

DOI: 10.13288/j.11-2166/r.2016.16.009

方法学与临床评价

中医药疗效评价队列研究的方法学质量评价

王梅¹, 王建华¹, 张抗², 梁宁², 刘建平^{2*}

(1. 辽宁中医药大学, 辽宁省沈阳市皇姑区崇山东路 79 号, 110847; 2. 北京中医药大学循证医学中心)

摘要 目的 评价目前我国中医药队列研究的文献质量, 为今后高质量的队列研究设计提供方法学参考。方法 在中国知网、万方数据库、维普数据库和中国生物医学数据库, 系统检索应用队列研究评价中医药疗效的文献, 检索时间为各数据库建库至 2015 年 12 月 31 日。两名研究者独立筛选文献并进行资料提取, 对纳入文献的一般情况、疾病分类与治疗方法进行描述, 依据 Newcastle-Ottawa Scale (NOS) 量表对纳入文献的方法学质量进行评价, 并分析其队列研究的方法学要点报告情况。根据纳入文献的实际情况对其研究类型进行重新分类。结果 本研究共获得 113 篇中医药队列研究文献, NOS 评分多为 4~7 分 (103 篇, 91.15%), 平均 (5.12 ± 1.19) 分。经过对研究设计的评估并重新分类, 其中完全符合队列研究原则的仅 41 篇 (36.28%); 5 篇 (4.42%) 研究同时提到“队列研究”和“随机分组”, 应为随机对照试验; 67 篇 (59.29%) 不符合观察性队列研究的设计原则, 应为临床对照试验。结论 我国中医药临床疗效评价的队列研究方法学质量不高, 存在诸多设计与方法学问题, 并极易与其他临床试验设计相混淆, 应加强研究人员的方法学培训, 从而合理、科学地应用队列研究评价中医药的疗效。

关键词 队列研究; 中医药; 疗效评价; 方法学; 文献计量

中医药临床实践多以复杂疗法为主^[1], 中医医师综合患者的病情、体质及患者的意愿等, 结合自身的临床经验, 给予患者不同的治疗方案, 而且治疗方案常会随着疾病的发展或证候的变化而改变。目前开展的大量中医药随机对照试验 (RCT) 中多制定了严格的纳入、排除标准, 只适用于高度同质的群体, 治疗方案亦标准化, 研究环境与常规治疗相去甚远, 这些研究得出的结论应用到临床实践中具有很大困难^[2]。此外, 随机分配形成的治疗措施可能与患者的意愿相违背, 造成 RCT 在中医药复杂干预的疗效评价中具有一定的局限性^[3-4], 因此, 国内外学者开始尝试采用流行病学观察性研究对中医药临床疗效进行评价, 以提供更广泛、可靠的诊疗证据^[2,5]。

队列研究最早用于探索疾病的病因和危险因素, 近年来开始用于医疗防治措施的效果评价, 治疗性队列研究是将特定患病人群根据是否接受某种

治疗措施分为观察组和对照组, 或接受不同类别的治疗措施分为不同的组别, 然后追踪观察一定时间, 比较各组间结局事件的发生率 (如病死率) 的差异^[6]。队列研究疗效评价在循证医学的证据等级中仅亚于 RCT^[7], 有以下特点: 1) 队列研究适用于评价复杂干预的综合效果; 2) 允许个体化治疗; 3) 评价结局指标简单且易于测量, 如死亡率、生存率等; 4) 尤其适用于随机分组难以操作, 患者对治疗措施有选择偏爱的疗效评价研究^[8]。基于以上这些优势, 队列研究在中医药疗效评价中逐渐受到重视并被采用, 尤其是国家的一些重大中医药临床研究课题, 已经开始应用队列研究的方法。但国内目前所发表的中医药队列研究的现状如何, 队列研究是否被合理应用, 是中医药行业普遍关注的问题。本研究全面检索国内中文数据库发表的中医药队列研究文献, 并按照国际标准对其进行方法学质量评价, 以便了解当前我国大陆地区中医药疗效评价研究中队列研究的现状和问题, 为今后进一步开展高质量的队列研究提供依据。

基金项目: 国家中医药管理局 2012 年中医药行业科研专项 (201207007)

* 通讯作者: jianping_l@hotmail.com, (010) 64287002

1 方法

1.1 文献检索

检索国内中文数据库: 中国知网 (CNKI)、维普数据库 (VIP)、中国生物医学数据库 (Sino Med)、万方数据库 (Wan Fang)。检索词包括: 针灸, 针刺, 电针, 艾灸, 耳穴, 穴位按压, 按摩, 中医, 中西医结合, 中药, 草药, 气功, 太极, 刮痧, 拔罐, 中成药, 汤, 方; 队列研究, 随访研究, 前瞻性观察性研究, 队列观察, 注册登记研究。语言限中文, 发表形式不限。检索时间为各数据库建库至 2015 年 12 月 31 日。

1.2 纳入标准

在文献的题目或摘要中提到该研究为“队列研究”的文献; 用于中医药疗效评价, 所治疗疾病不限; 观察组的治疗措施为中医药疗法, 包括药物疗法 (中成药、中药提取物、中药制剂等) 和/或非药物疗法 (如针灸、拔罐、刮痧、太极等); 对照组治疗措施不限; 发表语种限中文。

1.3 排除标准

非疗效比较的中医药队列研究; 重复发表的或仅有摘要而无法获得全文的文献。

1.4 资料提取分析

采用 Epidata 3.1 软件构建资料提取表。提取的资料内容包括发表年份、发表类型 (期刊、学位论文、会议论文)、文献报告的队列研究类型 (包括前瞻性队列研究、回顾性队列研究和双向性队列研究)、基金资助、样本量、观察组的治疗措施、所研究的疾病病种, 并按《疾病和有关健康问题的国际统计分类 (ICD-10)》^[8] 分类。队列研究方法学评价依据 Newcastle-Ottawa Scale (NOS) 量表^[9] 逐条评价, 分值越高质量越好。除此之外, 参考观察性流行病学研究报告规范 (STROBE)^[10] 补充评价队列研究设计方法学要点的条目及分类依据, 主要包括: 队列形成的依据, 治疗方案 (辨证论治: 仅限定到“治则”或“治法”; 标准化的治疗方案: 固定了具体药物组成和用量、服用剂量和疗程等细节), 主要结局指标 (终点事件: 生存率、病死率、癌症的复发率等; 生物学指标: 生理、生化等实验室指标; 临床结局: 治愈率、有效率等), 样本量估算的依据, 统计分析方法 (单因素统计分析方法: t 检验、单因素方差分析、卡方检验等; 校正协变量的统计分析方法: 重复测量的方差分析、协方差分析、生存分析、倾向指数评

分、其他校正协变量的统计模型等)。

由两名研究者独立对纳入的文献进行资料提取及方法学质量评价, 遇有分歧时与第三位作者协商解决。采用 Excel 2007 软件对所提取资料进行描述。

2 结果

2.1 纳入文献一般情况

本研究共纳入 113 篇文献。图 1 示, 国内第一篇中医药疗效评价队列研究文献发表于 2004 年, 且自 2010 年后呈增长趋势直至 2013 年。期刊论文 65 篇 (57.52%)、学位论文 34 篇 (30.09%)、会议论文 14 篇 (12.39%); 根据纳入文献自我报告的研究类型进行分析: 前瞻性队列研究 81 篇 (71.68%)、回顾性队列研究 22 篇 (19.47%)、双向性队列研究 1 篇 (0.88%)、因文献报告不充分无法判断的 9 篇 (7.96%); 标注有基金资助的队列研究为 52 篇 (46.02%)。本研究纳入的 113 篇文献共计 21470 例观察对象, 平均 190 例。 ≤ 99 例的 35 篇 (30.97%), 100 ~ 199 例的 36 篇 (31.86%), 200 ~ 299 例的 19 篇 (16.81%), 300 ~ 399 例的 13 篇 (11.50%), ≥ 400 例的 10 篇 (8.85%)。

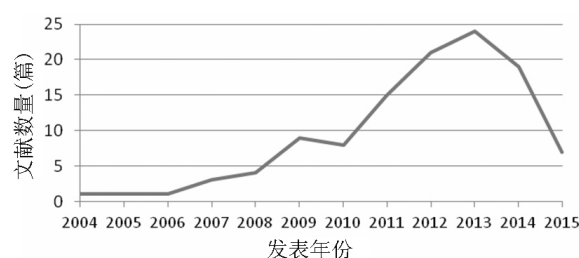


图 1 我国中文发表的中医药疗效评价队列研究趋势

2.2 疾病分类和治疗方法

本研究将纳入文献所关注的疾病按照 ICD-10 进行分类, 队列研究涉及病种广泛, 共 12 类疾病, 以肿瘤 (39 篇, 34.51%) 和循环系统疾病 (24 篇, 21.24%) 为主, 详见表 1。

本研究纳入文献观察组的治疗方法包括: 中西医结合 71 篇 (62.83%), 中药汤剂及中成药 22 篇 (19.47%), 针灸 13 篇 (11.50%), 中医综合疗法 4 篇 (3.54%), 穴位埋线 2 篇 (1.77%), 八段锦 1 篇 (0.88%)。

2.3 方法学质量评价

将纳入文献采用 NOS 量表对方法学质量进行逐条评价, 本研究纳入的 113 篇文献有 64 篇 (占 56.64%) 进行了随访, 因此, 随访时间的充分性

表 1 纳入的 113 篇文献涉及的疾病分类

ICD-10 编码	疾病分类	文献 篇(%)]
C00-D48	肿瘤	39 (34. 51)
I00-I99	循环系统疾病	24 (21. 24)
A00-B99	某些传染病和寄生虫病	12 (10. 62)
M00-M99	肌肉骨骼系统和结缔组织疾病	8 (7. 08)
N00-N99	泌尿生殖系统疾病	10 (8. 85)
J00-J99	呼吸系统疾病	7 (6. 19)
K00-K93	消化系统疾病	5 (4. 42)
D50-D89	血液及造血器官疾病和某些涉及免疫机制的疾病	3 (2. 65)
G00-G99	神经系统疾病	2 (1. 77)
F00-F99	精神和行为障碍	1 (0. 88)
R00-R99	症状、体征和临床与实验室异常所见，不可归类在他处者	1 (0. 88)
S00-T98	损伤、中毒和外因的某些其他后果	1 (0. 88)

和随访的完整性两个条目仅对此 64 篇文献进行评价，其他 49 篇关于随访的两个条目均记为 0 分。结果表明，NOS 评分 9 分的 0 篇，8 分的 1 篇 (0. 88%) ， 7 分的 15 篇 (13. 27%) ， 6 分的 20 篇 (17. 70%) ， 5 分的 49 篇 (43. 36%) ， 4 分的 19 篇 (16. 81%) ， 3 分的 7 篇 (6. 19%) ， 2 分、1 分的各 1 篇 (0. 88%) ， 平均 (5. 12 ± 1. 19) 分。项目报告情况见表 2。

2.4 方法学要点评价

针对队列研究的主要方法学设计要点提取资料并进行归类，真正意义上的队列研究仅 41 篇 (36. 28%) 。 5 篇 (4. 42%) 文献同时提到“队列研究”和“随机分组”，显然混淆了“队列研究”和“随机对照试验”的概念或进行了错误的描述。 67 篇 (59. 29%) 不具备观察性队列研究的设计原则，其中 37 篇文献剔除了依从性差的或违背治疗方法的病例，30 篇文献因样本量小，且经过严格的纳入，排除标准筛选病例，未进行随访或随访不充分，采用单因素的统计分析方法，被判定为临床对照试验。

11 篇 (9. 37%) 文献报告了研究队列的形成依据参与者自愿选择的疗法； 65 篇 (57. 52%) 文献采用了辨证论治； 46 篇 (40. 71%) 文献报告了终点事件； 10 篇 (8. 85%) 文献有明确的样本量估算公式； 33 篇 (29. 20%) 文献采用了校正协变量的统计分析方法； 4 篇 (3. 54%) 文献采用了连续收集病例的方法， 52 篇 (56. 02%) 文献报告了安全性结局指标， 3 篇 (2. 65%) 采用了意向性分析 (ITT) 的方法， 7 篇 (6. 19%) 报告了缺失数据是如何处理的。

表 2 纳入的 113 篇文献 NOS 条目报告情况

条目	评分依据	文献 篇(%)]
暴露组的代表性	真正代表人群中暴露组的特征 (多个社区)	1 (0. 88)
	一定程度上代表了人群中暴露组的特征(单个社区)	5 (4. 42)
	选取某类人群 (住院或门诊病例)	102 (90. 27)
非暴露组的选择	没有说明来源	5 (4. 42)
	与暴露组的来源相同	90 (79. 65)
	与暴露组的来源不同	18 (15. 93)
暴露的确定	没有说明非暴露组的来源	5 (4. 42)
	固定的档案记录	112 (99. 12)
	参与者的自我报告	1 (0. 88)
研究开始前无预期事件发生	无预期事件发生	47 (41. 59)
	因文章报告不充分无法判断	66 (58. 41)
组间可比性	研究控制了最重要的因素，如年龄	105 (92. 92)
	研究控制了其他重要的混杂因素	85 (75. 22)
结局事件评估	两组不可比	7 (6. 19)
	未进行两组比较	11 (9. 73)
	盲法独立评价	97 (85. 84)
随访时间是充分*	有档案记录	2 (1. 77)
	自我报告	4 (3. 54)
	两者结合	3 (2. 65)
随访的完整性	未描述	7 (6. 19)
	充分	32 (23. 32)
	不充分 (无随访)	8 (7. 08)
随访的完整性	不确定	24 (21. 24)
	全部随访	12 (10. 62)
	少数失访，但不大可能引入偏倚	30 (26. 55)
随访的完整性	多数失访	1 (0. 88)
	未描述随访情况	21 (18. 58)

注：*，共 64 篇进行了随访，故随访时间是否充分仅对此 64 篇文献进行评价

3 讨论

近几年文献发表量逐渐增多，研究所关注的疾病和疗法较为广泛，但方法学质量不高，本研究共纳入 113 篇中医药队列研究疗效评价的文献，NOS 评分平均为 (5. 12 ± 1. 19) 分，且通过队列研究主要设计要点的分析，一部分文献的研究设计并非真正的队列研究，应为临床对照试验。

在实际的临床科研中易与临床试验相混淆，主要体现在以下几个方面： 1) 队列研究的分组是自然形成的，参与者是否接受某种治疗是研究者不能控制的，本研究纳入文献仅 11 篇报告是按参与者意愿选择的治疗措施，其余文章没有说明分组依据，因此，不能判断医生本人作为研究者是否在分组时有人为干预的影响。 2) 队列研究是在自然状态下

“真实世界”的研究,病例的收集不应剔除违背方案者^[2],此时连续收集病例能够真实反映客观实际,本研究中仅 4 篇文献是连续收集病例,37 篇文献剔除了依从性差的病例,违背了观察性研究的原则。3) 队列研究宜选择便于客观测量的、明确的终点事件作为主要结局指标,如生存、死亡等。但本研究中报告终点事件的文献仅 46 篇(40.71%),另有一部分以生物学指标改变作为主要结局指标,不能排除研究者出于研究目的增设了不必要的临床检查项目来评价其治疗效果,同时这些指标不能充分反映出中医学整体调节对人群远期健康的影响。4) 队列研究在研究过程中不宜施加过多的人为控制,极易受到混杂因素的影响,因此,在统计分析方法上也不同于临床试验,需进行协变量校正,可采用 Logistic 回归或 Cox 比例风险模型^[2]。对于评价终点事件的长期随访数据应采用生存分析,对于基线不均衡的情况可采用配对或倾向指数评分进行分析^[3]。本研究纳入的文献仅 23 篇在统计分析过程校正了混杂因素的影响,其余文献仅使用了 *t* 检验、卡方检验等单因素统计分析方法,今后在设计高质量的队列研究中要加强对混杂因素影响的控制。5) 队列研究的样本量确定需要依据队列研究的样本量估算公式^[4],还需考虑统计分析方法的选择导致样本量的减少(如倾向指数评分)、长期随访引起的失访以及缺失数据等^[3],应当综合这些因素正确评估实际进入分析的样本量是否充分。本研究中仅有 10 篇文献报告了样本量估算过程,其余文献未提及,且纳入的文献样本量普遍偏少, <300 例为 90 篇。这些研究很难保证足够的样本量来回答研究提出的问题,因此建议未来治疗性队列研究应当进行科学的样本量估算,保证结论的可靠性、可信性。

NOS 量表由澳大利亚 Newcastle 大学和加拿大 Ottawa 大学联合起草,目的是评价观察性研究(队列研究和病例对照研究)的方法学质量^[4]。本研究纳入的 113 篇文献 NOS 评分主要集中在 4~7 分,得分较低的条目是暴露组代表性和随访情况。在本研究中,6 篇文献的样本来源于社区,具有很好的代表性,102 篇文献通过医院收集门诊或住院病例,会因为患者病情不同而引入选择性偏倚。对于随访情况,64 篇文献(56.64%)进行了随访,仅 12 篇文献报告全部参与者随访情况。在实际研究中,参与者住址的变动、依从性低、不愿意配合随访等因素会对随访的完整性带来困难。因此,为

尽可能控制失访偏倚,一方面,随访方式应简化,如参与者来院复查或电话、电子邮件、信件随访;另一方面,研究者应完整地报告全部参与者的随访情况,并将退出或失访的病例纳入统计分析。

本研究是根据文献的报告情况来判断方法学质量,未联系原作者获取研究方案或询问设计中相关的条目,可能存在实际研究方法与文章报告不符合的情况。文章报告的质量不完全等同于其设计的质量或实施的质量,有些文章报告质量好并不一定能说明其方法学质量也好,有些文章可能由于杂志版面限制的影响,研究细节在正文中报告不全面,这也可能导致本研究结果代表性受限。

综上所述,中医药临床疗效评价队列研究的设计存在较大问题,研究者极易将其与临床试验相混淆,甚至存在“队列研究”方法误用的情况,建议加强队列研究方法学的培训,提升研究人员的科研能力,从而使中医药疗效评价合理地采用队列研究的设计并确保其质量。

参考文献

- [1] 刘建平. 循证中医药临床研究方法 [M]. 北京: 人民卫生出版社, 2009: 144-145.
- [2] CLAUDIA M. WITT, 黄文静, 劳力行, 等. 我们需要什么研究来支持将结合医学运用于临床的决策? 效益比较研究可以填补其间的沟壑吗? [J]. 中国中西医结合杂志, 2013, 33(8): 1030-1035.
- [3] NORRIS SL, ATKINS D. Challenges in using nonrandomized studies in systematic reviews of treatment interventions [J]. Ann Intern Med, 2005, 142(12 Pt 2): 1112-1119.
- [4] CHOU R, ARONSON N, ATKINS D, et al. AHRQ series paper 4: assessing harms when comparing medical interventions: AHRQ and the effective health-care program [J]. J Clin Epidemiol, 2010, 63(5): 502-512.
- [5] NORRIS SL, ATKINS D, BRUENING W, et al. Observational studies in systemic reviews of comparative effectiveness: AHRQ and the Effective Health Care Program [J]. J Clin Epidemiol, 2011, 64(11): 1178-1186.
- [6] 刘建平. 队列研究的设计、实施及方法学要点 [J]. 中国中西医结合杂志, 2008, 6(4): 331-336.
- [7] 刘建平. 传统医学证据体的构成及证据分级的建议 [J]. 中国中西医结合杂志, 2007, 27(12): 1061-1065.
- [8] 李赞华, 万霞, 刘建平. 队列研究与随机对照试验的方法学比较 [J]. 北京中医药大学学报(中医临床版), 2008, 15(5): 11-14.
- [9] 世界卫生组织. 疾病和有关健康问题的国际统计分类 (ICD-10) [M]. 北京: 人民卫生出版社, 1996.
- [10] GA WELLS, B SHEA, D O'CONNEL, et al. The

Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomized studies in meta-analyses [M/OL] . [2016 - 02 - 01] . <http://www.ncbi.nlm.nih.gov/books/NBK126179/>.

- [1] VON ELM E, ALTMAN DG, EGGER M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for reporting observational studies [J] . Int J Surg, 2014, 12 (12) : 1495-1499.
- [2] 费宇彤, 张颖, 刘建平. 再论“队列研究”在中医药临床疗效评价中的应用 [J] . 世界中医药, 2014, 9 (10) :

1261-1263.

- [3] 邢建民, 费宇彤, 陈薇, 等. 观察性研究在中医临床研究中的应用(1) : 队列研究方法与设计 [J] . 中医杂志, 2008, 49 (6) : 502-503.
- [4] 王家良. 临床流行病学 [M] . 4 版. 上海: 上海科学技术出版社, 2014: 180.
- [5] VELENTGAS P, DREYER NA, NOURJAH P, et al. Developing a protocol for observational comparative effectiveness research: a user's guide [M/OL] . (2013 - 02 - 12) [2015 - 07 - 30] . <http://www.ncbi.nlm.nih.gov/books/NBK49468>.

Evaluation of Methodological Quality of Cohort Study on Chinese Medicine Effect Assessment

WANG Mei¹, WANG Jianhua¹, ZHANG Kang², LIANG Ning², LIU Jianping²

(1. Liaoning University of Chinese Medicine, Shenyang, 110847; 2. Evidence-based Medicine Centre, Beijing University of Chinese Medicine)

ABSTRACT Objective To evaluate quality of current literature on Chinese medical cohort studies in China, and to provide methodological reference for further high-quality cohort study design. **Methods** Literature on effect evaluation of Chinese medicine by cohort study was systematically searched from China National Knowledge Infrastructure, Wan Fang Data, VIP Data and Chinese Biological Medical Literature Database on disc from the date of database inception until December 31, 2015. Two researchers screened the papers independently and extracted data. General conditions, classification of diseases and treatment methods were described. The methodological quality of included papers was evaluated with Newcastle-Ottawa Scale (NOS) . The reported methodological key points of cohort studies were analyzed. The research types were reclassified according to contents of included papers. **Results** Totally 113 papers on Chinese medicine cohort study were included in this research. NOS scores of most papers were 4 - 7 (103 papers, 91. 15%) , and (5. 12 ± 1. 19) on average. After the research design was evaluated and papers were reclassified, only 41 papers (36. 28%) were completely in line with the principles of cohort study. Five papers (4. 42%) mentioned both “cohort study” and “randomization”, which should be randomized controlled trials. Sixty-seven papers (59. 29%) were not in accord with the design principles of observational cohort study, which should be clinical controlled trials. **Conclusion** The methodological quality of cohort study on Chinese medicine clinical effect evaluation in China was poor. There were many problems in study design and methodology. It was easy to confuse cohort study with other clinical trial design. The methodological training for researchers should be enhanced. Thus, Chinese medicine effect might be evaluated by cohort study reasonably and scientifically.

Keywords cohort study; Chinese medicine; effect evaluation; methodology; bibliometrics

(收稿日期: 2015 - 11 - 29; 修回日期: 2016 - 04 - 16)

编辑: 焦 爽