

· 思路与方法学 ·

临床研究中结论偏倚及其评价的研究现状

冯 硕 孙 瑾 吴若菡 韩 梅 刘建平

摘要 临床研究中,从获得研究结果到形成研究结论的过程中,可能出现结论偏倚。本文对结论偏倚的定义、产生原因、影响及分类进行概括,并介绍结论偏倚评价工具和已有的结论偏倚评价研究。

关键词 临床研究; 结论偏倚; 评价

Conclusion Bias and Corresponding Evaluation Studies for Clinical Researches FENG Shuo, SUN Jin, WU Ruo-han, HAN Mei, and LIU Jian-ping Center for Evidence-based Chinese Medicine, Beijing University of Chinese Medicine, Beijing (100029)

ABSTRACT Bias in conclusion may exist in clinical researches, in the process from the data obtained to drawing conclusion. This study generally introduced the concept of conclusion bias, as well as its cause, influence and classification. The evaluation tools for conclusion bias and the corresponding studies about it were also introduced.

KEYWORDS clinical researches; conclusion bias; evaluation

科研论文的结论是文章的精髓部分。干预性临床研究中,结论应包括以下内容:研究结果说明的临床问题,干预措施的疗效和安全性是否实现了原来的假设;结论有无研究数据支撑;干预效应对临床实践有何意义,如是否推荐在临床实践中使用;以及研究的局限性和对未来的展望。然而,从获得结果到形成结论的过程是人为而主观的,因此这一过程也会产生偏倚。在国外的方法学研究中,将其称为“spin”(结论歪曲)、“distortion”(结论曲解)、“bias in conclusion”(结论偏倚)、“interpretive bias”(解释偏倚)等,可意译为结论中含有对结果的歪曲报告、解读或偏离,在此称之为“结论偏倚”,以下将对国内外临床研究结论偏倚的现状进行介绍。

1 结论偏倚的定义

结论偏倚属于非技术原因造成的主观偏倚^[1],指作者在报告研究结论时有意识地或无意识地使用特殊的手段,用来说服读者,让读者信服干预措施是有效而安全的,因而夸大了干预措施实际的疗效和安全性^[2]。

早先提出结论偏倚的是一些国外的方法学专家和杂志社编辑,源于其评阅稿件的过程。Lazarus C

等^[3]发现 BMC 系列杂志两年半内收到的稿件中,超过半数存在结论偏倚。科学需要严谨的态度,尤其是对新技术和疗法,但难免有来自研究者的主观判断。然而,研究数据并不会为自己代言^[4]。从得出数据到做出其效果的结论判断,这一主观过程中产生的偏倚常常被忽略。

2 结论偏倚产生的原因

(1)研究者对于干预措施抱有乐观和期待的心理,因此下结论时容易主观臆断。尽管医务工作者应保持乐观心态,但科研是不能带有色眼镜的^[5]。有学者建议,不应当将阴性结果称之为“这项研究结果为阴性”,而应该是“研究结果支持这个无效的假设”^[5]。(2)研究本身可能存在方法缺陷,例如未预先规定假设检验的类型,未计算样本量,或未选择恰当的统计学方法^[6,7]。这些均造成了偏倚,因此也不可能得出真实而客观的结论。(3)研究者执迷于一些传统观点或看法,尽管有科学的证据来反驳,但仍对自己所持的观念深信不疑。比如“不吃早饭容易增重”这是个伪命题。有研究者检索了与此相关的观察性研究^[8],而后通过累积性 Meta 分析来证实,其并不具备因果关系。评价原始研究的结论后发现,在结论报告中存在因果关系语言表述不恰当,以及对结果的解释带有主观倾向。(4)研究可能会涉及利益因素。有学者评价结直肠腹腔镜对比传统开腹手术的研究^[9],发现 80% 以上的研究存在结论偏倚,如样

作者单位:北京中医药大学循证医学中心(北京 100700)

通讯作者:刘建平, Tel: 010-64286757, E-mail: jianping_l@hotmail.com

DOI: 10.7661/j.cjim.20180630.243

本量较小、随访期间较短的情况下就声称腹腔镜是安全的,并利用统计分析探讨了结论偏倚的相关因素,发现这与文章是否有利益声明相关。有研究者对比了同一领域有企业资助和非企业资助的临床研究,并进行了 Meta 分析,发现两者有相近的结果却有不同结论^[10],前者带有倾向性,认定试验药物优于对照药物,并推荐使用试验药物。

3 结论偏倚的影响

Boutron I 等^[11]做了一项方法学研究,以 300 位临床医生为研究对象,将其随机分为两组:一组给他们读带有结论偏倚的文章,但原文中结果为阴性;另一组让他们读同一篇修正了结论的文章。比较两组中临床医生对文章中干预措施的印象,主要是指治疗利弊和使用推荐。结果发现第一组医生更容易相信干预措施是有效的。这说明,结论中对结果作出歪曲、过度或不充分地解读,可误导读者对干预措施的态度。有研究者指出,部分摘要中仅包含有统计学意义的结果和结论,不含无统计学意义以及安全性相关的内容,要获得对评价措施的全面了解,建议阅读全文以合理指导临床决策^[12]。

4 结论偏倚的分类

Amos AJ^[13]将结论偏倚归为六类:(1)未对照就下优效性结论;(2)发现主要结局为阴性后更改主要结局,利用更改后的结局下结论;(3)当主要结局为阴性时,过多关注次要结局;(4)当结果为阴性时,使用亚组分析中的阳性结果下结论;(5)根据预先制定的统计分析计划得到阴性结果后,更改统计分析方法

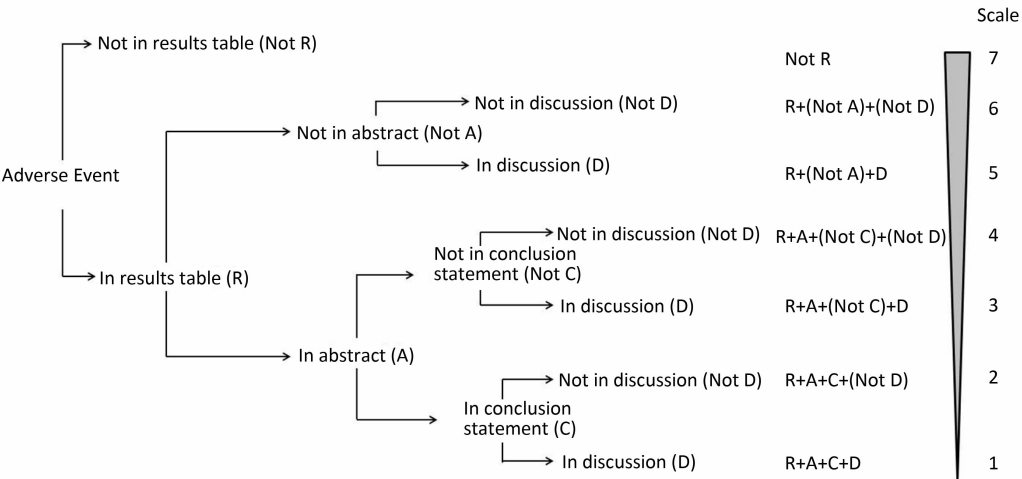
后下结论;(6)将 P 值接近 0.05 的结果(如 $P = 0.051$)报告为有显著差异。

此外,研究者在结论报告中会使用一些夸张的修辞手法或写作技巧^[14,15],也应属于结论偏倚的范围。英文文章强调对不同类型的临床研究使用不同强度的语言描述。例如在临床试验中可以说“干预措施能够降低某种风险”,而在同一领域的观察性研究中,应该改写为“研究发现使用干预措施后某项风险降低了”,或“某种风险降低可能与干预措施相关”。在中文的临床研究中,常见有“疗效尤其显著”,“值得广泛使用”这样的修饰语。

5 结论评价工具

Boutron I 等^[16]在 JAMA 上发表的研究中制定了一款结论偏倚评价工具,被之后的研究者多次借鉴使用,其将结论偏倚归为三类:(1)结论中只关注有统计学意义的结果(例如组内比较、次要结局、亚组分析、按受试者特征调整后的分析等);(2)结论中将无统计学意义的结果解释为疗效或安全性相等;(3)结果无统计学意义,结论中仍强调干预措施会为受试者带来获益。

此后,Vera-Badillo FE 等^[17]针对药物安全性,制定了一项不良事件结论偏倚刻度尺。根据文中对不良事件的记载是否出现在结果、讨论、结论、摘要中,将其评为 1~7 分。分值越高,结论偏倚的程度越高。如四部分均未提及不良事件,评为 7 分;仅在结果部分提及,评为 6 分;仅结果和讨论部分有提及,评为 5 分,依次类推,四部分中都有安全性的描述则评为 1 分(见图 1)。



注:Not R:结果部分无关于是否发生不良事件的报告;Not D:不良事件未在讨论部分提及;Not A:不良事件未在摘要中提及;Not C:不良事件未在结论部分提及;R:结果部分无关于是否发生不良事件的报告;D:讨论部分有提及不良事件;A:摘要部分有提及不良事件;C:结论部分有提及不良事件

图 1 Vera-Badillo FE 等^[17]研究中所使用的不良事件相关结论偏倚刻度尺

另外, Lazarus C 等^[18]制定了一款针对非随机对照试验的结论偏倚评价工具。评价内容为: (1) 误导性报告, 是对结果的选择性分析和解读(包括: ①不报告干预措施的安全性; ②对有效性只报告阳性结果, 对安全性只报告阴性结果; ③错误地报告研究设计类型; ④使用修辞技巧; ⑤不考虑研究的局限性; ⑥选择性引用文献)。 (2) 误导性分析, 指错误地分析研究结果或夸大干预措施的疗效(包括: ①对无显著性差异的结果做出了有效的结论; ②对无显著性差异的结果, 不考虑可信区间和精确性, 就做出了疗效相等的结论; ③对无显著性差异的安全性结果, 声称其是安全的, 而未考虑检验效能; ④无对照就做出优效性的结论; ⑤下有效性结论时只考虑统计学差异, 而未考虑临床意义; ⑥不恰当的因果关系推断; ⑦未进行统计学检验就下结论)。 (3) 结果外推不充分(包括: ①不恰当地外推至其他疾病、其他人群; ②不恰当地在临床应用中推荐使用)。此外, Lazarus C 等对结论偏倚的程度进行了划分。偏倚程度高意为作者在结论中没有表达“不确定”的意思, 如未使用 may、probably 一类的情态动词, 没有对今后研究的建议; 偏倚程度中等意为作者表达了一些“不确定”的意思, 并推荐今后开展进一步研究; 偏倚程度低意为作者对结论进行了不确定性的表达, 推荐今后开展类似的研究。

6 国外发表的结论偏倚评价研究

O'Connell NE 等^[19]分析了一项德国开展的针灸临床试验, 发现在报告结论时, 由于过分重视方法的科学性, 在解读数据时产生了偏差, 即结论中降低了研究结果的有效性。Derry CJ 等^[20]发现, 这类偏倚不仅存在于单项针刺临床试验中, 还在系统综述中。有系统综述在下结论时忽略了结果中大规模临床试验对结果的影响。也有系统综述只依据结果认为针刺具有效果(effectiveness), 而忽略了是否具备效力(efficacy)。

Amos AJ^[13]对“精神病是否应该早期干预”这一话题提出了质疑, 由此引发了对文献中“是否存在结论偏倚”的思考和研究。类似研究中的结论偏倚, 最常用的手法就是在主要结果为阴性的情况下使用次要结局和亚组分析结果来下结论。

Vera-Badillo FE 等^[17]评价了乳腺癌Ⅲ期临床试验, 发现有 67% 的研究存在安全性相关的结论偏倚, 大多数在研究结论中对不良反应描述不够充分, 同时也强调了在摘要中避免结论偏倚的重要性。

Ioannidis JP^[21]提出, 对不良反应的歪曲报告是更严重的一个情况。如某些研究在未观察到不良事件

时就声称“无不良反应”, 这些不恰当的报告方式, 可能会受到投资方市场运营和各方利益冲突的影响。随机对照试验中未观察到的不良反应有赖于非随机对照试验和个案报道去弥补, 而不能轻易地下安全的结论。

Ochodo EA 等^[22]评价了 2010—2011 年 PubMed 上发表的影响因子 4 分以上的诊断试验研究。其中 31% 在摘要中存在明显的结论偏倚, 包括下结论时过于乐观, 语气过于肯定, 选择性下结论, 利用亚组分析结果下结论, 研究目的与研究结论不一致等。而 90% 的诊断试验存在不易察觉的结论偏倚, 这与统计分析方法、样本量计算等相关。Lumbreras B 等^[23]评价了 2006 年发表的分子诊断研究, 发现有 56% 在结论部分夸大了分子诊断的临床应用, 且普遍夸大了这一诊断措施的临床意义。

Ioannidis JP^[24]检索了 2005 年高影响因子期刊中 400 篇被高频引用的文章, 只有 17% 在全文中谈论到了研究的局限性; 仅 1% 在摘要中谈到了研究的局限性。尽管 61% 的文章谈到了设计缺陷、偏倚、真实性、可重复性等问题, 但许多并未说明研究局限性对结果造成的影响。

Allan GM 等^[25]建议, 在下研究结论时应当同时考虑点估计和区间估计, 因为 P 值并不能反映临床意义, 应当同时使用 RR 值及其 95% 可信区间。Glujovsky D 等^[26]检索了 2014 年 3 种高影响因子的生殖期刊上发表的随机对照试验, 32 篇文章中没有一项使用可信区间下结论。国外已有学者提出, 在报告效应量时, 绝对效应和相对效应都要报告。国内也有学者指出, “ $P < 0.05$ ”不应当解释为“关系密切”或“显著相关”, 而应是“差异(或差别)有统计学意义”^[27]。

7 小结

结论偏倚评价是方法学研究中的一项重要内容。结论的客观公正影响了对治疗方法或诊断技术的认识和采纳。国外已发表的方法学研究着眼于制定评价工具来评价结论偏倚, 或针对特殊的临床问题探索是否有结论偏倚。鉴于这项工作能够指导文献阅读, 帮助临床医生制定医疗决策, 建议在国内的中医药领域开展结论偏倚评价研究。通过开展类似的研究, 制定相关的评价工具并评价中药临床研究中结论偏倚的出现情况, 如此可以警惕作者为追求阳性结论而带来的主观偏见。同时, 倡导客观地报道结论, 才能有益于医学科研的发展和进步。

参 考 文 献

[1] Kaptchuk TJ. Effect of interpretive bias on re-

- search evidence [J]. *BMJ*, 2003, 326 (7404): 1453 – 1455.
- [2] Fletcher RH, Black B. Spin in scientific writing: scientific mischief and legal jeopardy [J]. *Med Law*, 2007, 26(3): 511 – 525.
 - [3] Lazarus C, Haneef R, Ravaud P, et al. Peer reviewers identified spin in manuscripts of nonrandomized studies assessing therapeutic interventions, but their impact on spin in abstract conclusions was limited[J]. *J Clin Epidemiol*, 2016, 77: 44 – 51.
 - [4] Elstein AS. Clinical judgment: a critical appraisal [M]. Dordrecht: Reidel, 1979: 17 – 28.
 - [5] Sullivan GM. Is there a role for spin doctors in medical research[J]. *J Grad Med Educ*, 2014, 6 (3): 405 – 407.
 - [6] Marco CA, Larkin GL. Research ethics: ethical issues of data reporting and the quest for authenticity[J]. *Acad Emerg Med*, 2000, 7(6): 691 – 694.
 - [7] Scott IA, Greenberg PB, Poole PJ. Cautionary tales in the clinical interpretation of studies of diagnostic tests [J]. *Intern Med J*, 2008, 38 (2): 120 – 129.
 - [8] Brown AW, Bohan Brown MM, Allison DB. Belief beyond the evidence: using the proposed effect of breakfast on obesity to show 2 practices that distort scientific evidence [J]. *Am J Clin Nutr*, 2013, 98(5): 1298 – 1308.
 - [9] Patel SV, Van Koughnett JA, Howe B, et al. Spin is common in studies assessing robotic colorectal surgery: an assessment of reporting and interpretation of study results [J]. *Dis Colon Rectum*, 2015, 58(9): 878 – 884.
 - [10] Jørgensen AW, Hilden J, Gøtzsche PC. Cochrane reviews compared with industry supported meta-analyses and other meta-analyses of the same drugs: systematic review [J]. *BMJ*, 2006, 333 (7572): 782.
 - [11] Boutron I, Altman DG, Hopewell S, et al. Impact of spin in the abstracts of articles reporting results of randomized controlled trials in the field of cancer: the SPIIN randomized controlled trial [J]. *J Clin Oncol*, 2014, 32(36): 4120 – 4126.
 - [12] Dal-Ré R, Castell MV, García-Puig J. If the results of an article are noteworthy, read the entire article; do not rely on the abstract alone [J]. *Rev Clin Esp*, 2015, 215(8): 454 – 457.
 - [13] Amos AJ. A review of spin and bias use in the early intervention in psychosis literature [J]. *Prim Care Companion CNS Disord*, 2014, 16(1), doi: 10.4088/PCC.13r01586.
 - [14] Kohli P, Cannon CP. The importance of matching language to type of evidence: avoiding the pitfalls of reporting outcomes data [J]. *Clin Cardiol*, 2012, 35(12): 714 – 717.
 - [15] Horton R. The rhetoric of research [J]. *BMJ*, 1995, 310(6985): 985 – 987.
 - [16] Boutron I, Dutton S, Ravaud P, et al. Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes [J]. *JAMA*, 2010, 303(20): 2058 – 2064.
 - [17] Vera-Badillo FE, Shapiro R, Ocana A, et al. Bias in reporting of end points of efficacy and toxicity in randomized, clinical trials for women with breast cancer [J]. *Ann Oncol*, 2013, 24 (5): 1238 – 1244.
 - [18] Lazarus C, Haneef R, Ravaud P, et al. Classification and prevalence of spin in abstracts of nonrandomized studies evaluating an intervention [J]. *BMC Med Res Methodol*, 2015, 13(15): 85.
 - [19] O'Connell NE, Wand BM, Goldacre B. Interpretive bias in acupuncture research: a case study [J]. *Eval Health Prof*, 2009, 32(4): 393 – 409.
 - [20] Derry CJ, Derry S, McQuay HJ, et al. Systematic review of systematic reviews of acupuncture published 1996 – 2005 [J]. *Clin Med (Lond)*, 2006, 6(4): 381 – 386.
 - [21] Ioannidis JP. Adverse events in randomized trials: neglected, restricted, distorted, and silenced [J]. *Arch Intern Med*, 2009, 169 (19): 1737 – 1739.
 - [22] Ochodo EA, de Haan MC, Reitsma JB, et al. Overinterpretation and misreporting of diagnostic accuracy studies: evidence of "spin" [J]. *Radiology*, 2013, 267(2): 581 – 588.
 - [23] Lumbreras B, Parker LA, Porta M, et al. Overinterpretation of clinical applicability in molecular diagnostic research [J]. *Clin Chem*, 2009, 55 (4): 786 – 794.
 - [24] Ioannidis JP. Limitations are not properly acknowledged in the scientific literature [J]. *J Clin Epidemiol*, 2007, 60(4): 324 – 329.
 - [25] Allan GM, Finley CR, McCormack J, et al. Are potentially clinically meaningful benefits misinterpreted in cardiovascular randomized trials? A systematic examination of statistical significance, clinical significance, and authors' conclusions [J]. *BMC Med*, 2017, 15(1): 58.
 - [26] Glujovsky D, Sueldo CE, Borghi C, et al. Misleading reporting and interpretation of results in major infertility journals [J]. *Fertil Steril*, 2016, 105(5): 1301 – 1306.
 - [27] 徐波, 沈叔洪. 统计 *P* 值的意义及常见统计学表述错误辨析 [J]. *浙江医学*, 2011, 33(8): 1252 – 1253.
- (收稿: 2017-04-19 在线: 2018-08-29)
责任编辑: 汤 静