

# 疗效比较研究的方法学应用及其实施过程

李敏 时景璞

**【导读】** 疗效比较研究(CER)可有效比较诊疗措施的临床效果,为医疗决策提供证据。其设计除采用大规模临床随机对照试验、整群随机对照试验、类试验、数学模型外,以电子病历为基础的观察性研究备受重视。对研究课题的证据生成、综合、传播及应用的每一环节上,采用CER均需相应的统计学方法进行统计分析和质量控制。

**【关键词】** 疗效比较研究; 研究设计; 生物统计学

**The process of application and implementation on methodologies for comparative effectiveness research** LI Min<sup>1,2,3</sup>, SHI Jing-pu<sup>1,2,3</sup>. 1 Department of Clinical Epidemiology, the First Affiliated Hospital, 2 Cardiovascular Institute of China Medical University, Shenyang 110001, China; 3 Evidence-Based Center of Liaoning Province

Corresponding author: SHI Jing-pu, Email: sjp56@yahoo.com

**【Introduction】** Comparative effectiveness research is able to compare the clinical effects of diagnosis and treatment measures effectively, thus providing evidence for making medical decision. In addition that the large-scale clinical randomized controlled trials, cluster randomized trials, quasi-trial, and mathematical models have started to be used in the designing processes of comparative effectiveness research, observational studies on the foundation of electronic medical records have also been paid enough attention. For each research topic, every procedure on the following areas as: generation of a specific piece of evidence, as well as synthesis, communication, translation and application of related data need to have corresponding statistical methods to perform statistical analysis and quality control.

**【Key words】** Comparative effectiveness research; Research design; Biostatistics

## 一、疗效比较研究

随着新药及新技术(影像学、检验学及介入诊断治疗)不断涌现,导致医疗费用膨胀,但并未降低心脑血管病、肿瘤等慢性病的发病,也未提高人群平均寿命。以美国为例,新医疗技术研发居世界前列,而该领域投资是其他发达国家的2倍,但将这些研究成果和资金转化到改善民众健康水平上,却未见明显收益。面对愈发难以负荷的医疗费用,美国国会于2009年2月提出用于不同卫生保健措施的疗效比较,试图找到最有效和消费者最接受的治疗处方集<sup>[1]</sup>。即随后国际间兴起疗效比较研究(comparative effectiveness research, CER)。

美国政府对CER最初的定义是“严格评估针对某一患病人群不同治疗间的疗效差异,可以比较两种药物、不同治疗方式,也可以比较获益风险比或效价比”<sup>[2]</sup>。美国医学研究所(IOM)制定了100个主要

优先发展的研究领域,其中以卫生保健居首,其次是心血管和周围血管病,共涉及25类疾病和29个研究课题<sup>[3]</sup>。需要评价的干预措施包括药物制剂、设备、程序和诊断方法等。IOM的定义是“CER提供证据证实不同治疗方法在预防、诊断、治疗和临床监测方面的获益风险差异,使其有据可查”<sup>[4]</sup>。Slutsky和Clancy提出了该定义的速记法,即“在什么条件下,对谁何种治疗措施效果最好”<sup>[5]</sup>。CER又称临床疗效研究(clinical effectiveness research)和以患者为中心的研究(patient-centered research),其目标均是在人群、亚人群水平上,对两种以上治疗措施开展短期或长期的重要结局评定,产生和分析新证据,综合二级研究资料,证明治疗措施在预防、诊断、治疗、监控临床状况,及改善护理方面的获益和危害,为临床决策提供更多信息。效力研究(efficacy trials)以随机对照试验(RCT)为代表,关注干预措施能否在理想、严格控制环境下产生预期效果,着重于干预措施的内部有效性,相比之下,CER更侧重于效果研究(effectiveness trials),评价干预措施在真实临床环境下的治疗效果,重在外围有效性,在一些方面有异于

常见的效力研究,如效果研究限制较少,样本量较大,能更好地回答发病的亚人群问题,得到的结果原则上更代表真实世界中观察的结果<sup>[6,7]</sup>。

CER的提出引起广泛关注和激烈讨论。支持者认为:①应规范医疗市场。目前许多临床试验将有意控制风险,开展观察时间短样本量小的研究,且尽量纳入并发症少的患者。美国食品药品监督管理局(FDA)同样认可此类研究结果<sup>[8]</sup>。而这样的治疗措施很可能有较高的使用率。即使该治疗措施明显优于常规治疗,但由于缺乏有效收集和传播这些信息的渠道,也很难使其发挥功效。CER考虑的是减少较差临床决策的费用,将更多资源最大程度帮助患者。此外还应慎重判断安慰剂的使用。②为医疗决策提供证据。有研究估计,IOM约有一半的临床决策基于证据,CER将改善这一现象<sup>[9,10]</sup>。Cochrane中心支持开展疗效研究,认为强调CER是国家主动提高患者和临床医生决策制定的基础。③强调“个性化”治疗。CER强调针对不同类型的患者,比较不同治疗方法的优劣,找出最佳治疗方法,即通过评估临床疗效和成本效益,对特殊的患者或疾病提供最合适的治疗方案。

但CER考虑了成本效益因素,可能涉及到患者的权益<sup>[11]</sup>,故CER反对者认为<sup>[12]</sup>:①政府涉足医疗。民众担心一旦CER的研究结论被法制化,成为医生及医药公司等行业的行业法,将限制决策范围。Price称“这将会产生一个替代医生和患者定量配给处方药的永久性政府”,最终导致政府控制医患关系,出现“食谱医疗”和定量供给的局面<sup>[8]</sup>。CER支持使用电子医疗记录开展观察性研究,但有人对此担忧,称“资料电脑化,政府就可将监控个体病例的治疗决策,并惩罚未遵循治疗指南的临床医生”<sup>[13]</sup>。②其他方面包括民众担心CER可能影响个性化治疗<sup>[14]</sup>;成本效益或多或少与伦理发生碰撞,试想如果患者最佳治疗策略不符合成本效益的常规标准,那

么医生的道德责任是什么。总之,CER引起的争论已影响其实施,并揭示出很多问题。

## 二、研究方法

### 1. 试验性研究:

(1)RCT:在推动和促成CER中,RCT发挥重要作用也面临复杂的问题。首先,为评估研究结果间潜在的差异并具有统计学意义,该试验往往需要更大的样本量;其次,患者病期不同,相关指标随之变化,故在一组药物比较中,可能出现在不同疾病阶段有效药物差异的现象;此外,患者和医生关心的临床结局可能需要等待一定时间而由此带来一系列问题<sup>[15]</sup>。故理想化的RCT设计很难实施,但一个设计良好、效力充足的RCT仍可为干预措施效果评价提供严格的证据,如观察个体患者的干预措施及其效益时<sup>[16]</sup>。然而,当在人群水平研究干预措施效果时,如对某传染病筛查患者是否携带感染病毒,并实施隔离防止传播,该筛查项目均会对未被隔离者带来间接获益,故不宜使用RCT,可采用整群随机对照试验、类试验和数学模型<sup>[17]</sup>(表1)。

(2)整群随机对照试验:该方法对充分评估人群干预措施极其重要,但要求有足够的样本量。实施中保证严格、正确地随机化,避免研究对象退出;分析时除传统的亚组分析、意向性分析外,有时需使用调整分析等统计学方法处理整群设计效应。2004年发表了针对整群随机对照试验的CONSORT声明<sup>[18]</sup>。当存在纳入对象间相互影响或受到同样外部环境因素影响时,整群随机对照试验能更好地避免不同干预间的混杂,更适用于开展行为或环境暴露的干预研究。

(3)类试验:为单个RCT和整群随机对照试验的替代试验。类试验设计有明显的优势,如效力 and 普遍性,并具有评价非随机化事件的能力。类试验虽在一定程度上简化了试验研究,但在设计和统计分析上仍要谨慎,应有效控制混杂因素和选择性偏

表1 CER的研究方法比较

研究方法	适用条件	优点	制约
RCT	个体患者有效性相互独立	有效控制混杂;内部有效性强	结论外推性较差;结果可能模糊不清;样本量及时间跨度问题
整群随机对照试验	针对整个人群施加措施;个体患者间干预措施相互影响	避免不同干预间混杂	较个体随机设计更复杂;所需研究总人数较大;组间可比性差,常需要某种匹配和分层;分析时需调整群效应;统计效率比较低
类试验	较难做到随机分配时;评价非随机事件	简化试验性研究	忽略指标的变化趋势及对对照组指标或结局的自然变化;选择性偏倚
数学模型	群体中个体参与率低等无法开展其他试验性研究	综合考虑现有流行病学资料	建立在可得到数据及假设基础上
观察性研究 (基于电子病历)	国家甚至国际范围内研究;地区研究结果间比较的大样本量人群研究	多中心大样本;避免小样本研究产生的偏倚	电子病历记录的完整性;临床一体化网络;统一数据收集格式

倚<sup>[19]</sup>。首先可通过一个相对延长的基线观察期评估结局指标;再使用一些非等价的控制点,即设立非随机对照组,不施加任何干预措施,必要时在实施干预措施前对一些重要的特征,如结局测量指标等进行对照组与干预组匹配,以减少选择性偏倚;最后是收集混杂变量的数据。在统计分析上,为具体研究某一指标的变化趋势,可进行分层回归分析;为控制潜在的混杂偏倚和选择偏倚,可采用多因素回归分析、倾向分数等。

(4)数学模型:类试验和整群随机对照试验的共同点,就是可在某一单位水平进行干预。因此,保证单位内所有个体的参与很重要。当某一单位个体参与率很低,无法开展单位水平的干预,或当某项比较研究无法开展临床试验时,可采用数学模型进行鉴别、设置优先项及完善设计,甚至进行实际的疗效比较<sup>[20]</sup>。数学模型是在假定存在统计学联系的基础上,预测特定的结局。其局限性体现在数据资料和假设上,而其优势是考虑了现有流行病学资料,并通过假设,得出干预在何种条件下可获得最大的收益。

2. 观察性研究:进行CER开展大规模临床试验的同时,也应侧重于综合分析现有的医疗数据记录。医院建立的数据库拥有大量医疗记录,可作为试验外的主要数据来源。由于手工收集纸质记录和不同地方的资料费用巨大,且难以避免家庭或医生小样本研究产生的偏倚,如所有临床数据均被存入网络数据库,通过访问电子病历(electronic medical record),采用网络数据产生不同地区比较的多中心大样本研究结果。目前费城儿童医院(Children's Hospital of Philadelphia)2003年启动的儿科研究联盟(Pediatric Research Consortium, PeRC)和美国儿科学会(American Academy of Pediatrics)的电子儿科研究(Electronic-Pediatric Research in Office Settings, ePROS)两个大型儿科初级护理研究网络已积极发展应用电子病历进行CER。

电子病历实施CER的方法中其代表性有三种<sup>[21]</sup>。一是收集各地电子病历分析回答某一特定临床问题,如PeRC的一个研究组分析预防性使用抗菌素与再发尿路感染危险的相关性,应用基于网络的电子病历资料,共有74 974名儿童符合纳入标准。研究发现首发尿路感染后采用抗菌治疗并不会降低尿路感染再发风险,反而增加以后细菌耐药的风险<sup>[22]</sup>。为验证该结果的真实性和可靠性,研究人员回顾了纸质数据的研究结果,发现电子病历与纸质数据相比,敏感度和特异度均大于95%,证实了临床背景下

电子病历研究的完整性及可靠性。二是为某一项目招募研究对象。当患者年龄、性别、临床特点等符合纳入标准时,临床医生在获得患者知情同意后还要对受试者随访。该方法更适用于定期门诊患者,可解决研究对象居住地偏远随访困难和费用高等问题,如招募对该项研究感兴趣的患者,则可明显提高研究效率。使用电子病历招募受试者虽可提高患者/家庭对研究的参与率,但存在选择偏倚。三是验证干预有效性。如对某地区进行免疫宣传干预,医疗机构在临床实践中记录就诊人群的免疫情况,研究发现与干预前相比,干预后健康就诊者中接受至少一次疫苗注射人群占应接种疫苗人群的比例从78.2%上升至90.3%,患者从11.3%增加到32.0%,增加了近2倍,由此验证干预措施的有效性<sup>[23]</sup>。

尽管上述研究已证实应用电子病历进行CER的可行性,但在实际应用中仍存在多重障碍,如电子病历记录不完整,医疗机构间数据资源共享的网络系统尚未建立等。故在医生诊疗实践中应建立健全电子病历记录程序,应全面记录如药物治疗及过敏等重要的临床信息。考虑到网络系统在多中心研究中发挥着重要作用,应进一步完善医疗机构的网络系统,努力做到临床网络一体化,并强化各大网络系统的合作,在条件具备情况下,实现家庭电子病历网络建设。建立地区或全国范围的标准化临床数据总库,统一数据收集标准,从各医疗机构的网站上提取电子病历数据资料,在安全的服务器上将这些数据标准化及汇总,随着患者不断就诊,数据持续累积,并且在整个过程中涉及到敏感信息时均采用匿名形式。这样不仅可以开展常见病的CER,也可将其应用于罕见病、罕见不良反应、家族性疾病、前瞻性研究、地区间甚至国家间比较的多中心研究。

### 三、实施过程

CER的实施过程主要包括最初的证据生成和分析、二次资料分析、对现有证据综合及相关统计学方法应用。概括为生成、综合、传播及应用证据<sup>[7, 24]</sup>。其步骤为①明确问题,即找出证据不完善的某一研究课题,对现存证据特征进行描述和必要综合,明确数据质量、潜在偏倚及研究设计上存在的问题。②生成证据。设计符合决策制定者需要的试验,通过合理设计,开展临床试验、观察性研究。临床试验通过实施随机、盲法等措施,可相对有效控制偏倚,但结论外推性较差。观察性研究更接近临床实际,研究对象的纳入限制较少、人群的异质性较大、自主选择治疗措施等造成潜在的选择偏倚可采用多元风险

模型、倾向性匹配等分析方法进行调整。数据分析可选择单变量模型、回归模型、成本效益模型、贝叶斯模型。整合数据尤其是在合并不同来源数据时应谨慎。有学者认为按照统一的系统评价标准,观察性研究与临床试验性研究使用相同的纳入标准,将有价值的研究纳入 Meta 分析<sup>[25]</sup>。当存在两种类型研究数据时,为提高统计学效力和减少偏倚,使用 Meta 回归判定异质性来源<sup>[26]</sup>。对多种治疗措施的疗效评价,可使用间接比较的网络 Meta 分析<sup>[27]</sup>。③ 证据交流。疗效比较只是第一步,为使患者获益,证据需传播并应用于实践。目前一些“治疗指南”并未推广应用,不同地区不同医院间依从性差异很大,造成“指南”与临床脱节,故准确、清晰的证据信息交流对政策制定者、医疗服务人员、患者自身制定最优的决策均有必要。2009 年美国预防服务专责小组 (USPSTF) 建议,50 岁以上女性应进行常规乳房 X 线检查;2011 年 10 月提出建议反对无症状男性常规做前列腺特异抗原检查。这些建议引起对 CER 新一轮争议,也反映出 CER 在结果认同和接受上存在信息交流障碍。对 55~74 岁吸烟者应用低剂量螺旋计算机断层扫描筛查 3 年,肺癌死亡率下降 20%<sup>[28]</sup>。然而该研究也提出一个重要问题,在定期接受 CT 扫描的患者中,假阳性率接近 40%,而筛查本身可造成伤害,并产生巨大费用,故有效进行筛查将需要更多的研究确定风险更高的吸烟者,对高危人群筛查。当前证据信息交流的主要方式是将研究结果发表在学术期刊上,这样使科研与普通大众脱节,造成将结果转化为证据时,普遍难以接受。应发挥现代电子化的优势,将那些具有广泛影响的结果转化为公众可以理解的形式,尤其是要保证实施该措施的目标人群能充分理解和接受。

此外,在方法学创新上,Diana Miglioretti 认为 CER 应着重开展创新性的研究设计、使用观察性数据进行亚组分析和调整偏倚的统计学方法<sup>[29]</sup>。

#### 四、结语

RCT 等试验性研究重视效力研究,以电子病历为基础的观察性研究可能更注重效果研究,各具优势,故在开展 RCT 同时,灵活使用设计良好的类试验、整群随机对照试验和数学模型等补充研究方法,充分利用电子病历进行多中心大样本观察研究,可更好地完成 CER。与此同时在方法学上仍需思考以下问题:如何设计观察性研究才能有效控制偏倚,使 CER 收益最大?在保证内部有效性情况下,如何使 RCT 更贴近真实临床?如何使试验性研究、观察性

研究互相补充,应用全部证据答复某一类型的问题?此外,我国具有中西医并行的医疗特点,其中中医治疗重视个性化治疗,治疗方式也多样,经常出现同病异治现象。需要进行西药、中药、中西医结合等方法的疗效比较,这大大增加了 CER 在我国的实施难度。总之,CER 可为卫生保健部门和患者提供更多信息,使其作出更好的决策,在实施过程中应注意方法学、实用性及道德伦理问题。

#### 参 考 文 献

- [1] Perloth DJ, Goldman DP, Garber AM. The potential impact of comparative effectiveness research on U.S. health care expenditures. *Demography*, 2010, 47 Suppl: S173-190.
- [2] Editorial. Comparative effectiveness research in the USA. *Lancet*, 2009, 373(9665): 694.
- [3] Institute of Medicine (IOM). Initial National Priorities for Comparative Effectiveness Research. Washington DC: Institute of Medicine, 2009. Available at <http://www.iom.edu/Reports/2009/Comparative-Effectiveness-Research-Priorities.aspx>.
- [4] Sox HC, Greenfield S. Comparative effectiveness research: a report from the Institute of Medicine. *Ann Intern Med*, 2009, 151(3): 203-205.
- [5] Slutsky JR, Clancy CM. AHRQ's Effective Health Care Program: why comparative effectiveness matters. *Am J Med Qual*, 2009, 24(1): 67-70.
- [6] Godwin M, Ruhland L, Casson I, et al. Pragmatic controlled clinical trials in primary care: the struggle between external and internal validity. *BMC Med Res Methodol*, 2003, 22(3): 28.
- [7] Morton SC, Ellenberg JH. Infusion of statistical science in comparative effectiveness research. *Clin Trials*, 2012, 9(1): 6-12.
- [8] Avorn J. Debate about funding comparative-effectiveness research. *N Engl J Med*, 2009, 360(19): 1927-1929.
- [9] Goodman SN. Quasi-random reflections on randomized controlled trials and comparative effectiveness research. *Clin Trials*, 2012, 9(1): 22-26.
- [10] Witt CM, Manheimer E, Hammerschlag R, et al. How well do randomized trials inform decision making: systematic review using comparative effectiveness research measures on acupuncture for back pain. *PLoS One*, 2012, 7(2): e32399.
- [11] Elshaug AG, Garber AM. How CER could pay for itself—insights from vertebral fracture treatments. *N Engl J Med*, 2011, 364(15): 1390-1393.
- [12] Chandra A, Jena AB, Skinner JS. The pragmatist's guide to comparative effectiveness research. *J Econ Perspect*, 2011, 25(2): 27-46.
- [13] McCaughey B. Ruin your health with the Obama stimulus plan. *Bloomberg.com*. February 9, 2009. Available at <http://www.bloomberg.com/apps/news?pid=newsarchive&sid=aLzfDxfbwhzs>.
- [14] Garber AM, Tunis SR. Does comparative-effectiveness research threaten personalized medicine? *N Engl J Med*, 2009, 360(19): 1925-1927.

- [15] Pearson SD. Placebo-controlled trials, ethics, and the goals of comparative effectiveness research: comment on "lack of head-to-head trials and fair control arms". *Arch Intern Med*, 2012, 172(3):244-245.
- [16] Dunkelgrun M, Boersma E, Schouten O, et al. Bisoprolol and fluvastatin for the reduction of perioperative cardiac mortality and myocardial infarction in intermediate-risk patients undergoing noncardiovascular surgery: a randomized controlled trial (DECREASE-IV). *Ann Surg*, 2009, 249(6):921-926.
- [17] Perencevich EN, Lautenbach E. Infection prevention and comparative effectiveness research. *JAMA*, 2011, 305(14):1482-1483.
- [18] Campbell MK, Elbourne DR, Altman DG, et al. CONSORT statement: extension to cluster randomised trials. *BMJ*, 2004, 328(7441):702-708.
- [19] Harris AD, Bradham DD, Baumgarten M, et al. The use and interpretation of quasi-experimental studies in infectious diseases. *Clin Infect Dis*, 2004, 38(11):1586-1591.
- [20] Institute of Medicine. Redesigning the Clinical Effectiveness Research Paradigm: Innovation and Practice-Based Approaches-Workshop Summary. Washington, DC: Institute of Medicine, 2010. Available at <http://www.iom.edu/Reports/2010/Redesigning-the-Clinical-Effectiveness-Research-Paradigm.aspx>.
- [21] Fiks AG, Grundmeier RW, Margolis B, et al. Comparative effectiveness research using the electronic medical record: an emerging area of investigation in pediatric primary care. *J Pediatr*, 2012 Feb. 23. [Epub ahead of print].
- [22] Conway PH, Cnaan A, Zaoutis T, et al. Recurrent urinary tract infections in children: risk factors and association with prophylactic antimicrobials. *JAMA*, 2007, 298(2):179-186.
- [23] Fiks AG, Grundmeier RW, Biggs LM, et al. Impact of clinical alerts within an electronic health record on routine childhood immunization in an urban pediatric population. *Pediatrics*, 2007, 120(4):707-714.
- [24] Ellenberg JH, Mick R, Ellenberg SS. Proceedings of the University of Pennsylvania annual conference on statistical issues in clinical trials: early, translational and proof of concept studies: the 'go/no go' decisions. *Stat Med*, 2010, 29(10):1059-1060.
- [25] Shrier I, Boivin JF, Steele RJ, et al. Should meta-analyses of interventions include observational studies in addition to randomized controlled trials? A critical examination of underlying principles. *Am J Epidemiol*, 2007, 166(10):1203-1209.
- [26] Simmonds MC, Higgins JP. Covariate heterogeneity in meta-analysis: criteria for deciding between meta-regression and individual patient data. *Stat Med*, 2007, 26(15):2982-2999.
- [27] Cipriani A, Furukawa TA, Salanti G, et al. Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. *Lancet*, 2009, 373(9665):746-758.
- [28] National Lung Screening Trial Research Team, Aberle DR, Adams AM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*, 2011, 365(5):395-409.
- [29] Nelson B. Comparative effectiveness research takes a stand on cancer screening: even if the data are there, effective communication of results is key. *Cancer Cytopathol*, 2012, 120(1):1-2.

(收稿日期:2012-05-30)

(本文编辑:张林东)

## · 消息 ·

## 中华医学会系列杂志已标注数字对象惟一标识符

数字对象惟一标识符(digital object identifier, DOI)是对包括互联网信息在内的数字信息进行标识的一种工具。

为了实现中华医学会系列杂志内容资源的有效数字化传播,同时保护这些数字资源在网络链接中的知识产权和网络传播权,为标识对象的版权状态提供基础,实现对数字对象版权状态的持续追踪,自2009年第1期开始,中华医学会系列杂志纸版期刊和数字化期刊的论文将全部标注DOI。即中华医学会系列杂志除科普和消息类稿件外,其他文章均需标注DOI,DOI标注于每篇文章首页脚注的第1项。由中华医学会杂志社各期刊编辑部为决定刊载的论文标注DOI。

参照IDF编码方案(美国标准ANSI/NISO Z39.84-2000)规定,中华医学会系列杂志标注规则如下:“DOI:统一前缀/学会标识.信息资源类型.杂志ISSN.\*\*\*\*-\*\*\*\*.年期.论文流水号”。即:“DOI:10.3760/cma.j.issn.\*\*\*\*-\*\*\*\*.yyyy.nn.zzz”。

中华医学会系列杂志标注DOI各字段释义:“10.3760”为中文DOI管理机构分配给中华医学会系列杂志的统一前缀;“cma”为中华医学会(Chinese Medical Association)缩写;“j”为journal缩写,代表信息资源类别为期刊;“issn.\*\*\*\*-\*\*\*\*”为国际标准连续出版物号(ISSN);“yyyy”为4位出版年份;“nn”为2位期号;“zzz”为3位本期论文流水号。