCT 的方法。国内近 10 年来跟进得很快,绝大部分的医学工作者对 RCT 研究能够接受。作为军医大学的卫生统计学教研室及世界卫生组织/热带病专项(WHO/TDR)临床数据管理中心成员,笔者近几年来参加了大量的临床试验设计及统计分析工作,感到我国在临床医学研究的水平上有了一定的提高,取得了一定的进步。但在严格的科研设计、正确的统计分析以及结果的科学解释等方面仍存在一些不足,尚需不断改进。笔者拟就 RCT、设计原则、统计分析等问题做一阐述。

1 RCT 简介

世界上第一篇应用随机对照设计的论文诞生于 1948 年的《英国医学杂志》(*British Medical Journal*)上,作者是英国著名的生物统计学家 Bradford Hill。这篇论文首次将数理统计理论应用于临床医学的科研设计之中,评价了链霉素治疗肺结核的疗效,将 107 例肺结核患者随机分为两组,分别接受链霉素和常规治疗^[1,2],所得结果为链霉素优于常规治疗。

几十年过去了,目前 RCT 研究被认为是临床治

疗试验中论证强度很高的试验,可为临床实践和卫生决策提供更真实的科学依据,引导临床医师在实践中做出正确的决策。循证医学是当前临床医学进步的必由之路。过去临床医学中的回顾性研究比较多,现在没有进行 RCT 研究的临床医学论文,就难以被高水平的期刊接受。RCT 研究结果可靠,科学性强,重复性好,两组的均衡性好,可排除许多混杂偏倚,保证了结果的真实性。如果一项设计完美的 RCT 能证实某一因素与结果之间有重要的关系,这个证据就很有说服力。

RCT 的做法是将满足条件的受试对象采用随机化方法,分别分配到试验组和对照组中,使非研究因素在试验组与对照组中达到均衡,其疗效差异可归结于干预措施。这样的设计能够最大限度地减小受试对象选择及分配所带来的偏倚,因而 RCT 是目前国际上公认的评价干预措施效果的金标准方案,可用于评价两种或多种干预措施的优劣,确定某种干预措施的利弊,证实某种干预措施的有效性和安全性。

目前,为提高 RCT 报告质量,国际上有了临床试验注册规定与报告规范,科学家和编辑们共同制定了试验报告统一标准(Consolidated Standards of Reporting Trials, CONSORT)声明^[3],对临床科研

[作者简介] 贺佳,教授,博士生导师。

E-mail: hejia@smmu.edu.cn

工作者提出了新的要求,若作者准备向高水平的国际期刊投稿,就必须遵从此项规范。CONSORT 声明由清单和流程图组成,作者可用来报告 RCT。CONSORT 声明为作者提高临床试验报告质量提供了指南,使得评价和解释 RCT 结果更为可行。全球诸多顶级期刊自 2005 年 7 月 1 日起,仅发表经过注册的临床试验报告^[4]。目前,国内多家期刊社也开始优先刊发经过试验注册的文章^[5]。中国循证医学中心(四川大学华西医院)按 WHO 和 Cochrane 协作网要求组织成立了中国临床试验注册中心,并于 2005 年 10 月开始正式运行,任务就是帮助优秀的研究注册及补注册,力争使我国的临床试验更规范。目前国内外有多家机构可提供注册服务,研究者可按自己的需求自行选择。

2 RCT 的设计及常见问题

笔者 1998 年曾组织在校研究生对国内医学期刊中统计学方法的应用情况进行调查,应用试验(实验)设计的文章约占 1/4(24.17%),但严格遵循试验设计原则的文章数量不多。王家良教授^[6]抽样调查了 1985~1996 年 35 种中华医学会系列杂志发表的 RCT 的研究论著,共计 164 篇,其中交待了随机分组方法者仅为 15.2%;交待了随机分组方法,但有错误者为 5.5%;竟有 79.3%的论著提到研究为随机对照试验,但文中根本未见关于随机分组方法与方式的叙述。RCT 研究的报告质量不尽如人意。不恰当的试验设计使治疗的效果产生偏倚,这种系统偏差正严重损害了 RCT 本可摒弃系统偏差的特征^[7]。

RCT 的设计原则是随机化、对照及重复(样本量)的原则^[2,8,9]。

随机化的原则是指每位受试者都有同样的机会进入不同的试验组或对照组中。“随机”不等于“随便”或“随意”。随机化可通过随机数字表或计算机软件实现。未经过随机化分组的研究通常会产生偏倚,更为严重的是,统计学方法不能直接分析有偏倚的数据,因为统计学分析方法是建立在随机化原则基础之上的,统计学方法不能弥补先天不足。实际中常常存在着“随机”概念的误用或滥用情况。有的研究将医生分为两组,患者就诊时碰到哪组就接受对应的治疗,这不是随机,因为医生与患者都可以进行选择,而随机是指医生和患者无法进行选择。非随机化的结果就是夸大或缩小了疗效,无法获得真实的结果。为避免出现实施偏倚,必须限制患者和医护人员对分组情况的知晓,随机化隐藏和双盲就是针对这种偏倚设计的。随机化隐藏是指随机是隐

蔽的、不公开的,患者和研究者是不知分组情况的,具有隐蔽性。如果随机不隐藏,将夸大或缩小疗效。盲法是指患者及与治疗有关的人员不知道患者用药的真实情况,其中双盲是高质量的 RCT 所必备的要素,是避免测量偏倚的重要手段。

对照的原则就是使对照组与试验组的非处理因素相同,抵消其影响,使处理的效应得以显露,使处理因素和非处理因素的差异可以有个科学的比较。对照组的选择,一定要是公认、安全、有效、法定的治疗方法,这时试验组与对照组比较进行非劣效性或等效性检验。安慰剂也可用来做对照,这时的检验需做优效性检验。选安慰剂做对照容易显示统计学意义,且所需样本量较少。但安慰剂对照有时存在伦理学问题,需谨慎采用。一般功能性消化不良、精神科疾病、有自愈倾向的疾病(如伤口愈合)、目前没有有效治疗药物的疾病等,可考虑采用安慰剂做对照。实践中曾碰到,应该选择安慰剂为对照的,评价伤口愈合的药物,却以不符合伦理学为借口,而不采用安慰剂为对照,结果无法证明试验药物的疗效,最终未通过国家食品药品监督管理局的评审。

关于重复的原则,样本量究竟多大才合适?这是个令研究者非常关注的问题,应根据统计学原理计算要达到试验预期目的所需的病例数。我国临床试验法规中规定了临床试验的最低样本数,但同时指出必须符合统计学要求^[10]。若试验设计的样本量过小,则代表性差,下结论缺乏依据;若样本量过大,增加了工作量,造成了不必要的浪费。只要能下结论, P 值小于 0.05,样本量越小越好。就临床试验来说,若设计的是试验组与安慰剂比较,进行优效性检验,则所需样本最少,但可能有伦理学问题;若设计与阳性药比较,做非劣效性检验,所需样本较少;若设计与阳性药比较,做等效性检验,所需样本最多。临床试验中经常看到,不管做什么试验,样本量都是试验组和对照组各 120 例,导致研究结果得不出结论,两组比较既不是优效性也不是非劣效性,得不到证据,浪费了大量的人力、物力、时间和经费。

3 RCT 研究中统计学方法的误用

1964 年,英国著名统计学家 Yates 和 Healy 曾提醒:“非常痛心地看着,因为数据分析的缺陷和错误,那么多好的生物研究工作面临着被葬送的危险。”据国外对 20 世纪 60~80 年代医学期刊上所发表论文的调查,有统计错误的论文的比率在 20%~72%之间。第四军医大学 1996 年对 4 586 篇医学论文的调查显示,统计分析误用率达 55.7%^[2]。军事医学科学

院胡良平教授的调查结果显示,统计学误用率约为88.0%(315/358)。笔者1998年组织在校研究生对医学期刊中统计学方法的应用情况进行了调查,结果显示统计学分析方法的误用率在20%左右。

3.1 统计学方法的误用 虽然医学研究人员都学习过统计学知识,有的研究者曾多次学习医学统计学,但在实际应用中,用错的实例不在少数。如果所选用的统计分析方法是错误的,其结论常常是歪曲事实的或者是彻底错误的。从大量的RCT研究中,笔者观察到统计方法应用错误的具体表现是:(1)统计表绘制不规范,表中层次不清,结构混乱;(2)统计图的图形选择错误,条图、线图混淆;(3)相对数的率、比混淆;(4)统计学分析方法选择错误,应该用方差分析的应用了 t 检验,应该用秩和检验的应用了 t 检验,应该用协方差分析、重复测量设计的方差分析的而没有采用,应该用多元分析的仅采用了单因素分析等等。有些研究,由于统计学分析方法应用错误,导致结果、结论错误,非常可惜。更有甚者,有些人根本不进行统计学分析,即直接写上 $P<0.05$,没有写出统计学方法的具体名称,而有的虽然写出了统计学方法,却将计量资料写成用 χ^2 分析,计数资料用 t 检验分析。另外,应用什么统计学方法开始要规定好,不能到时再说,应该事先规定而不是事后选择,有的研究者,将同一份资料用不同的统计学方法去分析,选择能得出 $P<0.05$ 的那种方法。统计学不是数字游戏,数据分析必须遵循统计学原则。

3.2 对统计学结果的错误理解 统计学的 P 值是神秘的,不少人将研究的成败就寄托在这个 P 值上。 P 值的含义就是拒绝原假设犯错误的概率,也就是下结论为优效、等效和非劣效时犯错误的概率。通常 P 值以0.05或0.01为界,现在都采用统计学软件计算,应该写出确切的 P 值,这样可以更好地判断结论的可靠性。值得注意的是,实际中常常有人理解, P 值越小,所比较的组别之间差别越大,这是错误的,必须引起高度重视。正确的理解是, P 值越小,拒绝假设的理由越充分,越可以下“差别有统计学意义”的结论。还有人将 $P<0.05$ 理解为有显著差异, $P<0.01$ 有非常显著的差异,都是错误的。在统计学上没有量的区别,只判定有没有统计学意义,因为 P 值的大小与它的样本量有关,样本量越大, P 值越小。描述假设检验结果不仅要写出 P 值,同时还应写出统计量是多少,附上可信区间。假设检验和可信区间同时呈现,才能更好地理解和解释结果。同时,还应指出所用的统计学分析软件,注意统计分析软件的合法性。

实际中经常看到,完成一项试验后,经两组比较, $P>0.05$,认为两组疗效相当,这是非常错误的,必须引起高度重视。这时还需要考虑把握度,若把握度低,则需要扩大样本。如果试验组与阳性对照组进行比较,统计学检验的目的就是等效性或非劣效性(不对照组疗效差)试验,计算方法不同,但若下结论为等效或非劣效性,需要得出 $P<0.05$ 。因此,需要强调的是,统计结果的解释应着重考虑其临床意义。

值得注意的是统计学结果与医学结论矛盾的情况,实际情况时两组间疗效有差别,但由于样本量小,统计学可能得出差别无统计学意义的结论,这时需要扩大样本。但当样本量非常大时,即使比较组之间疗效差别非常小,也可能得出有统计学意义的结论,而该结论却无临床实际意义。所以,统计学结论与医学结论可能出现不一致。故笔者在此提醒科研工作者,不要迷信于统计学的“ P ”值,将精力放在获得小的“ P ”值上,而应该将重点放在严格、科学的设计,正确的统计学分析上。

2003年8月我国颁布的《药物临床试验质量管理规范》(GCP)中提到^[1],临床试验资料的统计分析过程及其结果的表达必须采用规范的统计学方法。临床试验各阶段均需有生物统计学专业人员参与。所以,统计学思想应该贯彻于医学科研实践的始终。

[参考文献]

- [1] 李强. 临床疗效研究的金方案[J]. 世界华人消化杂志, 1999,7:697-698.
- [2] 徐勇勇主编. 医学统计学[M]. 2版. 北京:高等教育出版社, 2004:7.
- [3] The CONSORT Group. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials [J]. Lancet, 2001, 357:1191-1195.
- [4] De Angelis C, Drazen JM, Frizelle FA, et al. Clinical trial registration: a statement from the International Committee of Medical Journal Editors[J]. JAMA, 2004, 292:1363-1364.
- [5] 李幼平,吴泰相,李静,等. 创建中国临床试验注册和发表机制的联合宣言[J]. 第二军医大学学报, 2006, 27: I-II.
- [6] 王家良. 重视科研设计提高临床科研质量[J]. 中华眼科杂志, 1999, 35:167-168.
- [7] Grimes DA, Schulz KF. An overview of clinical research: the lay of the land[J]. Lancet, 2002, 359: 57-61.
- [8] International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH Harmonised Tripartite Guideline E9. Statistical Principles for Clinical Trials[G/OL]. 1998-02-05. <http://www.ich.org/cache/compo/276-254-1.html>
- [9] 国家药品监督管理局. 药物临床试验质量管理规范[EB/OL]. 2003-08. <http://www.sda.gov.cn/cmsweb/webportal/>

W53384/A34639009.html

- [10] 国家药品监督管理局. 新药审批方法[EB/OL]. 2003-08. <http://www.sfda.gov.cn/cmsweb/webportal/W53384/A61246595.html>
- [11] International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use.

ICH Harmonised Tripartite Guideline E6(R1). Guideline for Good Clinical Practice[G/OL]. 1996-05-10. <http://www.ich.org/cache/compo/276-254-1.html>

[收稿日期] 2006-06-26

[修回日期] 2006-06-29

[本文编辑] 邓晓群