

ROBIS : 评估系统评价偏倚风险的新工具

吴琼芳^{1#}, 丁泓帆^{2#}, 邓围², 杨楠³, 王琪^{4,5,6}, 姚亮^{4,5,6}, 王小琴^{4,5,6}, 韦当^{4,5,6}, 陈耀龙^{4,5,6*}, 杜亮^{7*}

1. 北京大学第一医院麻醉科(北京 100034); 2. 兰州大学第二临床医学院(兰州 730000);
3. 兰州大学第一临床医学院(兰州 730000); 4. 兰州大学基础医学院循证医学中心(兰州 730000);
5. 甘肃省循证医学与临床转化重点实验室(兰州 730000); 6. GRADE 中国中心(兰州 730000);
7. 四川大学华西医院中国循证医学中心(成都 610041)

摘要 目前尚无专门针对系统评价设计、制作和分析过程中的偏倚风险进行评估的工具,而最新研发的 ROBIS 工具主要评估系统评价的偏倚风险,不仅用于评估包括干预性、诊断性、病因性、预后性等多种系统评价制作过程和结果解释过程中的偏倚风险,还用于评价系统评价问题与其使用者要解决的实践问题的相关性。本文旨在介绍 ROBIS 工具,为国内系统评价制作者、指南制定者和其他相关研究者提供理解和应用该工具的参考,从而提高制作系统评价的质量,促进其应用。

关键词 系统评价;偏倚风险;ROBIS

ROBIS: A New Tool to Assess Risk of Bias in Systematic Reviews

WU Qiong-fang^{1#}, DING Hong-fan^{2#}, DENG Wei², YANG Nan³, WANG Qi^{4,5,6}, YAO Liang^{4,5,6},
WANG Xiao-qin^{4,5,6}, WEI Dang^{4,5,6}, CHEN Yao-long^{4,5,6*}, DU Liang^{7*}

1. Department of Anesthesiology, Peking University First Hospital, Beijing 100034 China;
2. The Second Clinical Medical School, Lanzhou University, Lanzhou 730000, China;
3. The First Clinical Medical School, Lanzhou University, Lanzhou 730000, China;
4. Evidence-Based Medicine Center, School of Basic Medical Sciences, Lanzhou University, Lanzhou 730000, China;
5. Key Laboratory of Evidence Based Medicine and Knowledge Translation of Gansu Province, Lanzhou 730000, China;
6. Chinese GRADE Center, Lanzhou University, Lanzhou 730000, China;
7. Chinese Evidence-Based Medicine Center, West China Hospital, Sichuan University, Chengdu 610041, China

Abstract Currently there is no tool designed specifically to assess the risk of bias in the design, conduct or analysis of systematic reviews. ROBIS (Risk Of Bias In Systematic reviews), which was developed lately, aims mainly to assess the risk of bias in the conduct and result interpretation of systematic reviews relating to interventions, etiology, diagnosis and prognosis, as well as the relevance of the systematic review questions and the practice questions that their users want to address. This paper aims to introduce the ROBIS tool to Chinese systematic review developers, guideline developers and other researchers to promote the comprehension of it and its application, so as to improve the quality of systematic reviews in China.

Key words Systematic review; Risk of bias; ROBIS

1 背景

系统评价通过减少偏倚的策略以合成多个原始研究的结果^[1,2],是公认的临床决策中最佳的证据综合研究^[3]。但随着系统评价数量的快速增长,其参差不齐的质量引起越来越多国内外学者的关注和担忧^[4,5]。只有高质量的系统评价才能为临床医师、患者及其他利益相关者提供决策依据^[6]。系统评价的

质量包括方法学质量和报告质量。方法学质量存在问题可能会影响系统评价结果真实性,使针对同一问题的系统评价所纳入的原始研究和最终结论出现差异^[7]。目前已有许多工具评估系统评价方法学质量,其中广泛使用的是1991年制定的OQAQ(the Overview Quality Assessment Questionnaire)工具^[8,9]和2007年发布的AMSTAR(the Assessment of Multiple Systematic Reviews)工具^[10,11]。AMSTAR的制定是基于OQAQ清单^[12]。由于AMSTAR不能产生量化的评价结果,因此有研究人员在2010年制定发布了AMSTAR修订版(R-AMSTAR)^[13]。AMSTAR虽被认为是当前有效和实用的系统评价方法学质量评价工具之一,但目前主要用于评价纳入随机对

DOI: 10.7507/1672-2531.20150236

共同第一作者:吴琼芳,女(1991年~),硕士研究生,医师,以循证麻醉学为主要研究方向。Email: wuqf91@163.com
丁泓帆,男(1993年~),本科在读,临床医学专业。Email: dinghf14@163.com

* 共同通讯作者:陈耀龙,Email: chenyaolong21@163.com;
杜亮,Email: duliang0606@vip.sina.com

照试验(randomized controlled trial, RCT)的系统评价^[14],虽然有研究表明,其用于评价非 RCT 系统评价也具有较高的信度和实用性^[15],但在非 RCT 系统评价中的应用仍有限^[14],而 R-AMSTAR 的有效性仍有待进一步研究^[14]。2014 年英国布里斯托尔大学(University of Bristol)社会医学部制定了一种全新的评价工具——ROBIS (Risk of Bias in Systematic Review)工具^[16],其针对系统评价的偏倚风险,不仅用于评估包括干预性、诊断性、病因性、预后性等多种系统评价制作过程和结果解释过程中的偏倚风险,还用于评价系统评价问题与其使用者要解决的实践问题的相关性。本文旨在介绍 ROBIS 工具,为国内系统评价制作者、指南制定者和其他相关研究者提供理解和应用该工具的参考,从而提高系统评价的质量,促进其应用。

2 系统评价偏倚风险

如果系统评价在设计、制作和分析过程中出现缺陷或者局限性,严重影响了系统评价的结果,那么就会产生偏倚,这被称为系统评价的偏倚风险,其不同于“系统评价方法学质量”这一目前虽广泛使用但笼统抽象的概念。系统评价的偏倚风险需要与系统评价纳入原始研究的偏倚风险^[17]区别开来。原始研究的偏倚风险指单个研究在设计、实施和分析过程中产生的偏倚,即使原始研究偏倚风险都很高,只要系统评价在制作过程中充分考虑和处理了这些偏倚,那么系统评价的偏倚风险仍可为“低”。Whiting 等^[16]的研究发现,当前尚无专门针对系统评价设计、制作和分析过程中的偏倚风险进行评估的工具,研究人员通过严格的方法和流程,研发了 ROBIS 工具,填补了这一空缺。ROBIS 工具可用于评估多种类型的系统评价,不仅能评估其偏倚风险,还能评估系统评价的问题与其使用者关心的实践问题的相关性。

3 ROBIS 工具的制定过程

ROBIS 工具的制定分为四个阶段^[18],其整个过程科学、严谨和透明。

(1)确定范围:成立专家指导小组,通过视频会议制定 ROBIS 范围,之后举行面对面会议,通过德尔菲法(Delphi)^[19]最终确定 ROBIS 工具的要点、结构和性能。

(2)获取证据支持:制定者通过 3 种方法来获取证据以支持 ROBIS 制定。包括:参考和分析 Cochrane 系统评价方法学期望值(Methodological Expectations of Cochrane Intervention Reviews,

MECIR)条目^[20];评价 40 个现有的系统评价或 Meta 分析质量评价工具并进行条目分类;对使用 AMSTAR 工具的系统评价进行评价,以获取 ROBIS 工具使用者的具体要求。

(3)进行共识:所有利益相关者召开面对面会议,讨论并形成 ROBIS 工具的草案,之后进行小范围讨论再次确定。

(4)预试验和完善:通过三轮改良德尔菲法再次确定 ROBIS 工具的范围和内容。并举办 ROBIS 培训班,邀请参会者进行预试验,根据其使用 ROBIS 后的反馈,完善该工具。

4 ROBIS 工具的使用人群

ROBIS 工具的主要使用人群为:系统评价再评价的作者;指南制定者;系统评价作者:可在系统评价完成后评价其质量,或在系统评价研究设计阶段参考该工具以减少偏倚。其他可能的使用者包括决策支持机构(如英国国家卫生与临床优化研究所,NICE)、对循证医学感兴趣的临床医生、杂志编辑和评审人员等。

5 ROBIS 工具条目

应用 ROBIS 评估系统评价偏倚风险的过程包括三个阶段:评估相关性(根据情况选择);确定系统评价制定过程中的偏倚风险程度;判断系统评价的偏倚风险。ROBIS 工具清单详见表 1 和表 2。使用者可以从其网站(<http://www.robis-tool.info/>)获得 ROBIS 工具清单和使用指导。ROBIS 条目具体情况可参见《评估系统评价偏倚风险的 ROBIS 工具实例解读》一文。

5.1 阶段一:评估相关性(根据情况选择)

ROBIS 用户首先需要确定想要解决的问题(例如,系统评价再评价或指南的问题),即目标问题(target question)。第一阶段旨在评估目标问题与系统评价中拟解决的问题的吻合度。如果仅评估系统评价的偏倚风险,而没有目标问题,那么可跳过该阶段(即根据情况选择)。由于系统评价分为干预性、病因性、诊断性和预后性等不同类别,其对应的评估表也不同(见表 1)。

5.2 阶段二:确定系统评价制作过程中偏倚风险的程度

此阶段主要确定系统评价各制作过程中可能产生的偏倚,涉及制作系统评价的 4 个关键领域:研究的纳入标准、研究的检索和筛选、数据提取和质量评价、数据合成和结果呈现(见表 2)。领域 1 旨在评估

系统评价的纳入标准是否预先确定,是否清晰且适合于系统评价的问题。领域2旨在评估系统评价是否遗漏了满足纳入标准的原始研究,以及是否纳入了不符合纳入标准的研究。领域3旨在评估数据提取和原始研究质量评价过程是否产生了偏倚。领域4旨在评估系统评价制作者是否使用了合适的方法来合并原始研究的数据。每一领域都设置标志性问题以识别各领域引入的偏倚。每一领域的评估包括三个步骤:从系统评价中寻找支持偏倚风险程度判断的信息,回答标志性问题,判断偏倚风险程度。

5.3 阶段三:判断偏倚风险

第三阶段是判断系统评价整体的偏倚风险,主要识别结果解释部分引入的偏倚,旨在评估系统评价在结果解释中是否考虑或处理了阶段二中的偏倚风险,是否考虑到了纳入研究与系统评价研究问题的相关性,是否避免强调有统计学意义的结果(见表2)。同样涉及三个步骤:从系统评价中寻找支持

偏倚风险程度判断的信息,回答标志性问题,判断该系统评价的偏倚风险程度。

6 小结

调查显示,2010年系统评价或Meta分析发表量约4000篇,平均每天11篇^[21],如此庞大的数量并不代表每篇系统评价的质量相同。系统评价再评价者和指南制定者及其他研究人员在应用系统评价时需要一个当前最好的衡量其偏倚风险的工具。目前有40个质量评价工具^[16],其中AMSTAR工具应用广泛且被认为具有较好的信度、结构效度和实用性^[11],其制定者认为该工具可用于评价各种研究设计的系统评价,但实际上目前主要用于评价纳入RCT的系统评价,在非RCT系统评价中的应用仍有限^[14],其使用指导中也并没有对其他研究类型的系统评价的应用进行过多说明^[22]。虽然已有研究表明AMSTAR用于评价非RCT系统评价也具有较高的

表1 不同类型系统评价的评估表(阶段一)

系统评价类型				目标问题	系统评价问题
干预性	病因性	诊断准确性试验	预后性		
患者或人群	患者或人群	患者	患者		
干预措施	暴露因素和对照因素	待评价试验	要预测的结局		
对照措施		金标准试验	计划使用的模型		
结局指标	结局指标	目标疾病	计划的时间点		
系统评价要解决的问题与目标问题匹配吗?				是/否/部分/不确定	

表2 评估领域及标志性问题(阶段二、三)

领域1:研究的纳入排除标准	阶段二			阶段三	
	领域2:研究的检索和筛选	领域3:数据提取和质量评价	领域4:数据合成和结果呈现	系统评价的偏倚风险	
标志性问题*	1.1 系统评价遵循了预先确定的目的和纳入标准吗? 1.2 纳入标准适合系统评价的问题吗? 1.3 纳入标准明确吗? 1.4 纳入标准中所有基于研究特征的限制合适吗? 1.5 纳入标准中所有与研究来源相关的限制合适吗?	2.1 检索已发表和未发表的研究时所包含的数据库或电子资源的范围合适吗? 2.2 使用了除数据库检索以外的其他方法来确定相关研究吗? 2.3 检索策略的检索词和结构能尽可能多地检索到符合的研究吗? 2.4 基于时间、发表形式、语言的限制合适吗? 2.5 研究的筛选中尽可能地减小了误差吗?	3.1 数据提取中尽可能地减小了误差吗? 3.2 系统评价作者和读者能获取足够的研究特征来解读结果吗? 3.3 提取了所有相关的研究结果来进行数据合成吗? 3.4 使用了合适的工具来正规地评价偏倚风险(或方法学质量)吗? 3.5 偏倚风险评价中尽可能地减小了误差吗?	4.1 数据合成包括了所有应该包括的研究吗? 4.2 遵循了所有预先确定的分析吗?未遵循的部分解释了吗? 4.3 鉴于纳入研究的研究问题、研究设计和结局指标的性质和相似性,数据合成方法恰当吗? 4.4 数据合成中研究之间的差异(异质性)是最小的或者经过处理了吗? 4.5 结果稳定吗?例如是否通过敏感性分析来证明? 4.6 原始研究的偏倚最小吗?或者在数据合成中处理了吗?	A 结果解释中处理了领域1~4中所有偏倚风险吗? B 合理地考虑到了纳入研究与系统评价研究问题的相关性吗? C 评价者避免强调有统计学意义的结果了吗?
判断**	对纳入标准的描述的偏倚风险程度	研究检索和(或)筛选所使用方法的偏倚风险程度	数据提取和质量评价所使用方法的偏倚风险程度	数据合成和结果呈现的偏倚风险程度	系统评价的偏倚风险

* 标志性问题回答:是/可能是/否/可能否/无信息;** 偏倚风险程度判断:低/高/不确定

信度和实用性,但评价条目需要进一步修改和完善,而且其效度也未得到验证^[15]。

ROBIS 工具是目前值得推荐的一个系统评价偏倚风险评价工具,第 22 届和 23 届 Cochrane Colloquium 均对其进行了报道、讨论和相关培训。该工具的特点如下:第一,ROBIS 工具针对系统评价的“偏倚风险”,将系统评价设计、制作和结果解释中可能出现的偏倚以标志性问题一一提出,相较于以往“系统评价方法学质量”这一概念更为详细具体,并且最终的结果是以整个系统评价偏倚风险相关性的“高”、“低”、“不确定”来表示,这样比起 AMSTAR 工具以条目的答案数量或 R-AMSTAR 以最终分值来表示系统评价的方法学质量更明确;第二,制定者综合了大量证据,进行多轮德尔菲过程,开展面对面会议反复讨论,尽可能纳入所有的利益相关者参与制定,并经过了多次预试验不断进行完善,制定过程严谨、透明,基于证据;第三,ROBIS 工具不仅能够客观评估系统评价制定过程和结果解释过程中的偏倚风险、原始研究与系统评价研究问题的相关性,还特别评价目标问题与系统评价问题的相关性,以使系统评价的应用更准确;第四,ROBIS 工具不只是针对纳入 RCT 的系统评价,它可应用于所有干预性、诊断性、病因性和预后性等多种系统评价,而且其使用指导主要对非 RCT 系统评价做了实例解读^[23];第五,ROBIS 工具按照系统评价制定过程的顺序分别评价 4 个关键领域的偏倚风险情况,每个领域都有 5 ~ 6 个关键问题,每个问题都有相应的标准和对判断的说明,富有条理且十分详尽,避免了遗漏偏倚。

当然,ROBIS 工具并非十分成熟。例如,阶段一的结果未整合到阶段三中,即最终判断系统评价偏倚风险时并未考虑到目标问题与系统评价问题的相关性(如果目标问题存在的话);此外,评价条目(标志性问题)增多也意味着实际操作更为复杂耗时;另外,尽管工具制定过程中已经过多次预试验,ROBIS 工具的信度、效度和实用性及其推广应用情况仍有待时间的检验,而且这些预试验主要是由系统评价再评价作者开展,指南制定方面的预试验尚在进行中^[16]。目前该工具的 1.0 版已可获取,之后还会不断进行更新,我们相信 ROBIS 工具会日臻完善。

参考文献

- Mulrow CD. The medical review article: state of the science. *Ann Intern Med*, 1987, 106(3): 485-488.
- Cook DJ, Sackett DL, Spitzer WO. Methodologic guidelines for systematic reviews of randomized control trials in health care from the Potsdam Consultation on Meta-Analysis. *J Clin Epidemiol*, 1995,

- 48(1): 167-171.
- 3 Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. *Ann Intern Med*, 1997, 126(5): 376-380.
- 4 Moher D, Soeken K, Sampson M, et al. Assessing the quality of reports of systematic reviews in pediatric complementary and alternative medicine. *BMC Pediatr*, 2002, 2(2): 1-3.
- 5 Jadad A, Moher M, Browman G, et al. Systematic reviews and meta analyses on treatment of asthma: critical evaluation. *BMJ*, 2000, 321(7256): 537-540.
- 6 Jadad AR, Cook DJ, Jones A. Methodology and reports of systematic reviews and meta-analysis: a comparison of Cochrane paper-based journals. *JAMA*, 1998, 280(3): 278-280.
- 7 Schlosser RW, Goetze H. Effectiveness and treatment validity of interventions addressing self-injurious behavior: From narrative reviews to meta-analyses. *Advances in learning and behavioral disabilities*, 1992, 7: 135-175.
- 8 Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. *J Clin Epidemiol*, 1991, 44(11): 1271-1278.
- 9 Oxman AD, Guyatt GH, Singer J, Goldsmith CH, et al. Agreement among reviewers of review articles. *J Clin Epidemiol*, 1991, 44(1): 91-98.
- 10 Shea BJ, Grimshaw JM, Wells GA, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol*, 2007, 7: 10.
- 11 Shea BJ, Hamel C, Wells GA, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *J Clin Epidemiol*, 62(10): 1013-1020.
- 12 Sacks HS, Berrier J, Reitman D, et al. Meta-analyses of randomized controlled trials. *N Engl J Med*, 1987, 316(8): 450-455.
- 13 Kung J, Chiappelli F, Cajulis OO, et al. From systematic reviews to clinical recommendations for evidence-based health care: validation of revised assessment of multiple systematic reviews (R-AMSTAR) for grading of clinical relevance. *Open Dent J*, 2010, 4: 84-91.
- 14 Pieper D, Buechter RB, Li L, et al. Systematic review found AMSTAR, but not R (evised)-AMSTAR, to have good measurement properties. *J Clin Epidemiol*, 2014, 68(5): 574-583
- 15 Pieper D, Mathes T, Eikermann M. Can AMSTAR also be applied to systematic reviews of non-randomized studies? *BMC Res Notes*, 2014, 7: 609.
- 16 Whiting P, Savović J, Higgins JPT, et al. ROBIS group, ROBIS: A new tool to assess risk of bias in systematic reviews was developed. *J Clin Epidemiol*, 2015, S0895-4356(15)00308-X.
- 17 Hopp L. Risk of bias reporting in Cochrane systematic reviews. *International journal of nursing practice. Int J Nurs Pract*, 2015, 21(5): 683-686.
- 18 Moher D, Schulz KF, Simera I, et al. Guidance for developers of health research reporting guidelines. *PLoS Med*, 2010, 7(2): e1000217.
- 19 H.A. Linstone, M. Turoff. The Delphi method: Techniques and applications. Addison-Wesley, 1975.
- 20 Chandler J, Churchill R, Higgins J, et al. Methodological standards for the conduct of new Cochrane Intervention Reviews. The Cochrane Collaboration, 2012.
- 21 Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? *PLoS Med*, 2010, 7(9): e1000326.
- 22 熊俊, 陈日新. 系统评价/ Meta 分析方法学质量的评价工具 AMSTAR. *中国循证医学杂志*, 2011, 11(9): 1084-1089.
- 23 Whiting P, Savovic J, Higgins J, et al. ROBIS: Tool to assess risk of bias in systematic reviews Guidance on how to use ROBIS. Available at: <http://www.robis-tool.info/>.

收稿日期: 2015-10-24 修回日期: 2015-12-04
本文编辑: 樊斯斯、张永刚