

• 回顾与思考 •

流行病学实验研究发展历史

吴涛 詹思延 李立明

流行病学实验研究作为流行病学的重要研究方法之一,是通过比较给予干预措施后的实验组人群与对照组人群的结局,从而验证研究假设和考核干预措施效果的一种前瞻性研究方法,又称干预研究、实验流行病学研究等。其中包括临床试验、现场试验和社区干预项目等研究类型^[1]。这类研究方法在临床治疗和疾病预防措施的科学与评价和筛选、医疗卫生政策、健康教育及诊断技术效果评估等方面起着举足轻重的作用,已被视为评价干预措施有效性的标准方法。但流行病学实验研究在方法学发展和科学性不断完善方面经历了漫长的历程。

一、流行病学实验研究各要素的发展

在现代流行病学试验研究中,设立对照、随机分组、盲法和安慰剂的应用已成为在评价干预措施效果时减少偏倚发生的不可缺少的部分。

1. 对照实验研究:1865 年 Bernard 在实验研究中应用了“对照”一词,1890 年 Hankin 在动物试验研究中也提到了“老鼠对照”,但他们均未指出对照一词的确切含义^[2]。对照一词来源于法语 contre-rolle,原义为重复登记。1893 年,新英语词典将该词定义为“一种在比较研究中设立的标准对比组”。

最早记载的对照试验发生在公元前 605 年, Nebuchadnezzar 二世时期他命令子嗣们必须以酒肉为食,其中两个子孙每天食用面包和水,10 天以后他们表现出了比其他人更好的精神状态^[2]。1537 年,外科医生 Ambroise Paré 使用玫瑰、松油和蛋黄替代原有的疗法治疗外伤,并发现新的药膏有更好的疗效^[3]。1747 年,英国医生 James Lind 尝试用临床试验的方法来探讨坏血病(维生素 C 缺乏症)的病因及其治疗方法。他将 12 名病例分成 6 个小组,每组给予不同的治疗,其中一组病例每日服用橙子和柠檬作为治疗方法,其他组则给予苹果酒、醋或海水等。结果发现服用橙子和柠檬组的 2 名病例病情迅速好转,而其他病例病情变化不大^[4]。该研究初步显示出了平行对照研究的科学价值。1848 年 Semmelweis 关于产褥热的研究、1914 年 Goldberger 关于糙皮症的研究是早期采用平行对照研究的其他实例^[2]。早期采用历史对照进行比较的研究有 19 世纪 60 年代, Lister 使用消毒方法发明之前的医院记录,比较了消毒法使用与否对截肢患者死亡率的影响。1898 年,丹麦的 Fibiger 医生在白喉抗毒素血清的质量效果评价中,对隔日就诊患者给予不同治疗方法,进一步提高了对照研究的质量。

由于安慰剂效应和霍桑效应的发现,研究者更加意识到设立对照组进行实验研究的重要性。

(1) 安慰剂效应:安慰剂(placebo)一词为拉丁语,意为“我很满足”。安慰剂效应的发现始于 1784 年,路易十六下令调查是否真的存在“动物磁场”或是“催眠术”。Benjamin Franklin 带领其他研究者开展了一项研究,以评价催眠和磁性是否具有真正的疗效。他们将受试者的眼睛遮住,使用安慰剂如“催眠水”等进行干预。结果发现,只有在告知受试者使用磁场治疗的情况下,他们才能“感到”实际的效果,而事实上研究者并未采用相应的干预措施。1799 年, Haygarth 的研究更加证实了安慰剂效应的存在。他使用一套木棒代替原本声称具有治疗效果的金属棒,结果在治疗风湿疼痛时获得了与金属棒类似的治疗效果,而之前人们认为疼痛的缓解是由于金属的磁场效应带来的^[5]。19 世纪,临床试验中开始使用安慰剂。1863 年,伦敦 Guy's 医院的 Gull 医生在 21 例风湿热患者的治疗中使用了薄荷水,证实了安慰剂治疗在评估疾病自然病程和自发痊愈中的重要性。同年,美国的 Flint 医生在治疗 13 例风湿患者时同样使用了安慰剂,并得出了传统医学方法的“效果”不过是疾病自然发展的结果。

(2) 霍桑效应:1927~1932 年,哈佛大学商学院的 Mayo 教授在芝加哥郊区的 Hawthorne 西区电力工厂开展了一项研究^[6]。最初,研究者只是希望了解工作环境对生产力带来的影响。他采用了一系列手段改善女工的工作环境,包括改善灯光、休息时间、工作时间、送餐服务等。结果发现无论如何改变,女工的工作效率均显著提高。在将这些改善措施全部撤销、恢复到试验之前的状态后,工作效率更是达到了最高的状态。结果表明,受试者的感受和与干预措施的特异性无关,而是由于其成为研究中受关注的目标而改变其行为、心理感受的一种趋向。

2. 随机化概念的引入:流行病学实验研究的另一重要特征为随机化概念的应用。早在 19 世纪初, Louis 就提出如果有完全相似的两组受试者接受干预,评价措施有效性的可信度会大大提高。随机化概念于 1923 年由 Fisher 最早引入农业实验研究,并发展了基于随机化的方差分析方法^[2]。

早期在医学实验研究中使用单纯随机分组方式进行试验的研究包括:17 世纪 Van Helmont 使用了抽签法随机分配发热患者分别至放血疗法组和对照组,1929 年 Colebrook 采用掷硬币、1931 年 Doull 等采用掷筛子法、1937 年 Theobald 采用玻璃珠涂色法等来分配干预措施等^[7]。1931 年, Amberson 在治疗肺结核的临床研究中亦采用了分层随机的方法分配治疗措施^[2,8]。当时,研究者选择了 24 名病例,在个体匹配后,采用掷硬币的方法决定了治疗的分组。

在试验组中采用无机金盐硫代硫酸金钠 (sanocrysin) 进行治疗。同时,只有 2 名设计者和主管病房的护士了解分组情况,而患者并不知情。该项研究不仅运用了随机分组方式,同时也采用了多个条件对研究对象进行匹配,且提到了参与试验者并不知道治疗方法。消除在选择研究对象时发生偏性的另一方法是对参加试验的决策者隐匿随机分组的方案。该方法最好基于随机数字法实行。在 1948 年英国医学研究委员会开展的使用链霉素治疗肺结核的临床试验中,详细描述了隐匿法的实施过程,这也被认为是随机方法在实验研究中发展的里程碑。

3. 盲法的使用:盲法是流行病学实验研究的重要特征之一。其作用是为了防止参与试验者了解治疗方法后对研究结果造成影响。1799 年 Haygarth 木棒代替金属棒的研究被认为是早期开展的单盲试验之一。

早期开展的双盲试验包括,如 1880 年 Potter 和 Storke 在一项对比研究中,使用了安慰剂干预对照组,且主治医生亦不了解分组情况^[5]。如前所述,1931 年 Amberson 治疗肺结核的研究中亦采用了双盲法^[2,8],即患者和医生并不了解试验的分组情况。1911~1914 年, Bingel 在白喉治疗试验中采用了盲法,患者和参与的医生(除他本人外)均不了解分组方式。20 世纪后半叶,盲法开始被应用到现场试验中。

二、各类流行病学实验研究的发展

1. 临床试验:1912 年,美国国会禁止药品说明书上有与疗效不符的介绍,这在一定程度上促进了临床试验的开展。20 世纪 30 年代,英国治疗试验委员会成立以管理新药的临床试验。1938 年,107 人服用磺胺类药物后死亡。美国食品药品监督管理局(FDA)加强了对药物安全使用的管理^[3]。1946 年,英国医学研究委员会开展了一项使用链霉素治疗肺结核的临床试验。该试验由英国著名统计学家 Bradford Hill 设计进行。试验并未使用安慰剂治疗对照组,但患者和 2 名独立对 X 线片进行判断的放射科医生均不了解患者的分组情况。另外,此项研究还考虑到了可能涉及到的医学伦理学问题,由于链霉素在英国并不容易得到,因此研究并不违背医学伦理学原则。1948 年,该研究第一个现代随机化双盲临床试验发表于英国医学杂志,其研究设计包含了流行病学实验研究的每个重要特征^[2]。

1962 年英国 Hill 编撰的《临床和预防医学中的统计学》一书问世,可以作为临床试验研究发展的一个重要里程碑。同年,美国立法要求新药必须经临床试验评价后由 FDA 批准才能上市,由此,极大促进了临床试验的开展。

国际间多中心协作研究是临床试验发展的另一个重要里程碑。第一篇多中心临床试验的报告发表于 1944 年,英国医学研究委员会开展了一项多中心对照实验研究,以评价棒曲霉素(Patulin)对感冒的治疗效果,结果显示棒曲霉素的疗效很小或根本无效^[9]。现代成功的多中心临床试验实例还有 ISIS(International Study of Infarct Survival)协作组于 20 世纪 80 年代中期开展的急性心肌梗死临床试验,目前其协作范围已扩大到三十多个国家上千家医院,2 年时间可入

选五六万名病例,极大的提高了研究效益,其成果对急性心肌梗死的治疗产生了积极影响^[10]。其后相继开展了大量抗生素治疗或疫苗预防传染性疾病的临床试验和现场试验,而对慢性非传染性疾病的多中心试验最早开始于 20 世纪 60 年代。1970 年,美国 Finch 地方法院强制执行了 1962 年通过的 Kefauver-Harris 药品修正法案,指出药品商业的成功并不代表药品真正有效性与安全性。1979 年临床试验协会成立,1980 年“Controlled Clinical Trial”杂志首期发行。

2. 现场试验:1938 年, Harold Diehl 报告了在明尼苏达大学开展的感冒疫苗研究是早期的现场试验之一。1945~1955 年美国牙医 Ast 在纽约州开展了为期 10 年的人群试验,探索饮水中人工加入氟化钠对龋齿的预防效果^[11]。1951 年,英国医学研究委员会在其开展的一项百日咳疫苗试验中采用了隐匿的方法随机分配干预措施^[12]。迄今规模最大的人群实验是 1955 年 Francis 进行的疫苗现场试验,实验对象为近百万在校儿童,此项研究为脊髓灰质炎的预防奠定了坚实的基础。

3. 社区试验:20 世纪 70 年代初期,美国加利福尼亚州和芬兰北 Karelia 州开始出现社区试验。这些大规模的社区试验主要目标为降低慢性非传染性疾病在社区内的发病率与死亡率。芬兰北 Karelia 计划始于 1972 年,其中心任务是通过社区干预改变整个北 Karelia 人群心血管的危险因素现状^[13]。由于该计划的目标是从社区公共卫生的需求中演变而来,干预的策略是在相关的理论指导下设计的,根据现有条件和情况开发并调整计划的措施,因此,计划得到了成功的实施。在合理评估其经验后,北 Karelia 计划作为全芬兰的“试点”和“示范”方案,在全国推广,并为世界各国所借鉴。

三、医学伦理学观念

随着实验研究的蓬勃发展,由于涉及到以人类作为试验对象,研究者越来越意识到在研究中面临着种种伦理学方面的困惑,还有一些研究者为了研究目的而侵害了受试者的权益。例如美国公共卫生部 1932 年开始 Tuskegee 梅毒研究,目的在于确定慢性梅毒损伤的原因^[14]。在 1932 年,梅毒的治疗需要重金属如砷、铋、汞。1945 年,青霉素开始广泛应用,众所周知,青霉素是治疗梅毒既安全又有效的药物。所以如果在 1945 年后继续 Tuskegee 试验就违背了伦理学原则。但是,直到 1971 年华盛顿邮报的一名调查记者揭露出此事,试验才停止下来。研究者不仅未取得受试者的同意,而且向他们撒了谎。例如,他们谎称对患者做脊髓穿刺是“一种特别的新疗法”,但实际上,却是为研究需要而掌握脊髓的蛋白质水平。

知情同意并非当代的新概念。早在古希腊时期,医学之父希波克拉底在他的《传染病学[卷]》一书中提到,为了治疗疾病,患者必须与医生协调一致并知情同意。但那时的知情同意与医生对患者的尊重毫不相干,医生这么做是由于他们认为这样可以使患者参与自己的治疗措施,从而提高疗效^[15]。

在实验研究发展的进程中,一些治疗方法和疫苗的效果评价是以一些弱势群体如避难者、犯人的生命为代价的,而

这些受试者并不了解试验的内容与目的。二战中纳粹德国医生在纽伦堡进行了惨无人道的人体试验;这提醒医学界应吸取教训并制定相应的法案防止惨剧的再度发生。1948 年,联合国大会通过了《纽伦堡法典》,这是第一部国际上关于试验研究的伦理学法典,其中特别强调了知情同意原则,即一切治疗或实验都必须向患者或受试者说明情况,包括所施程序的依据、目的、方法及潜在损伤、风险和不可预测的意外等,然后在没有威胁利诱的条件下获得患者主动的同意,或在可能的多种选择办法中做出自由的选择^[16]。1964 年,世界医学联合会在芬兰赫尔辛基举行的大会上通过了新的伦理学法典——赫尔辛基宣言。研究机构、医学与基础科学杂志开始要求从事试验研究者签名声明遵守赫尔辛基宣言,否则就不能在相应的会议上宣布其研究,其文章也不会被发表。之后,各国相继成立了相应的医学伦理委员会,其职责是在试验研究开始前对其伦理学问题进行审查。

四、近年来流行病学实验研究的发展

过去的 20 年,流行病学试验进入了蓬勃发展时期,仅以 PubMed 检索为例,1981~1985 年收录的临床试验类文章有 29 142 篇,1986~1990 年为 47 963 篇,1991~1995 年达到 77 495 篇,1996~2000 年已高达 110 961 篇,几乎以每 5 年翻一番的速度增长,并广泛覆盖心脑血管疾病、肿瘤、糖尿病、出生缺陷等非传染性疾病和意外伤害及传染病的防治研究,以及对卫生管理、保健设施和保健项目的评价。20 世纪 90 年代,循证医学的思想和理论体系形成并逐渐发展成熟,在浩如烟海的医学研究成果中,随机化对照试验的结果被公认为临床和公共卫生实践领域的最佳证据。

尽管流行病学实验研究中随机化临床试验 (randomized clinical trial, RCT) 已成为检验临床干预安全性和有效性的理想方法,但其中存在的主要偏倚退出,包括不合格、不依从、失访等会导致原定的样本量不足,使研究工作效力降低;如果试验组和对照组退出不均衡,更会对研究结果的真实性产生更大的影响。为克服这种偏倚,在 RCT 设计和分析中采用了多种方法,如意向治疗分析 (intention to treat analysis, ITT) 等。ITT 首次应用是在 1961 年,它是指所有病例被随机分入 RCT 中的任意一组,不管他们是否完成试验,或者是否真正接受了该组治疗,都保留在原组进行结果分析。ITT 的目的在于避免选择偏倚,并使各治疗组之间保持可比性^[17]。

虽然在 1950~1980 年间,实验研究已在医学界得到了广泛的应用,可是许多研究结果却不尽人意。许多研究出现假阴性结果,且对于同样问题的试验结果往往不一致。1971 年,美国学者 Light 和 Smith 提出,应当在全世界收集某疫病中各种疗法的小样本、单个临床研究试验结果,对其进行系统评价和统计分析,将尽可能真实的科学结论及时提供给社会,以促进推广真正有效的治疗手段,剔除尚无依据的、无效的甚至有害的方法。这样 Meta 分析应运而生。该术语则是由社会心理学家 Glass^[18] 于 1976 年首先命名。Meta 分析有利于增大统计功效、评价结果的一致性、增强干预措施的可靠性和客观性并能够帮助寻求新的假说。它还能够评价各

项研究结果以及各亚组结果的不一致,探索异质性的来源,定量综合估计处理效应的大小,改善效应估计值。但也有研究者提出其方法学本身具有缺陷,易受到出版偏倚和原始研究报告质量的影响,并有可能阻碍研究进一步的深入。

五、我国的流行病学实验研究

国内早期的流行病学现场试验研究有:1963 年天津医学院、河北医学院、河北地方病防治所在承德市郊开展了以碘盐作为干预措施防治地方性甲状腺肿的现场试验^[19];1979 年前后,中国医学科学院卫生研究所在东北克山病地区开展了向人群投掷制剂以预防克山病的现场实验;1979 年苏德隆等在江苏省启东县进行的关于水源与肝癌发生关系的类实验 (quasi-experiment);同期,武汉、北京及长春等生物制品研究所组织了全国各省、市卫生防疫站进行的关于流行性脑脊髓膜炎多糖体菌苗、细菌性痢疾活菌苗及腮腺炎疫苗的人群流行病学实验等。

中国早期随机对照试验开始于 1970 年代末。检索《中国生物医学文献数据库》(CBMDISC) 早期发表的随机化双盲对照临床试验包括 1981 年发表的《用 Nd-OFS 治疗胃溃疡的随机双盲试验》,1983 年中医药领域的中药注射剂治疗心绞痛等。采用 CBMDISC 检索,1979 年至今国内发表的随机对照试验研究文章总数已达 8549 篇。20 世纪 80 年代以来,我国开始开展一些大规模的多中心临床试验,如中国老年收缩期高血压试验等。

尽管近十余年来随机对照试验发表的数量在逐渐增多,但研究质量仍然参差不齐。1990 年代中后期国内外学者对我国的临床试验进行了深入的方法学质量评价,结果发现国内的临床试验除了质量不稳定,还存在发表偏倚、随机概念的误解和滥用、极少使用盲法;随机误差与系统误差使试验报告的结果夸大了干预措施的效果等。一些学者提出成立相应的高水平试验研究中心,促进研究治疗的提高,帮助研究者创造高水平的实验研究成果。

病例对照研究、队列研究和随机对照试验的出现称为现代流行病学的开端,它们也同时形成现代流行病学研究方法的三大基石,标志着流行病学已经发展成为一套以研究方法为主要内容的学科。随机对照试验方法的形成简单、完美地解决了长期困扰干预研究的混杂问题,从而为流行病学研究开拓了一个更为广阔的领域。

(衷心感谢曹家琪教授审阅本文,并提出宝贵的修改意见)

参 考 文 献

- 1 施伯元. 实验流行病学研究. 见: 李立明, 主编. 流行病学. 第 5 版. 北京: 人民卫生出版社, 2003. 106.
- 2 Lilienfeld AM. The Fielding H. Garrison Lecture: *ceteris paribus*; the evolution of the clinical trial. *Bull Hist Med*, 1982, 56: 1-18.
- 3 History of clinical trials; how it all began. URL: http://www.healthandage.com/html/res/clinical_trials/content/section1.htm. 2004.
- 4 Robert HF. Evaluation of interventions. *J Clin Epidemiol*, 2002, 55: 1183-1190.

5 Kaptchuk T. Control of observer biases; masking (blinding) and placebos. URL: http://www.jameslindlibrary.org/essays/bias/observer_bias.html. 2004.

6 The Hawthorne Effect. URL: http://www.envisionsoftware.com/articles/Hawthorne_Effect.html. 2004.

7 Chalmers I. Control of selection biases; comparing like with like. URL: http://www.jameslindlibrary.org/essays/bias/selection_bias.html.

8 Lindquist EF. Statistical analysis in educational research. URL: http://www.jameslindlibrary.org/trial_records/20th_Century/1940s/lindquist/lindquist_commentary.html. 2004.

9 Chalmers I, Clarke M. The 1944 patulin trial: the first properly controlled multicentre trial conducted under the aegis of the British Medical Research Council. *Int J Epidemiol*, 2004, 32 : 253-260.

10 ISIS-4 Collaborative Group. ISIA4: a randomised factorial trial assessing early oral captopril oral mononitrate, and intravenous magnesium sulphate in 58 050 patients with suspected acute myocardial infarction. *Lancet*, 1995, 345 : 669-685.

11 张衍文. 氟与龋齿. 见: 钱宇平, 主编. 流行病学实例. 第 1 卷. 北京: 人民卫生出版社, 1984. 50-58.

12 Chalmers I. MRC Therapeutic Trials Committee's report on

serum treatment of lobar pneumonia. URL: http://www.jameslindlibrary.org/trial_records/20th_Century/1930s/MRC_trials/MRC_trials_commentary.html. 2004.

13 Lindgren A, Pukkala E, Nissinen A, et al. Blood pressure, smoking, and the incidence of lung cancer in hypertensive men in North Karelia, Finland. *Am J Epidemiol*, 2003, 158 : 442-447.

14 Robert JL, 张健. 美国人类受试者使用的历史. *医学与哲学*, 2001, 22 : 1-5.

15 姜萍, 殷正坤. 知情同意再探. *中国医学伦理学*, 2002, 15 : 20-21.

16 The history of informed consent and the system of protections <http://cancer.gov/clinicaltrials/conducting/informed-consent-guide/page4>. 2004.

17 李晓晖, 詹思延, 李立明. 随机化临床试验结果分析和解释的几个问题. *中华医学杂志*, 2000, 80 : 475-477.

18 Glass GV. Primary, secondary, and Meta analysis of research. *Educ Res*, 1976, 5 : 328.

19 于志恒. 地方性甲状腺肿人群干预实验. 见: 钱宇平, 李立明, 主编. 流行病学研究实例. 第 3 卷. 北京: 人民卫生出版社, 1996. 9-20. (收稿日期: 2004-05-24) (本文编辑: 张林东)

• 疾病控制 •

一起乙型溶血性停乳链球菌感染爆发的检测报告

韩秀兰 李云 柳连顺 董永辉 杜敏 魏秀萍 刘丽花

2003 年 5 月下旬石家庄市某高校发生一起不明原因发热(体温 37.5~39℃)病例,伴有咽痛,全身不适症状,后经实验室检测证实为停乳链球菌感染所致,检测结果报告如下。

患者咽拭子涂抹物 17 份,未发病人群涂抹物 30 份,均在采样后即刻划血平板 5%CO₂ 37℃ 培养。检测材料与方法见参考文献[1]。结果从 17 份患者咽拭子及 2 份未发病人群咽拭子中均检出形态、菌落特征、生物学性状完全相同的溶血性链球菌。鉴定结果如下:①形态特征:为革兰染色阳性球菌,无芽孢,在 1%葡萄糖肉汤培养基中呈长链状排列,链长一般为 5~6 个,有的长 12~14 个,还有的形似项链状排列。②培养特征:在血琼脂平板上经 5%CO₂ 24 h 37℃ 培养后菌落形态为:乙型溶血,菌落较小、圆形突起、呈半透明状、边缘整齐、光滑湿润、菌落周围有较宽的透明溶血环,在 1%葡萄糖肉汤中底部呈颗粒状浑浊生长。③链激酶及杆菌肽敏感试验:均参照中华人民共和国国家标准食品卫生检验方法微生物学部分 GB4789. 11-94 进行操作。链激酶试验结果:所有样本均在 10 min 内全部凝固,并于 40 min 开始溶解,6 h 后完全溶解,结果为阳性。杆菌肽敏感试验 37℃ 5% CO₂ 培养 24 h,观察无抑菌带出现,结果为阴性。④生化特征:为提高准确度和检验速度,使用了 Vitek 生化系统进行了鉴

定,GPI 鉴定结果为:停乳链球菌,生化反应见表 1。⑤药敏试验:试验菌株对氨苄青霉素/舒巴坦、羧苄青霉素、羧噻吩青霉素、环丙沙星、头孢噻肟、头孢西丁、头孢唑啉、头孢三嗪、呋喃妥因、庆大霉素敏感,对头孢他啶、四环素、红霉素、氨苄青霉素、妥布霉素耐药。

表1 19 株乙型溶血性停乳链球菌 Vitek 生化鉴定结果

项目	结果	项目	结果
蛋白胨基质	+	水杨素	+
半纤维素酶	-	海藻糖	+
40%胆汁	-	淀粉	+
尿素	-	松三糖	-
右旋糖、葡萄糖	+	木糖	-
棉子糖	-	奥青托琴	+
蔗糖	+	10%胆汁	-
丙酮酸	-	精氨酸	-
密二糖	-	新生霉素	-
核糖	+	甘露醇	-
杆菌肽	-	山梨醇	-
6%氯化钠	-	阿拉伯糖	-
七叶灵	-	菊糖	-
红四氮唑	+	纤维二糖	-
乳糖	-	触酶	-

通过流行病学调查、细菌学鉴定、生化反应等结果判断,确认此次区域性感染爆发是由 β-乙型溶血性停乳链球菌引起。

参 考 文 献

1 韩秀兰,柳连顺,李云,等. 一起金黄色葡萄球菌 A、C 肠毒素引起食物中毒的检测分析. *中华流行病学杂志*, 2004, 25 : 547.

(收稿日期: 2003-12-16)

(本文编辑: 张林东)

作者单位: 050011 河北省石家庄市卫生防疫站(韩秀兰、李云、柳连顺、董永辉、杜敏、刘丽花); 正定县卫生防疫站(魏秀萍)