

医学科研中的系统评价

张小年

[ 关键词 ] 医学 ; 科研 ; 系统评价  
中图分类号 : R 33 文献标识码 : B 文章编号 : 1006-9771 ( 2005 ) 01-0076-02  
[ 本文著录格式 ] 张小年 . 医学科研中的系统评价 [ J ] . 中国康复理论与实践 , 2004 , 10 ( 8 ) : 76 - 77 .

随着医学科学的迅速发展 , 时刻在更新的医学信息使医疗服务的提供者、研究人员、决策者及患者感到越来越无所适从。首先 , 单一的一项研究可能因方案设计、样本量等方面存在缺陷而使结果不够可靠 ; 其次 , 关于同一临床问题的多项研究可能得出截然相反的结论。由此产生的后果是 , 一方面 , 现存研究已经证实的结论难以为大多数临床医师所知、所用 ; 另一方面 , 这些结论又可能导致某些临床医师依据片面的信息做出错误的决策。系统评价 (systematic review, SR) 针对具体的临床问题 , 系统收集全球相关的临床研究文献 , 用统一的科学评价标准筛选出符合标准、质量好的文献 , 经统计学综合分析 , 获取定量结果并加以说明 , 最终得到可靠的结论 , 并强调根据新的临床研究结果对结论及时更新、补充。SR 采用的明确、系统的方法可以限制偏倚并减少偶然因素产生的影响。

由于一级文献数量非常多 , 而且每个研究的质量参差不齐 , 因而得出的结论可能不一致 , 容易使临床医生费时、费力查阅大量一级文献后 , 面对不同的研究结论十分困惑。这就需要高质量的 SR 为临床医生提供正确的实践指导。例如 , 针刺疗法在我国常被用于卒中康复治疗 , 但其有效性一直受到争议。Park 等对此进行了 SR 研究。他们搜索到 9 个有关针刺疗法治疗卒中的随机对照试验 (random control test, RCT) 研究 , 共包括 538 例病例 , 其中 6 个研究提示有效 , 而其他 3 个则与对照组无显著性差异 , 但通过 Meta 分析 , 只有 2 个研究 Jadad 指数超过 3 , 而方法学最好的几个研究均显示阴性结果 , 最后得出的结论是阴性的 , 但仍需要进一步大规模、高质量的 RCT 研究<sup>[1]</sup>。根据 RCT 或 SR 制定的临床诊治指南 (Guideline) 能更直接、全面地指导实践。例如 , 最新的欧洲脑卒中治疗指南就是完全根据循证医学的要求制定的。

系统评价不同于传统的叙述性文献综述。见表 1。

表 1 系统评价与传统叙述性文献综述的区别<sup>[2]</sup>

	文献综述	系统评价
针对问题	涉及面较广	只针对某一个要解决的问题
资料检索	无严格规定 , 往往不全面	全面收集 , 有明确的检索策略和要求
资料筛选	无严格规定 , 易受主观因素影响	有严格的方法学评价 , 较少混入人为因素
评价	无规定标准	有严格的评价指标
数据合成	通常为定性归纳	以定量的 Meta 分析为主
推论	可能是有根据的	通常是有根据的

作者单位 : 100068 北京市 , 北京博爱医院。作者简介 : 张小年 (1978-) , 男 , 浙江台州市人 , 医学博士 , 主要研究方向 : 神经康复及睡眠神经生理。

当然 , SR 也存在一定缺陷 , 如它是一种观察而不是具体临床研究 , 无法对条件进行控制 , 结论正确性受到原始研究质量的直接影响。此外 , 一些偏倚如发表偏倚 (publication bias) 等有时很难完全避免。尽管有缺陷 , SR 仍是重要的临床决策依据来源。

1 系统评价的来源

SR 的来源主要来自 Cochrane 协作网。此外 , 也有许多其他 SR 发表于众多期刊杂志。比较而言 , Cochrane 协作网的 SR 方法学质量较高 , 可信度也更大。Cochrane 的 SR 和研究方案以电子出版物的形式发表于 Cochrane 图书馆 , 同时 , Cochrane 协作网主张作者在杂志上以书面形式发表 , 并与国际著名医学杂志 , 包括柳叶刀 (Lancet)、英国医学杂志 (BMJ)、美国医学会杂志 (JAMA)、神经病学杂志 (Neurology)、脑卒中杂志 (Stroke) 等签订协议。因此 , 可以在医学搜索引擎如 Pub Med 中检索 Cochrane Systematic Review , 以“Cochrane”为检索词限制于 [title] 即可。另一方面 , Cochrane 图书馆也搜集了部分非 Cochrane 协作网的 SR , 保存于疗效评价文摘库中。

2 系统评价的步骤与方法

2.1 形成需要解决的临床问题 提出的问题应具有科学性 , 可通过研究结果回答。但更重要的是要有实用意义 , 应是临床决策真正需要的证据。问题的覆盖面过宽可能不利于最后形成可靠的结果 ; 反之 , 问题过窄则会影响评价结果的可推广应用性。形成问题时要对重要因素进行精确的描述 , 包括以下几点 :

2.1.1 研究对象 首先 , 明确要研究的疾病 , 如脑卒中的分型、分期 , 然后限定脑卒中患者的其他特征 , 如年龄、性别、存在的危险因素、合并症等。对人群特征的限定一定要有合理的生物学、社会学依据 , 具有一定的代表性。

2.1.2 干预类型或暴露因素 所比较的干预措施与试验的干预措施如有类似或重叠效应 , 则应当避免。辅助的干预措施也尽可能避免。

2.1.3 研究设计的类型 针对需回答的问题 , 选择最佳研究设计类型。如设计疗效评价问题 , 随机对照试验是最佳设计方案 ; 研究病因或危险因素则多采用病例对照研究或队列研究。研究设计中需要考虑的其他问题有 , 试验是否为安慰剂对照、评价结局的方法有无偏倚、是否有随访及随访的期限等。评价者对设计类型的限制越多 , 收集的资料就越局限。然而 , 选择的研究如果不能提供可靠资料 , 整个 SR 可能是无益的甚至会得出有误导性的结论。例如 , 要阐述某一治疗措施对慢性病患者生存率的影响 , 采用短期研究结果显然是不恰当的。

2.1.4 评价的结局 各项研究可能分别提供几种结局指标 ,

SR 要避免给读者过多结局指标,但不要漏报重要资料,尤其是终点结局。有的结局指标难以获得,但对决策者十分重要,也应当纳入。

最后,制定研究方案时提出的问题有可能在评价过程中被发现有不恰当之处,因此,进行适当的修改是必要的。但要注意,任何修改都应避免可能产生的偏倚,尤其要避免根据获得的资料提出问题。在修改问题之前,需要注意以下几个问题:修改问题的动机是什么?这种修改是否在了解某一研究的结果之后或受到某一研究结果的影响而作出?资料收集和检索策略是否恰当?

## 2.2 研究的检索与选择

2.2.1 检索工具 电子数据库,如 MEDLINE 是检索的主要工具。但如果检索仅限于 MEDLINE,则可能受发表偏倚(阳性结果的研究容易得到发表的倾向)和语言偏倚(英语为主,发展中国家的文献少)的影响。为了防止这种偏倚并尽可能收集相关研究,应使用多种检索工具系统地选择研究,最好使用多个电子数据库(如 CINAHL、EMBASE、AMED 等)。当然,不应忽略手工检索。此外,还可采用个人通信。这种非正式的交流形式有时是获取未发表资料的唯一途径。其他的所谓灰色文献,如技术报告、会议文章汇编、研究生论文等也可能含有一些并未全文发表的研究。最后,评价者应当对检索到的所有文章(包括以前发表的 SR)附录的参考文献进行筛查,以获得未能检索到的研究。

2.2.2 制定的检索策略 制定的检索策略应避免潜在的参考文献引用偏倚(作者优先选择支持自己观点的研究的倾向),一般要求一个以上评价者独立选择并鉴定文献,有时需要图书管理员参与制定检索策略,保证检索策略的全面、准确、具体。

2.2.3 选择研究 评估所有可能合格的研究报告是否满足 SR 的纳入标准。一般要求两人独立选择纳入的研究,出现不一致情况时,由第三者或双方讨论协商解决。有时,评价者需要对选择的研究作者、作者单位、发表出处进行设盲,交由他人评价。

2.3 评估每个研究的质量 研究的质量在这里主要指研究的设计和实施的防止系统误差或偏倚的程度。如果纳入低质量的研究,由于该研究有很大的偏倚就会导致 SR 的结论错误。评估前应制定明确的评估标准,以保证对所有研究的评估是一致的。制定评估标准主要根据以下几点:①是否随机分组,不仅要求分组方案随机化,更重要的是应采取恰当的措施隐藏随机方案;②受试对象、研究人员及其他研究参与者是否在研究过程中实行“盲法”;③患者的随访率是否理想,如何定义及处理失访;④各组患者的年龄、性别、职业等是否相似;⑤治疗效果的评价是否准确合理等。通常可以选择几篇文献进行初评,以检验制定的质量标准是否恰当,评价结果是否一致。

2.4 资料收集 评估者应设计一个适合本研究的数据收集表格,主要包括分组情况、每组样本数和研究效应的测量指标,此外,还有各项研究的真实性评估、研究的特征如方法、研究对象、干预措施、结局及不良反应等,尽量不遗漏有用的信息。有时,各研究间效应测量指标不一致,需要根据情况尽可能转化为统一指标或分别分析。如果从发表的研究报告中难以得到 SR 所需的全部资料,需要与研究报告者本人取得联系获取原

始资料。

Meta 分析将另文介绍。

2.5 结果讨论 在 SR 的讨论中,以下几方面有助于人们的决策:证据强度、结果可应用性、其他与决策有关的信息如费用问题和临床实践现状,以及干预措施的利弊、费用权衡等。评价的主要目的是提供信息,而不是提出建议或推荐性意见。因此,评价者应力求客观。讨论与结论应有助于人们正确理解证据的含义及其与实际决策的关系。此外,还应考虑所有重要的干预结局,包括不良反应的结局。最后,需要对 SR 发现的临床实践意义进行总结,对该评价结果对未来的科学研究具何种价值进行概括。当该评价结果不足以指导临床实践时,应阐明如何进行深入的临床试验研究。

2.6 偏倚控制 发表偏倚一直是 SR 中存在的问题之一,它是指阳性结果的研究容易得到发表的倾向,而对阴性结果的研究,作者一般不愿投稿或投稿后不易被发表,因而造成对某一干预措施的效果片面夸大<sup>[4]</sup>。所以,强调研究检索的范围要包括未发表文献的目的就在于避免发表偏倚。检查 SR 是否存在发表偏倚的方法之一是看其是否具有“倒漏斗”图形(funnel plot),即采用单个研究的效应估计值(X 轴)对应各个研究样本大小的量值(Y 轴)构成的散点图。对真实治疗效果估计值的精确性随着研究样本量的增加而增加。因此,小样本研究的效应值呈散在地分布于图形的底部,而大样本研究的分布随样本量增大而逐渐变窄,形状类似于一个倒置的漏斗。在无偏倚的情况下,该图形呈对称的倒漏斗状;如果存在偏倚,该图形会显示底部缺角。产生图形缺角的原因除发表偏倚外还有很多,如语言偏倚、引用偏倚等。研究方法学质量差异也可能导致图形不对称,因为小样本研究方法学质量一般较低,而方法学质量试验显示的治疗效应反而可能更大。因此,如果发现存在不对称,应对可能的原因进行探讨。

## 3 系统评价的可信度评估

对于临床医生,一般只要求能检索、阅读并利用已有的 SR,指导自己的临床实践。由于不同的 SR 存在质量差异,因此,进行 SR 可信度评估是很重要的。评估主要针对以下几个问题:①是否提出设计良好的临床问题?②是否明确前瞻性的入选和排除标准?③研究策略是否全面、明确?④是否对个体研究做了真实性评估?⑤研究的选择和资料收集过程是否可靠?⑥个体研究之间是否有足够的相似程度而被合并?⑦每个研究的关键信息是否在表格或图表中有明确显示?关键信息包括研究的设计和样本量、研究对象的选择和特征、失访数、干预措施的量和时间、评价指标和研究结果等;⑧个体研究之间的结果差异程度如何?⑨结论是否代表样本量最大、执行最严格的研究的结果?⑩结论与自己的临床实践经验是否一致?

## [参考文献]

- [1] Park J, Hopwood V, White AR, et al. Effectiveness of acupuncture for stroke: a systematic review[J]. J Neurol, 2001, 248(7): 558—563.
- [2] 李强. 循证医学——临床证据的产生、评价与利用[M]. 北京: 科学出版社, 2001. 32—33.
- [3] Meijer R, Ihnenfeldt DS, de Groot IJ, et al. A systematic review of the literature[J]. Clin Rehabil, 2003, 17(2): 119—129.

(收稿日期: 2004-10-31)