

• 循证理论与实践 •

Meta 分析系列之三: 随机对照试验的质量评价工具

曾宪涛, 包翠萍, 曹世义, 刘菊英

[中图分类号] R4 [文献标识码] A [文章编号] 1674-4055(2012)03-0183-03

Meta 分析是对原始研究的二次综合分析与评价, 其质量受纳入原始研究质量、评价的方法、评价者的知识水平及观点的影响。若 Meta 分析纳入的原始研究质量低, 且未对原始研究方法学质量进行正确评价, Meta 分析的结果和结论可能是不正确的, 从而对临床实践造成误导^[1]。因此, 对 Meta 分析纳入的原始研究质量进行严格的评价尤为重要。原始研究质量评价工具主要根据研究设计、实施、结果分析整个过程中可能出现偏倚的各个因素而确定, 目前出现的评价工具较多, 主要有单个评价条目 (components/items)、清单 (checklist/list) 和量表 (scale) 三种类型。前面我们已经对 Meta 分析的类型^[2]、Meta 分析的软件^[3]进行了介绍, 本文将在简要回顾原始研究类型的基础上, 对随机对照试验 (randomized controlled trial, RCT) 常用的质量评价工具进行简要分析。

1 临床研究类型简介

临床研究按照是否存在人为干预, 分为实验性研究 (experimental study) 和观察性研究 (observational study) 两大类, 观察性研究亦称为非实验研究 (non-experimental study)。图 1 展示了两类研究包括的类型^[4-5]。此外, 为适应实际情况的需要, 又发展出了新的研究类型, 如: 巢式病例-对照研究 (nest case-control study) 病例-队列研究 (case-cohort study), 病例-病例研究 (case-case study), 自身交叉对照研究 (self-cross controlled study) 单纯病例研究 (case study), 前瞻性随机对照试验 (prospective randomized controlled study), 回顾-前瞻性队列研究 (retrospective-prospective study) 等。

2 常用的 RCT 质量评价工具

2.1 Cochrane 风险偏倚评估工具 长期以来, 研究质量评价和偏倚风险评价被认为是等同的, 但 Cochrane 系统评价手册认为“研究质量”和“研究偏倚”是有区别的, “偏倚”能更真实的反映研究存在的缺陷, 故在 5.0 及以上的版本中命名为“Cochrane 风险偏倚评估工具 (Cochrane collaboration's tool for assessing risk of bias)”, 要求 Cochrane 系统评价的制作采用, 并不再推荐任何一种评价 RCT 质量的清单或量表^[1]。该工具主要从 6 个领域对偏倚风险进行评价 (domain-based evaluation), 对每条指标采用“低度偏倚”、“不清楚”、“高度偏倚”进行判定 (表 1)。

2.2 PEDro 量表 PEDro (物理治疗证据数据库, Physiotherapy Evidence Database, PEDro) 是由位于乔治中心的物理治疗循证中心 (CEBP) 的一群临床和物理治疗师于 1999 年所建立,

旨在帮助使用最佳证据于临床应用来强化物理治疗服务的有效性。PEDro 量表 (PEDro scale) 是 CEBP 基于 Delphi 清单^[6]制作的 RCT 评价量表, 上次修订的时间是 1999 年 6 月 21 日, 包括 11 条, 每个条目为 1 分^[7]: (1) 受试者的纳入条件有具体说明; (2) 受试者被随机分配到各组 (在交叉研究中, 受试者的治疗顺序是随机安排的); (3) 分配方式是隐藏的; (4) 就最重要的预后指标而言, 各组在基线都是相似的; (5) 对受试者全部设盲 (实施盲法); (6) 对实施治疗的治疗师全部设盲 (实施盲法); (7) 对至少测量一项主要结果的评定者全部设盲 (实施盲法); (8) 在最初分配到各组的受试者中, 对 85% 以上的人进行至少一项主要结果的测量; (9) 凡是有测量结果的受试者, 都必须按照分配方案接受治疗或者对照条件, 假如不是这样, 那么应对至少有一项主要结果进行“意向治疗分析”; (10) 对至少一项主要结果的组间统计结果作出报告; (11) 研究将提供至少一项主要结果的点测量值和变异测量值。

2.3 Delphi 清单 Delphi 清单 (Delphi list) 由马斯特里赫特大学 Verhagen 及其同事 1998 年依据 Maastricht 和 Chalmers 量表^[8]遵照 Delphi 法制作的, 用以评价 RCT 质量, 共 8 个条目, 其中第一条又包括了两个部分^[6], 均采用“是”、“否”及“不知道”进行判定: (1) 治疗分组: ①是否采用了随机分组? ②是否采用了分配隐藏? (2) 各组最重要的预后指标的基线是否相似? (3) 是否具体说明了受试者的入选条件? (4) 是否对研究结果评价者实施盲法? (5) 是否对研究实施者实施盲法? (6) 是否对受试者实施盲法? (7) 主要结局指标的点测量值和变异测量值是否提供? (8) 分析中是否包含了“意向治疗分析”?

2.4 CASP 清单 CASP (Critical Appraisal Skills Programme) 于 1993 年在英国牛津大学成立, 由 NHS 专项基金资助, 旨在倡导运用循证的方法来阅读论文, 以真正服务于医疗卫生保健及社会保健研究, 并制作了用来评价研究质量的 CASP 清单 (CASP Checklist)。其中, 用于评价 RCT 的清单包括 11 个条目^[9], 其中前 3 条是筛选问题, 1~6 和 9 条均用“是”、“否”及“不知道”判定, 第 10 和 11 条用“是”、“否”判定 (表 2)。

2.5 Jadad 量表 Jadad 量表 (Jadad scale)^[10] 由 Jadad 等 1996 年发布, 其最初目的是为了评价疼痛治疗的 RCT 的质量, 从随机方案及其隐匿、盲法、退出与失访病例的原因及例数这 3 个方面进行评价, 采用 0~5 分记分法, ≤2 分认定为低质量研究, ≥3 分则认为质量较高 (表 3)。

2.6 Chalmers 量表 Chalmers 量表 (Chalmers scale)^[8] 是由 Chalmers 等 1981 年首次尝试用评分量表的方式对临床试验质量作评价, 分为一般情况记录、研究方案、统计学分析、结果陈述四个部分共 36 个条目。其中一般情况记录不参与计分, 试验设计、统计分析和结果三个部分按 6:3:1 权重记分, 共有 27 个条目, 计 100 分, 报告条目的选择主要根据作者累积的经验, 因其现在应用较少, 故不具体列出。

2.7 CONSORT 声明 CONSORT 声明 (Consolidated Standards of Reporting Trials statement)^[11] 由报告 RCT 必备的基本项目清单和描述整个试验过程中受试者流程的流程图组成, 主要

基金项目: 湖北医药学院 2011 年度优秀中青年科技创新团队项目 (2011CZX01)

作者单位: 442000 十堰, 湖北医药学院附属太和医院口腔医学中心 (曾宪涛); 麻醉科 (刘菊英); 天津医科大学公共卫生学院流行病学与卫生统计学教研室 (包翠萍); 华中科技大学同济医学院流行病学与卫生统计学系 (曹世义)

通讯作者: 刘菊英, E-mail: liu6119@163.com

doi: 10.3969/j.issn.1674-4055.2012.03.003

表 1 Cochrane 风险偏倚评估工具

领域	评价内容
选择偏倚 (selection bias)	
随机序列产生	详细描述了产生随机分配序列的方法 ,以便评估组间可比性
分配隐藏	详细描述了隐藏随机分配序列的方法 ,以便判断干预措施分配情况是否能预知
实施偏倚 (performance bias)	
对研究者和受试者施盲	详细描述了对研究者和受试者实施盲法的方法 ,以防其知晓受试者的干预措施。提供了判断盲法是否有效的信息
测量偏倚 (detection bias)	
研究结果盲法评价	详细描述了对研究结果评价者实施盲法的方法 ,以防其知晓受试者的干预措施。提供了判断盲法是否有效的信息
随访偏倚 (attrition bias)	
结果数据的完整性	完整性的报告了每个主要结局指标的数据 ,包括失访及退出的。是否明确报道了失访及退出 /每组人数 (与随机入组的总人数相比) ,失访/退出的原因 ,以便系统评价者行相关的处理
报告偏倚 (reporting bias)	
选择性报告研究结果	描述的信息可供系统评价者判断选择性报告研究结果的可能性及相关情况
其他偏倚 (other bias)	
其他偏倚来源	除上述偏倚外 ,提供的信息是否可评估存在其他引起偏倚的因素。若是先在计划书 (protocol) 中提到某个问题或因素 ,需给出对应的回答

表 2 评价 RCT 质量的 CASP 清单

领域	条目	提示
研究结果可靠吗?	研究是否提出了清晰明确的问题?	研究的人群
		采取的干预措施
		比较的方法及对照的选择
		可能的结局
	干预组和对照组的分配是否随机?	/
研究结果是什么?	纳入研究的所有患者能否恰当的解释试验的结论?	随访是否完整
		纳入分析的患者是否随机抽取
	是否对研究对象、研究者、分析者采取盲法?	患者盲法
		分析者盲法
		全体研究者盲法
研究结果有用吗?	各组之间在试验开始时的基线数据是否相似?	主要是影响试验结局的因素,比如年龄、性别、社会地位等
		/
	除研究的干预措施外,试验组和对照组的处理因素是否相同?	测量了哪些结局指标
		其可信区间是多少
研究结果有用吗?	干预效果有多大?	纳入试验的研究人群是否与你所研究的人群相似
	干预效果评价的精确性?	如果没有,是否会影响决策
	试验结果能否适用于局域人群?	这个问题不大可能在研究中提及,但你是怎么考虑的

表 3 Jadad 量表

评价指标	内容	记分
随机	未随机/不清楚/假随机	0
	提及“随机”但未描述具体的随机方法	1
	采用“随机”并描述了正确的随机方法	2
盲法	未实行双盲/假双盲	0
	提及“双盲”但未描述具体的随机方法	1
	采用“双盲”并描述了正确的施盲方法	2
失访/退出	未提及	0
	对退出/失访的例数和理由进行了详细的描述	1

针对的是两组平行设计的 RCT。该声明的目的是为了提高 RCT 的报告质量,让读者了解试验的操作过程,评价其真实性;读者、审稿人和编辑亦可使用 CONSORT 来帮助评估和解释 RCT 报告。特别是 2010 年修订版增加了选择性报告结果的条目,我国边振甲等在此基础上发展了针对中医药的

CONSORT 声明^[12]。故 CONSORT 不是专用于评价 RCT 质量的工具,但可以用来对 RCT 进行质量评价,在此不作具体列出。

3 结语

Moher 等^[13]在 1996 年鉴定出了用于评价 RCT 质量的 9 种清单和 60 余种量表,分别有 3~57 个条目,需要耗时 10~45 分钟完成。这些工具中,Chalmers 量表^[6]是相对成熟的,故在一定时间内得到了推广应用,但因实施起来繁琐,需要时间较长;其它工具将研究的设计和实施的真实性与报告质量相混淆,甚至根据作者报告的内容而非根据实施的方法是否恰当来评分,因此易受文献报告质量的影响及主观因素的制约。在此基础上,Jadad 于 1996 年发布了 Jadad 量表^[10],其最大的优点在于直接评价那些经过验证的、与试验效应估计中的偏倚有直接关系的试验特征,且简单明了,故迅速得到了广泛认可及使用,成为评价 RCT 质量的重要工具。但

Cochrane 系统评价员手册 5.0 及以后版本中指出了该量表的诸多不足,如该量表没有考虑到随机研究中很重要一个潜在偏倚-“分配隐藏”的重要性,而不推荐使用^[1]。Delphi 清单^[6]是严格遵循了 Delphi 法,从原本比较公认的量表^[8]中选择合适的条目,制作出的包括内部有效性、外部有效性和统计分析三个方面的 RCT 质量评价工具,但此量表最大的不足在于仅基于专家共识 (expert consensus) 而非实验数据 (empirical data)。PEDro 量表^[7]是基于 Delphi 清单^[6]制作的专用于评价物理治疗方面的 RCT 质量的工具,比 Delphi 清单多了第 8 条和第 10 条,在干预性研究方面应用也越来越广泛。CASP 清单有多个,RCT 只是其中之一^[9],目前应用最多的国家是英国及加拿大。CONSORT 声明^[11]主要是用来规范报告 RCT 的格式且繁琐,虽可用来评价 RCT 的质量,但使用的较少。由 Cochrane 协作网的方法学家、编辑和系统评价员共同制作及不断完善的 RCT 质量评价工具在 5.0 版本中命名为“偏倚风险评估”工具^[1],不仅要求采用文字和表格描述,还要求采用更形象、直观的图形来反映偏倚的情况;同时,减少了评估者主观因素的影响,保证了评估的可靠性。

总之,这些常用的 RCT 评价工具各有优势,我们建议在针对 RCT 的质量评价中选择 Cochrane 风险偏倚评估工具,在针对物理疗法的 RCT 中还可选择 PEDro 量表;在评价原始研究质量时则推荐选择 CASP 清单,因为它考虑了研究的价值。

参 考 文 献

- [1] Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [EB/OL]. The Cochrane Collaboration, 2011 [2012-03-30]. <http://www.cochrane-handbook.org>.
- [2] 曾宪涛,冷卫东,郭毅,等. Meta 分析系列之一: Meta 分析的类型[J]. 中国循证心血管医学杂志, 2012, 4(1): 3-5.
- [3] 曾宪涛, Kwong JSW, 田国祥, 等. Meta 分析系列之二: Meta 分析的软件[J]. 中国循证心血管医学杂志, 2012, 4(2): 89-91.

- [4] Grimes DA, Schulz KF. An overview of clinical research: the lay of the land [J]. *Lancet*, 2002, 359(9300): 57-61.
- [5] 王家良. 临床流行病学: 临床科研设计、测量与评价 [M]. 3 版. 上海: 上海科学技术出版社, 2009.
- [6] Verhagen AP, de Vet HC, de Bie RA, et al. The Delphi list: a criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus [J]. *J Clin Epidemiol*, 1998, 51(12): 1235-1241.
- [7] Centre for Evidence-Based Physiotherapy at The George Institute for Global Health. Physiotherapy Evidence Database (PEDro) [CP/OL]. (1999-06-21) [2012-03-30]. <http://www.pedro.org.au/english/downloads/pedro-scale>.
- [8] Chalmers TC, Smith H Jr, Blackburn B, et al. A method for assessing the quality of a randomized control trial [J]. *Control Clin Trials*, 1981, 2(1): 31-49.
- [9] Critical Appraisal Skills Programme (CASP). RCT checklist_14. 10.10 [EP/OL]. [2012-03-30]. http://www.casp-uk.net/wp-content/uploads/2011/11/CASP_RCT_Appraisal_Checklist_14oct10.pdf.
- [10] Jadad AR, Moore RA, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary [J]? *Control Clin Trials*, 1996, 17(1): 1-12.
- [11] Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials [J]. *Int J Surg*, 2011, 9(8): 672-677.
- [12] Bian Z, Liu B, Moher D, et al. Consolidated standards of reporting trials (CONSORT) for traditional Chinese medicine: current situation and future development [J]. *Front Med*, 2011, 5(2): 171-177.
- [13] Moher D, Jadad AR, Nichol G, et al. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists [J]. *Control Clin Trials*, 1995, 16(1): 62-73.

(收稿日期: 2012-04-02; 修回日期: 2012-05-26)

· 循证视窗 ·

老年多支冠脉病变患者血管重建策略疗效的对比研究

《新英格兰医学杂志》(*New England Journal of Medicine*) 在 2012 年 4 月 19 日发表了一项关于老年多支冠脉病变患者血管重建策略疗效研究的结果。该项观察性研究表明: 对于不需要急诊治疗的多支冠脉病变的老年患者, 采取冠状动脉旁路移植 (CABG) 治疗的患者, 其生存期长于经皮冠脉介入 (PCI) 的患者 (*N Engl J Med*, 2012, 366: 1467-1476)。

PCI 和 CABG 的临床疗效孰优孰劣一直困扰着医学界, 美国心脏病学会基金会 (ACCF) 和胸外科医师协会 (STS) 合作对 PCI 和 CABG 术后患者长期存活率进行了比较研究。

研究者链接了 ACCF 国家心血管数据登记库和 STS 成人心脏手术数据库, 查取了 2004 年至 2008 年医疗补助和医疗保险服务中心的相关数据。研究者采用了倾向评分及逆

向概率加权法来修正试验结果以减少治疗选择性偏倚。

研究纳入了年龄 65 岁及以上患有两支或三支冠状动脉疾病, 但没有急性心肌梗死的受试者, 其中 86 244 例患者接受 CABG 治疗, 103 549 例患者接受 PCI 治疗。随访期中位数为 2.67 年。在第 1 年, 两组间修正后的死亡率没有显著性差异 (CABG 组: 6.24%; PCI 组: 6.55%; 危害比: 0.95; 95% 可信区间 [CI]: 0.90 ~ 1.00)。第 4 年, CABG 组的死亡率较 PCI 组低 (16.4% vs. 20.8%; 危害比: 0.79; 95% CI: 0.76 ~ 0.82)。研究者采用敏感性分析评估了残余混杂 (residual confounding), 并运用几种不同的分析方法对若干亚组进行分析也得出了相似的结论。

(田国祥编译)