

如何做一个好的系统综述和 Meta 分析

詹思延[△]

(北京大学公共卫生学院流行病与卫生统计学系, 北京大学循证医学中心, 流行病学教育部重点实验室, 北京 100191)

[关键词] 系统综述; Meta 分析

[中图分类号] R181.2 [文献标识码] A [文章编号] 1671-167X(2010)06-0644-04

doi: 10.3969/j.issn.1671-167X.2010.06.008

随着循证医学的兴起, 如何系统地总结既往的研究成果, 为循证决策提供高质量的证据日益受到重视, 系统综述(systematic review, SR)和 Meta 分析(meta-analysis, MA)已被公认为客观评价和合成针对某一特定问题的研究证据的最佳手段, 通常被视为最高级别的证据^[1]。过去 20 年间这种合成证据的方法在医学研究领域得到了广泛的应用, 截至 2010 年 9 月 PubMed 已经收录 Meta 分析的论文 2 万 5 千余篇, Cochrane 系统综述也从 2006 年开始被 SCI 收录, 目前影响因子高达 5.653。国内发表的系统综述和 Meta 分析的数量从 2001 年以后呈现持续增长的趋势^[2], 但文章质量良莠不齐, 在研究设计、资料处理和论文写作等方面仍需要进一步完善和规范。

1 系统综述和 Meta 分析的概念

系统综述又叫系统评价, 属于二次研究, 是在复习、分析、整理和综合原始文献的基础上进行。一个系统综述研究可能只包括一种类型的研究, 也可以是不同研究方法的综合。Meta 分析是系统综述中使用的一种统计方法, 是以综合研究结果为目的, 通过查阅文献收集与某一特定问题相关的多个研究, 并对这些研究的结果进行统计分析。通常情况下, 针对同一研究目的可能有多篇研究报道, 单独任一研究都可能因为样本量太少或研究范围过于局限而很难得到一个明确的或具有一般性的结论, 将这些结果进行整合后所得到的综合结果(证据)无疑比任何一个单独的研究结果更有说服力, 因此, 系统综述和 Meta 分析是循证决策的良好依据^[3]。

Cochrane 系统综述是国际循证医学组织 Cochrane 协作网发布的系统综述, 该组织要求, 在开始系统综述前综述者首先要联系 Cochrane 50 余个工作组中相关的一个评价组, 以注册系统综述的题目; 进一步将研究方案提交该组评审、修改, 合格后将在 Cochrane 系统综述数据库中发表, 并接受来自用户的评论或批评; 系统综述全文在研究方案发表后的两年内完成, 否则, 该研究方案将从发表的 Cochrane 图书馆中撤除。发表后的系统综述也要定期更新。Cochrane 系统综述均采用统一的格式, 主要使用协作网免费提供的 Review Manager (RevMan) 软件进行, 在 Cochrane 图书馆 (Cochrane Library, 一本电子光盘杂志, 12 期/年) 上发表^[4]。

目前, 系统综述常和 Meta 分析共同或交叉使用, 如果精确区分的话, 当系统综述采用了定量合成的方法对资料进行统计学处理时可以称为 Meta 分析; 而未使用统计学方法的则称为定性的系统综述 (qualitative SR)。

2 正确认识和合理使用 Meta 分析方法

Meta 分析从本质上讲是定量化综述, 但与传统文献综述又有区别。传统文献综述往往是定性的, 且依赖于综述者的主观分析; 在复习文献时缺乏共同遵守的原则和步骤, 同类文献由不同的研究者进行综述, 结果可能大相径庭; 此外, 综述者常常注重研究结果统计学上是否“有意义”, 而统计学是否“有意义”取决于研究样本的大小, 许多小样本的研究可能得到的是假阴性的结果。Meta 分析则克服了传统文献综述的上述缺陷, 而具有如下功能: 定量综合; 对同一问题可提供系统的、可重复的、客观的综合方法; 通过对同一主题多个小样本研究结果的综合, 提高原结果的统计效能, 解决研究结果的不一致性, 改善效应估计值; 回答原各研究未提出的问题。

正是由于上述的特点, Meta 分析被视为循证医学的高级别证据, 但对此要正确认识。由于 Meta 分析是

对既往针对同一问题的研究结果的回顾,实际上是一种观察性研究^[5],它不仅不能排除原始研究中存在的偏倚,当原始研究质量不高时,合并的结果会遭受“垃圾进、垃圾出”的质疑;而且在文献查找、选择、资料提取和统计分析过程中,如果处理不当,还会引入新的偏倚,导致合并后的结果歪曲了真实的情况。

近些年来,Meta 分析在医学研究领域受到了日益广泛的重视,但作为系统综述中使用的一种统计学方法,能否采用取决于原始研究的结果是否可以合并。目前认为 Meta 分析主要适用于随机化对照试验(randomized controlled trial, RCT)结果的综合,尤其存在以下指征:(1)需要做一项紧急决定,而又缺乏时间进行一项新的试验;(2)目前没有能力开展大规模的临床试验;(3)有关药物和其他治疗,特别是副作用评价方法的研究;(4)研究结果矛盾时。即便对于 RCT,一般来说,各个试验在研究对象、干预措施和结局方面足够相似的情况下进行结果的合并才有意义;如果存在异质性,但合并资料仍然具有临床上的意义,则可采用随机效应模型;如果存在严重异质性,建议不要进行 Meta 分析。

观察性研究结果的 Meta 分析也开展了很多,几乎占已发表 Meta 分析的一半,主要是队列研究和病例对照研究在病因假设的检验或医学干预中的应用,但由于观察性研究很难证明排除了一切偏倚,也不可能完全去除混杂效应,如果研究过程中确实存在相同的系统误差,Meta 分析只会加大这些偏倚,产生统计学上的假象,因此观察性研究资料的 Meta 分析应慎重使用,对 Meta 分析的结果亦应采取科学的态度进行解释,重点应当放在检查研究结果异质性的可能来源上。

3 科学设计和严格实施系统综述及 Meta 分析

一个好的系统综述应该具备如下特征:清楚地表明题目和目的;采用综合检索策略;明确研究入选和排除标准;列出所有入选的研究;清楚地表达每个入选研究的特点并对其方法学质量进行分析;阐明所有排除研究的原因;如果可能,使用 Meta 分析合并合格的研究结果,并对合成的结果进行敏感性分析;要采用统一的格式报告研究结果。要达到这些标准,必须科学设计和严格实施系统综述及 Meta 分析,其中有几个环节尤其要注意。

3.1 研究的选题要有比较重要的临床意义,而且目前没有肯定一致的结论

研究的问题要宽窄适宜,研究目的简单明确。例如,“药物治疗能预防儿童偏头痛的发作吗?”这样的问题过大很难进行 Meta 分析,对此可以采用 PICO(patient, intervention, comparator, outcome)格式将研究问题结构化,即对研究对象的特征、采取什么干预措施、与什么进行比较、观察的结局指标明确定义进行结构化,从而精练研究目的,并提出一个明确的检验假设^[6]。

3.2 要多途径、多渠道、最大限度地收集相关文献

如利用多种电子资源数据库、临床试验注册登记系统、参考文献的追溯、手工检索等,还要注意那些未正式发表的所谓“灰色文献(grey literature)”,以减少发表偏倚的影响^[7]。

3.3 要根据研究目的确定文献的入选和排除标准

如疾病的诊断标准,研究对象的特征,暴露或干预的明确定义,是否排除伴发疾病;研究类型是仅限于随机化对照试验还是包括观察性研究;是选用已发表的文章还是包括未发表的研究;对文章的语种有无限制;观察时间和终点是什么。通常可以定义一个基本的入选和排除标准,待收集资料后进行彻底的敏感性分析,以估计不同的入选标准所得结果的稳定性(robustness)。

3.4 复习每个研究并进行质量评估

Meta 分析结果的真实性与原各研究的质量密切相关,只有从高质量的独立研究中才可能获得高质量的综合结论。一般来说,至少应从以下 3 方面来评估一个研究的质量:(1)方法学质量:研究设计和实施过程中避免或减小偏倚的程度;(2)精确度:即随机误差的程度,一般用可信限的宽度来表示;(3)外部真实性:研究结果外推的程度。目前已经发表了上百种质量评价工具来评价各种设计类型的文献,没有哪一个是金标准,研究者应该根据研究目的仔细选择恰当的评价工具^[8];从实践的角度出发,参考已发表的文献,选择本领域常用的评价工具也是可行的,例如,JADAD 评分就是文献中最常用到的评价 RCT 质量的简易工具^[9]。

3.5 重视异质性检验,不盲目追求统计学合并

异质性检验即统计量的齐性检验,是 Meta 分析的重要一环。一般认为当 $P > 0.10$ 时,各独立研究结果一致性较好,采用固定效应模型(fix effect model)进行分析。如果存在异质性,调查者对资料的汇总就要慎重,若合并资料仍然具有临床意义,可采用随机效应模型(random effect model)分析;但如果异质性严重,建

议不要进行 Meta 分析, 而应寻找异质性的来源, 异质性的来源一般存在三方面: (1) 临床异质性: 如对象特征、诊断、干预、对照、研究地点、评价结局等不同; (2) 方法学异质性: 研究设计与质量不同; (3) 统计学异质性: 不同试验中观察到的效应, 其变异性超过了机遇本身所致的变异性。因此, 当异质性严重时, 研究者可以根据异质性的来源进行亚组分析, 或进行敏感性分析, 或考虑协变量的影响进行 Meta 回归分析等。

3.6 尽可能做敏感性分析

敏感性分析是检查一定假设条件下所获结果的稳定性的方法, 其目的是发现影响 Meta 分析研究结果的主要因素, 解决不同研究结果的矛盾性, 发现产生不同结论的原因。敏感性分析最常用的方法是分层分析, 即按不同研究和发表特征, 如不同的研究设计类型、统计方法、研究的方法学质量高低、样本量大小、是否包括未发表的研究, 以及发表类型、发表语种等, 将各独立研究分为不同组后, 按 Mental-Haenszel 法进行合并分析, 再比较各组及其与合并效应间差异有无统计学意义。如果敏感性分析改变了结果, 下结论必须更加谨慎。

3.7 努力识别和减少证据合并过程中的偏倚

由于 Meta 分析是文献的二次研究, 在文献查找和选择过程中, 如果处理不当, 会引入偏倚, 如发表偏倚 (publication bias) 和定位偏倚 (location bias) 等, 导致合并后的结果歪曲了真实的情况。发表偏倚是指阳性的研究结果较阴性的研究结果被报告和发表的可能性更大。此外, 阳性结果的文章更容易以英文发表在国际性杂志, 被引用的次数可能更多, 重复发表的可能性更大, 从而带来文献定位中的偏倚。如果系统综述和 Meta 分析只是基于已经发表的研究结果和部分数据库的结果, 可能会夸大疗效, 甚至得到一个虚假的疗效。根据不同的入选标准进行彻底的敏感性分析是检查上述偏倚的最佳途径。此外, 还可以采用漏斗图分析 (funnel plots)、剪切-添补法和计算失效安全数 (fail-safe number, Nfs) 来评估发表偏倚对研究结果的影响, 但解决发表偏倚的根本途径是在医学伦理委员会或其他机构批准研究之际就将所有的 RCT 进行登记, 一些国际组织已经建立这类登记系统。中国循证医学中心也于 2006 年发起成立“中国临床试验注册和发表协作网 (Chinese Clinical Trial Registration and Publishing Collaboration, ChiCTRPC)”, 这种临床医学研究管理新模式的创建和应用, 将对提高中国临床试验信息透明度和质量、提高医学研究公信力发挥极其重要的作用。

3.8 采用标准规范的格式撰写总结报告

近年来多位学者指出, 国内发表的 Meta 分析文章的共同问题是报告不规范^[10-14], 导致无法判断是研究本身质量不高还是报告书写的问题。可以参考 2009 年国际上提出的系统综述和 Meta 分析的优先报告条目 PRISMA (preferred reporting items for systematic reviews and meta-analyses) 的要求进行总结报告^[15-16]。

3.9 重视 Meta 分析过程的质量控制

通常在文献检索、纳入/排除、质量评估和资料提取、数据录入等过程都要求双人独立进行, 如果出现分歧, 就需要讨论并由第三人 (一般为项目负责人) 裁定。

4 灵活选择不同类型的 Meta 分析方法

除了前述常规的系统综述和 Meta 分析之外, 近年来还出现了其他一些合成证据的方法^[4], 可以根据研究目的和潜在拥有的文献资料灵活选用。

4.1 累积 Meta 分析 (cumulative MA)

传统的 Meta 分析是对原始研究文献的一次性合并, 但累积 Meta 分析是按原始研究发表的时间顺序及时进行的重复 Meta 分析, 即每当一项新的研究得以鉴定, 则进行一次新的资料分析, 并按一定的顺序排列累积的结果, 用图表示, 从而反映研究结果的动态变化趋势, 而且可以评价各研究对综合结果的影响^[17]。这种方法的特殊功能在于, 当研究某一疗法有效或有害的趋势时, 可以指出在某一选定水准下, 疗效或安全性出现统计学显著性的最初时间, 为开展新的研究和制定相关政策提供方向和科学依据。例如, 利巴韦林联合干扰素与单用干扰素疗效比较的累积 Meta 分析发现, 虽然先后开展了 82 个 RCT, 但在 1998 年第 9 个试验纳入后, 对持续性病毒应答 (sustained virological response) 显示了有益的效果 ($RR: 0.74; 95\% CI: 0.64 \sim 0.85, P < 0.0001$), 之后的 73 个研究只是缩窄了可信区间并降低了 P 值; 但对临床结局是否有益尚无充分证据, 需要针对这些指标开展进一步的研究^[18]。再如, 环氧酶抑制药罗非昔布 (rofecoxib, Vioxx, 万洛) 因心血管不良作用在 2004 年 9 月被撤市, 但对 30 个 RCT 的累计 Meta 分析显示, 早在 2000 年 12 月就可以看出使用该药与心血管危险增加有关, 到 2001 年 6 月已经达到统计学显著性水平, 几乎可以提早 3 年半发现问题^[19]。

4.2 单个病例资料的 Meta 分析 (individual patient data, IPD MA)

传统的 Meta 分析主要基于文献中的总结性资料,是对文献的统计合成。虽然 Meta 分析者对其中的某些亚组特别感兴趣,但经常遇到的问题是很难在原文中找到相应的数据,因此从每个研究的设计组织者处获取补充资料的要求日益增多。一些国际性协作组织的成员已经开始分享各自的研究数据,从而使个体患者的资料得以充分利用,由此形成了单个病例资料的 Meta 分析,又叫“pooling 分析”,即对原始研究数据的合成分析。例如,阿司匹林在心血管疾病一级和二级预防中的应用:有关随机试验个体化资料的 Meta 分析^[20]就是对 6 个一级预防试验和 16 个二级预防试验中的个体资料进行汇总,对比长期使用阿司匹林与安慰剂对严重血管事件(心肌梗死、卒中或血管性死亡)的预防作用。这种分析方法可以充分利用原始数据进行生存分析,根据患者基线特征的不同开展亚组分析,还可以仔细分析并调整混杂因素的影响。

4.3 前瞻性 Meta 分析(prospective MA)

传统的 Meta 分析通常是回顾性的,因为纳入的研究已经完成,甚至结果已经报道,综述者对于这些研究结论的理解可能会影响他们对系统综述问题的确定,以及纳入排除标准、被评价的干预措施、对比组和研究结局的选择;而前瞻性 Meta 分析是在临床研究结果知道之前,即确定哪些研究可被纳入,这种方法的特色在于事先确定待检验的假设、前瞻性地确定筛选标准、事先确定统计学分析方法(包括亚组分析方法),因此可避免由资料决定的不可靠亚组分析方法,克服回顾性 Meta 分析某些缺点。在需要进行大样本研究以保证检验效能但现实又不可行的情况下,前瞻性 Meta 分析是一种非常有用的方法。研究者独自进行同类型研究,有一定的自主性,但又互相合作,以便合并试验结果,同时按计划进行前瞻性 Meta 分析。近年来在心血管疾病和儿童白血病研究领域对这种方法有所尝试^[21-22]。

总之,Meta 分析作为一种定量综合既往研究资料的新方法,为我们从整体角度把握事物的本质提供了一个有用的工具,但其设计、实施要科学和规范,避免误用和滥用。即使是最好的证据也要结合临床经验与患者个体情况进行恰当的应用;而对于目前质量欠佳的临床证据,应充分认识其局限性,辩证地对待,并有针对性地开展高质量的临床研究,不断完善和丰富证据资源,为循证临床决策奠定基础。

参考文献

- [1] 陈耀龙,李幼平,杜亮,等. 医学研究中证据分级和推荐强度的演进[J]. 中国循证医学杂志, 2008, 8(2): 127-133.
- [2] 卫茂玲,刘鸣,苏维,等. 中文发表系统评价、Meta 分析 18 年现状分析[J]. 华西医学, 2007, 22(4): 697-698.
- [3] 詹思延. 系统综述和 Meta 分析//李立明. 流行病学. 6 版. 北京: 人民卫生出版社, 2007: 180-195.
- [4] Higgins JPT, Green S. Cochrane handbook for systematic reviews of interventions version 5.0.2. The Cochrane collaboration, 2009[EB/OL]. (2009-09-20)[2010-08-30]. <http://www.cochrane-handbook.org>.
- [5] Egger M, Smith GD, Pjillips AN. Meta-analysis: principles and procedures[J]. BMJ, 1997, 315(7121): 1533-1537.
- [6] 詹思延. 循证医学与临床研究概述[J]. 中国循证儿科杂志, 2009, 4(2): 142-144.
- [7] Hopewell S, McDonald S, Clarke M, et al. Grey literature in meta-analyses of randomized trials of health care interventions[J/OL]. (2007-04-18)[2010-08-30]. <http://www.ncbi.nlm.nih.gov/pubmed/17443631>.
- [8] Katrak P, Bialocerkowski AE, Massy-Westropp N, et al. A systematic review of the content of critical appraisal tools[J]. BMC Med Res Methodol, 2004, 4: 22.
- [9] 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.
- [10] 叶蓁,翁源红,许东航. 2000~2002 年国内精神类药物荟萃分析的质量[J]. 药学服务与研究, 2004, 4(2): 121-123.
- [11] 刘建平,夏芸. 中文期刊发表的中医药系统综述或 Meta-分析文章的质量评价[J]. 中国中西医结合杂志, 2007, 21(4): 306-311.
- [12] 李廷谦,刘雪梅,张鸣明,等. 中文期刊发表的中医药系统评价/Meta 分析现状调查[J]. 中国循证医学杂志, 2007, 7(3): 180-188.
- [13] 仇瑶琴,贺佳. 国内护理期刊有关 Meta 分析论文的质量分析[J]. 解放军护理杂志, 2008, 25(23): 28-31.
- [14] 詹思延. 正确使用和合理解读 Meta 分析[J]. 中华儿科杂志, 2009, 47(12): 883-886.
- [15] Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses; the PRISMA statement[R/OL]. (2009-07-21)[2010-08-30]. <http://www.prisma-statement.org/usage.htm>.
- [16] Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration[J]. J Clin Epidemiol, 2009, 62(10): e1-34.
- [17] Lau J, Schmid CH, Chalmers TC. Cumulative meta-analysis of clinical trials builds evidence for exemplary medical care[J]. J Clin Epidemiol, 1995, 48(1): 45-57; discussion 59-60.
- [18] Brok J, Gluud LL, Gluud C. Meta-analysis: ribavirin plus interferon vs. interferon monotherapy for chronic hepatitis C—an updated Cochrane review[J]. Aliment Pharmacol Ther, 2010, 32(7): 840-850.
- [19] Ross JS, Madigan D, Hill KP, et al. Pooled analysis of rofecoxib placebo-controlled clinical trial data; lessons for postmarket pharmaceutical safety surveillance[J]. Arch Intern Med, 2009, 169(21): 1976-1985.
- [20] Baigent C, Blackwell L, Collins R, et al. Aspirin in the primary and secondary prevention of vascular disease: collaborative meta-analysis of individual participant data from randomised trials[J]. Lancet, 2009, 373(9678): 1849-1860.
- [21] Simes RJ. Prospective meta-analysis of cholesterol-lowering studies: the Prospective Pravastatin Pooling (PPP) Project and the Cholesterol Treatment Trialists (CTT) Collaboration[J]. Am J Cardiol, 1995, 76(9): 122C-126C.
- [22] Valsecchi MG, Masera G. A new challenge in clinical research in childhood ALL: the prospective meta-analysis strategy for intergroup collaboration[J]. Ann Oncol, 1996, 7(10): 1005-1008.