

偏倚风险评估系列:(一)概述

杨智荣 孙凤 詹思延

CB18RN 英国剑桥大学临床医学院初级医疗中心(杨智荣); 100191 北京大学公共卫生学院流行病与卫生统计学系(孙凤、詹思延), 北京大学循证医学中心(孙凤、詹思延)

通信作者:詹思延, Email:siyan-zhan@bjmu.edu.cn

DOI:10.3760/cma.j.issn.0254-6450.2017.07.027

【摘要】 本文回顾了偏倚风险的概念,将系统综述中的偏倚风险评价与证据质量、方法学质量、报告质量、精确性、外部真实性等方面的评价作了对比,然后讨论了偏倚风险评估工具的进展、工具本身存在的不足、以及工具使用的常见问题等,以帮助我国系统综述者更好地理解偏倚风险评估及其工具的使用。

【关键词】 偏倚风险; 评估工具; 流行病学研究; 系统综述

基金项目:国家自然科学基金(81473067)

Risk of bias assessment: (1) overview Yang Zhirong, Sun Feng, Zhan Siyan
Primary Care Unit, Department of Public Health and Primary Care, School of Clinical Medicine, University of Cambridge, Cambridgeshire CB1 8RN, UK (Yang ZR); Department of Epidemiology and Biostatistics, School of Public Health (Sun F, Zhan SY), Center of Evidence-based Medicine and Clinical Research (Sun F, Zhan SY), Peking University, Beijing 100191, China
Corresponding author: Zhan Siyan, Email: siyan-zhan@bjmu.edu.cn

【Abstract】 This paper reviews the concept of risk of bias, followed by demonstrating why assessment of risk of bias in systematic reviews should be different from that of quality of evidence, methodological quality, reporting quality, precision, and external validity. We also discuss the recent development of tools for risk of bias assessment, the problems with the tools themselves, and the challenges in using these tools. This review may help systematic reviewers understand risk of bias assessment and the use of assessment tools.

【Key words】 Risk of bias; Tool for assessment; Epidemiological research; Systematic review

Fund program: National Natural Science Foundation of China (81473067)

循证医学强调任何诊疗决策的制定都应该全面考虑最佳证据、临床经验和患者选择,证据是影响现代医学实践的关键要素之一。与循证临床决策最为相关的证据主要来自流行病学,尤其临床流行病学研究,包括原始研究和二次研究。系统综述和Meta分析作为二次研究,在汇总原始研究结果和评价证据质量方面起着重要的作用。然而,当原始研究存在较高的偏倚风险时,对研究结果的盲目合并不仅不能排除原始研究中的偏倚,而且还很有可能提高有偏估计的精确性,从而进一步歪曲真实的效应。为了减少原始研究偏倚对结果汇总的影响,所有系统综述都应该对纳入研究的偏倚风险进行评估。评估的结果也为进一步的证据质量分级提供参考依据。

一、偏倚风险评估及其相关概念

偏倚(Bias)是流行病学里非常重要的概念,又

称系统误差(Systematic error),是指研究结果与真实情况之间所存在的系统性偏差^[1]。偏倚的来源有多种,任何在设计、实施、分析或报告等研究环节中出现的,都有可能高估或低估真实的情况,影响研究的内部真实性(Internal validity)。同一种偏倚来源,在不同的研究中所产生的偏倚大小和方向可能是不同的,加之真实的效应通常不能确切知道,因此我们难以定量评价偏倚的程度,而只能定性判断偏倚风险(Risk of bias)是否存在及其可能对结果产生影响的方向^[2]。例如,在某个随机对照试验(Randomized controlled trial, RCT)中未采取分配隐藏,并且失访的人数和原因在组间不可比,这样的偏倚来源是否真的会对研究结果产生系统性的影响,通常不能完全确定。但可以肯定的是,该研究在随机分组和数据完整性方面存在很高的偏倚风险。偏倚风险评估的主要内容是判断偏倚来源是否存在。

偏倚来源可以是研究水平的,如随机序列的产生、分配隐藏等;也可以是结局水平的,如结局测量者的盲法、结局的选择性报告等。如果某种偏倚来源存在,则认为偏倚风险高;如果不存在,则认为偏倚风险低。除了对单个偏倚来源作出判断以外,评估者还可以对整体的偏倚风险作出评价。过去常用量表的合计得分作为整体评价的手段,但由于这种评分方式存在权重分配不合理的问题,现在已不推荐使用,取而代之的是综合评价各关键偏倚来源是否存在。这种整体的评价通常是针对特定结局而言的,不同结局的偏倚评估结果可能不同^[2]。

偏倚风险的概念与内部真实性相对应,评估时应注意其与证据质量(Quality of evidence)、方法学质量(Methodological quality)、报告质量(Reporting quality)、精确性(Precision)、外部真实性(External validity)等概念的区别与联系,不要相互混淆^[2-3]。

在系统综述中,证据质量是指效应估计值能够正确反映真实情况的把握程度。根据 GRADE 分级系统,证据质量可分为高、中、低、极低 4 个等级。评价的是某个特定结局的整个证据体,而不是针对单个研究(也可能在证据体中仅有一个研究,那是特例)。证据质量的 5 个降级因素有研究的局限性(即偏倚风险)、不一致性、间接性、精确性和发表偏倚,而 3 个升级因素则包括效应大小、剂量反应关系和残余混杂。由此可见,偏倚风险不等于证据质量,但它是影响证据质量分级的重要因素之一。

方法学质量是经常跟偏倚风险互换使用的表述,方法学质量高往往意味着偏倚风险低。但两者实质仍然存在一定差别,前者是指研究的某重要环节是否达到所预期的最高标准。例如,在一项评价针灸干预效果的 RCT 中,即使设计和实施的质量再高,也不可能对施灸者实施盲法,因此我们不能在盲法这一方面判断为方法学质量低,但仍然可以判断为偏倚风险高。

报告质量是与偏倚风险非常相关但明显不同的概念。由于偏倚风险的评估很大程度上是基于研究发表的信息来判断的,因此报告质量可影响偏倚风险评价的结果,但跟偏倚风险实质的高低并无直接关系。例如,在一个 RCT 中,作者采用了分配隐藏,但没有在文章里对此进行报告;若仅仅依据作者所报告的信息,在这种情况下,偏倚风险只能评为不清楚,但实际上应该是低风险。此外,报告质量的评估有专门的工具,如 RCT、诊断试验、观察性研究、系统综述相应的报告规范:CONSORT、STARD、

STROBE 和 PRISMA^[4-7]。

与精确性相对应的概念是随机误差。假设在一个总体中多次抽样对同一个问题进行研究,由于随机误差的存在,所得到的结果会不尽相同。精确性受样本量影响,样本量越小,精确性越低,95%CI 越宽。若不存在偏倚的影响,那么多个重复样本的估计值的平均水平就会与总体情况接近。因此,样本量的大小和 95%CI 的宽窄不宜作为偏倚风险评估的内容。

外部真实性是指研究结果的外推性或适用性,关注的是研究问题本身是否与实际决策的问题相关。获得良好的外推性的前提条件之一,就是研究有较好的内部真实性。内部真实性关注的是研究本身是否“正确”回答了所提出的研究问题。而至于研究中存在“错误”回答的可能性,那就涉及本篇所讲的偏倚风险评估。

另外需要注意的是,虽然偏倚风险评估内容没有涵盖上述的精确性和外部真实性,但这三者都是在应用 GRADE 分级系统给出证据质量评级和推荐意见之前,必须考虑的重要因素^[3]。

二、偏倚风险评估工具的进展

不同流行病学研究设计的偏倚来源不尽相同,针对同一研究设计不同的学术研究机构所制定的偏倚风险评估标准也有所差别。国际上已经陆续发布了上百个偏倚风险评估工具(“工具”),其中有些工具已经在循证医学实践中得到了广泛的应用,如针对 RCT 的 Cochrane 工具^[8],针对诊断试验研究的 QUADAS^[9],针对队列研究和病例对照研究的 NOS 量表^[10]、针对系统综述的 AMSTAR^[11]等。由于工具种类繁多、质量参差不齐、部分条目设置不合理,工具使用者(尤其是系统综述的制作者)在进行工具选择和偏倚风险评价时仍存在很大困扰。因此,近三年这类工具又陆续进行了淘汰、更新或重新制定,主流的工具包括本次“偏倚风险评估系列”中要介绍的 RoB2.0^[12]、ROBINS-1^[13]、QUADAS-2^[14]、QUIPS^[15]和 ROBIS^[16]。从这些工具的制定或历次更新来看,其发展呈现出以下特点。

1. 偏倚来源的覆盖更全面,重点更突出。早期广为使用的 Jadad 量表只涵盖 RCT 的 3 个方面:随机、双盲和退出。与之相反,有些工具把 RCT 的方法学质量、报告质量、精确性、外部真实性等方面也纳入评价体系,其条目数高达 57 项^[17]。这些过于简单或复杂的工具后来逐渐被 Cochrane 工具取代^[8],该工具只选取并突出了 RCT 偏倚风险相关的 6 个

方面,全部是偏倚的主要来源。该工具虽未明确纳入干预的依从性、基线特征的不可比等重要偏倚来源,但这些方面在新一代的RoB2.0中都作了补充^[12]。

2. 工具的适用范围开始以研究问题为主要切入点,兼顾多种研究设计类型。过去大多数常用的评估工具都只是考虑某类研究设计,如NOS量表^[10],适用于多种研究问题的队列研究和病例对照研究,而且两者分别有不同的量表。最近发布的ROBINS-1^[13],围绕干预有效性的问题,同时适用于类实验和队列研究。又如QUADAS-2^[14],针对诊断试验准确性评价的问题,同时适用于横断面研究、队列研究和病例对照研究。ROBIS^[16]的适用范围更是覆盖了干预、诊断、病因、预后四大类临床研究问题。

3. 工具的主流模式由传统的量表式和清单式向领域评估式(Domain-based evaluation)转变。这种领域评估的模式在2011年更新的Cochrane工具中初见端倪^[8],该工具把偏倚来源划分成6个领域,每个领域有1~2个条目,然而该工具尚未形成与传统清单的明显区别。这种模式随后在QUADAS-2中开始变得成熟^[14],它要求评价者在偏倚来源的4个领域中分别设置数个信号问题(Signalling questions),然后依据对这些问题的回答,按照事先制定的评价标准对每个领域的偏倚风险作出判断。自此至今,主流工具中无一不是采取领域评估这种模式。

4. 评价者可根据具体研究的问题,在正式开始评估前对信号问题进行合理的增减或修改。过去常用的工具,不管是量表里的评分规则,还是清单里的评价项目,都有严格的限定。但这种限定在主流工具中首先被2011年版的Cochrane工具打破^[8]。它加入了“Other bias”这一领域,允许评价者对条目进行合理的设置。此外,某些领域如不完整数据、结局选择性报告,也不再设有固定的评价标准。在随后所有采用领域评估式的主流工具中,不仅信号问题可以适当调整,对各领域偏倚风险的判断标准也可由评价者自己合理制定。有的工具还明确设置了外部真实性的领域供研究者选择评价,如QUADAS-2^[14]。

5. 评价流程变得更复杂。这一点在RCT的历代工具中体现得尤为明显。最初的Jadad量表^[18],评价者只需要对随机、双盲和退出三方面进行评分。2008年版的Cochrane清单式工具^[19],评价者需要对6个条目进行“Yes”、“No”或“Unclear”的回答。2011年版的Cochrane领域评估式工具^[8],对盲法和其他偏倚这两个方面进行了完善,并把偏倚来源明确划分

为6个领域,要求评价者对各领域作出“Low risk”“High risk”或“Unclear risk”的回答。而如今2016年版的Cochrane RoB2.0^[12],则需要对5个领域作出判断(每个领域需判断为Low risk、High risk或Some concerns),每个领域要回答数个信号问题,每个问题有5个选项(Yes、Probably yes、No、Probably no或No information),问题之间可能涉及跳转,在对这些信号问题进行回答前还要求作者对所关心的效应(干预的分配效应还是依从效应)和纳入的研究类型(平行、交叉、整群)进行定义。其评价流程的复杂程度已不可同日而语。

三、工具本身和工具使用的问题

尽管已历经多年的发展和完善,这些工具至今仍存在很多不可忽视的问题。我们把这些问题归纳为两大方面:工具本身和工具使用。

1. 工具本身的问题:

(1) 工具的可重复性(Reproducibility)尚不理想。有多个研究对2008和2011年版的Cochrane工具的可重复性作了评价,整体的Kappa值介于0.21~0.55之间,各条目或领域的Kappa值在0.19~0.86之间不等^[20-24],其中可重复性较低的领域有不完整数据、结局的选择性报告等。造成评估结果难以重复的原因,可能是最新版Cochrane手册上只是提供了各条目的实例^[2],辅以原理解释,并没给出各条目的统一判断标准,不同研究者对这些实例的理解可能会有所差别。而可重复性问题在最近制定的工具仍然存在,如QUIPS的整体Kappa值为0.56~0.82^[15],QUADAS-2的整体Kappa值为0.03~1.0^[14]。有学者因此呼吁,偏倚风险的评估需要更客观可靠的工具^[25]。

(2) 灵敏度和特异度无法评价,不存在公认准确的工具。在评价筛检或诊断试验的准确性时,可以将其跟参考试验,即“金标准”进行比较。但在偏倚风险评估领域,由于不可能存在这样的“金标准”,所有的工具均无法进行灵敏度和特异度的评价,因此没有公认准确的工具^[26]。某些工具的选用更普遍,并非因为它们能够更准确地识别偏倚风险,而是因为它们的制定可能更“循证”,偏倚来源的设置可能更合理。

(3) 制定流程和方法尚未标准化,发布后仍需适时更新。工具的制定步骤通常包括适用范围的确 定、文献综述、共识会议、预实验和内容修订等。但每一步的实施,现在还不像临床循证指南的制定那样有标准的流程和方法,工具的制作质量也因此参

差不齐。工具发布以后需要不断接受使用者的反馈,及时更新和完善,保持生命力。但这样的工具并不多,Cochrane 工具是其中非常经典的一个。

(4)结局水平的评价模式仍有待改进。一个系统综述可能会考量多个结局,而每个结局在同一个研究中受偏倚来源影响的程度可能是不一样的。最理想的偏倚风险评估应该是针对结局水平的,即对每个结局(包括不同的随访时间点)分别进行相应的偏倚风险评估。虽然现有的主流工具已经可以实现这一评价模式,但这种模式会涉及不同结局的界定以及大量信息的重复利用,工作量将大大增加,致使该模式至今仍未在系统综述(包括 Cochrane 系统综述)中推广应用。但可以预见的是,这模式将是今后工具改进的重点内容之一^[12]。

(5)评估流程复杂化的必要性。领域评估式工具的评估流程变得较为复杂,以期更“精准”地评价偏倚风险。但这种复杂化是否能实质性地改变评估结果,是否能在系统综述中推广使用,最后是否能真正帮助循证决策,都是有待实践检验和进一步研究的问题。

2. 工具使用的问题:

(1)工具选用的多样化,评价结果受主观影响大。工具的选择当前尚无统一要求,大部分情况下都是由研究者根据纳入研究类型而自行决定的。有研究纳入了 210 个干预相关的系统综述,发现其中有 38 个工具用于偏倚风险的评估,使用最多的是 Cochrane 工具(36.6%)和 Jadad 量表(11.1%),而在 84 个纳入观察性研究的系统综述中,共有 24 个工具被使用,其中 NOS 量表占首位(35.1%)^[27]。虽然工具很多,但不管选用什么工具,偏倚风险评估都是一个依据原始研究作者所报告的客观事实来进行主观判断的过程,因而带有一定的不确定性。即使是选用同一个工具,评价结果也会受工具本身可重复性和评价者主观性的影响。例如,同是使用 Cochrane 工具对相同的研究进行评价,结果的不一致率可高达 65%^[28]。有研究提示,一致率低的原因是 Cochrane 工具的不合理使用,工具内容的设置和指导说明的文件也有待改进^[29]。

(2)评价标准的报告不透明。由于领域评估式的主流工具允许对评价标准进行修改,如果评估者不能清楚地报告修改的内容和原因,很容易造成对评估结果的选择性报告^[30]。这问题通过系统综述的报告规范^[7]以及系统综述的事先注册^[31]可能在一定程度上得以解决。然而,除非是 Cochrane 系统综述

的研究方案,其他杂志上发表的研究方案和 PROSPERO 的方案注册目前均未明确要求对工具的判断标准进行全面的报告。即使是在众多非 Cochrane 系统综述的全文和附件中,也难以找到描述。而对于我国学者发表的系统综述,偏倚评估的报告质量也不容乐观,亟需改善^[32-33]。

(3)评估结果的使用仍不充分。在系统综述中,评估结果的意义不仅在于展现原始研究的偏倚风险概况,更重要的还在于去伪存真,帮助提高数据分析的科学性和结果解读的合理性^[2]。处理评估结果有多种方式,常见的包括纳入标准制定、敏感性分析、亚组分析、Meta 回归、结果讨论、GRADE 证据分级等。然而,大部分系统综述在评估结果的使用方面还做得不够充分^[29,34-35],对评估结果仍只是作简单的文字描述或图表展示。

(4)知识投入和时间成本高。由于领域评估式工具的复杂化,评价者若不具备扎实的流行病学专业基础,很难在短时间内掌握工具使用的正确方法。有针对性地进行指导和培训可能有助于提高偏倚风险评估的质量^[36]。但即使是经验丰富的系统综述者,在使用所熟悉的工具时也可能需要较长的时间。例如预后研究的系统综述者使用 QUIPS 评估一篇文章的中位时间为 20 min,个别评估者需要 1 h 以上^[15]。

四、本系列讲座的导读

由于存在上述各种问题和挑战,偏倚风险评估一直是系统综述制作中的难点之一。不管是在国内还是国外,工具的使用和结果的报告仍然有很大改进的空间。为进一步规范和帮助我国相关研究者正确理解和使用这批新工具,北京大学循证医学中心和中国医师协会循证医学专委会青年委员会专门组织撰写了这一系列。本系列分为 9 个讲座,本讲为概述,从第二讲到第八讲分别就随机对照试验的 3 种常见研究设计(平行、交叉和整群)、非随机的干预研究、诊断试验、预后研究、系统综述等的工具进行介绍,最后一讲介绍如何在系统综述中合理使用偏倚风险评估的结果。本系列将体现最新的研究进展,从实用性的角度进行介绍,强调各工具使用时的重点和难点,以促进偏倚风险评估在方法和报告质量上的提高。通过对这些评估工具的学习,也可以帮助读者更好地理解、识别和控制流行病学研究中各种常见的偏倚来源,从而为高质量证据的生产提供进一步的保障。

利益冲突 无

参 考 文 献

- [1] Rothman KJ, Greenland S, Lash TL. Modern Epidemiology [M]. 3rd ed. Philadelphia: Lippincott Williams & Wilkins, 2008.
- [2] Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [EB/OL]. The Cochrane Collaboration, 2011-03. www.handbook.cochrane.org.
- [3] Guyatt G, Oxman AD, Akl EA, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables [J]. J Clin Epidemiol, 2011, 64 (4) : 383-394. DOI: 10.1016/j.jclinepi.2010.04.026.
- [4] Schulz KF, Altman DG, Moher D, et al. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials [J]. BMJ, 2010, 340: c332. DOI: 10.1136/bmj.c332.
- [5] Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies [J]. BMJ, 2015, 351: h5527. DOI: 10.1136/bmj.h5527.
- [6] von Elm E, Altman DG, Egger M, et al. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies [J]. BMJ, 2007, 335 (7624) : 806-808. DOI: 10.1136/bmj.39335.541782.AD.
- [7] Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and Meta-analyses: the PRISMA statement [J]. BMJ, 2009, 339: b2535. DOI: 10.1136/bmj.b2535.
- [8] Higgins JP, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials [J]. BMJ, 2011, 343: d5928. DOI: 10.1136/bmj.d5928.
- [9] Whiting P, Rutjes AW, Reitsma JB, et al. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews [J]. BMC Med Res Methodol, 2003, 3: 25. DOI: 10.1186/1471-2288-3-25.
- [10] Wells G, Shea B, O'Connell D, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in Meta-analyses [EB/OL]. 2013 [2017-01-10]. http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp.
- [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]. J Clin Epidemiol, 2009, 62 (10) : 1013-1020. DOI: 10.1016/j.jclinepi.2008.10.009.
- [12] Higgins JPT, Savović J, Page MJ, et al. A revised tool to assess risk of bias in randomized trials (RoB 2.0) [EB/OL]. (2016). <https://sites.google.com/site/riskofbiastool/welcome/rob-2-0-tool>.
- [13] Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions [J]. BMJ, 2016, 355: i4919. DOI: 10.1136/bmj.i4919.
- [14] Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies [J]. Ann Intern Med, 2011, 155 (8) : 529-536. DOI: 10.7326/0003-4819-155-8-201110180-00009.
- [15] Hayden JA, van der Windt DA, Cartwright JL, et al. Assessing bias in studies of prognostic factors [J]. Ann Intern Med, 2013, 158 (4) : 280-286. DOI: 10.7326/0003-4819-158-4-201302190-00009.
- [16] Whiting P, Savović J, Higgins JP, et al. ROBIS: A new tool to assess risk of bias in systematic reviews was developed [J]. J Clin Epidemiol, 2016, 69: 225-234. DOI: 10.1016/j.jclinepi.2015.06.005.
- [17] 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. DOI: 10.1016/0197-2456(94)00031-W.
- [18] 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. DOI: 10.1016/0197-2456(95)00134-4.
- [19] Higgins JPT, Green S. Cochrane handbook for systematic reviews of interventions [M]. Chichester: Wiley, 2008.
- [20] Hartling L, Ospina M, Liang YY, et al. Risk of bias versus quality assessment of randomised controlled trials: cross sectional study [J]. BMJ, 2009, 339: b4012. DOI: 10.1136/bmj.b4012.
- [21] Hartling L, Hamm MP, Milne A, et al. Testing the risk of bias tool showed low reliability between individual reviewers and across consensus assessments of reviewer pairs [J]. J Clin Epidemiol, 2013, 66 (9) : 973-981. DOI: 10.1016/j.jclinepi.2012.07.005.
- [22] Hartling L, Bond K, Vandermeer B, et al. Applying the risk of bias tool in a systematic review of combination long-acting beta-agonists and inhaled corticosteroids for persistent asthma [J]. PLoS One, 2011, 6 (2) : e17242. DOI: 17210.11371/journal.pone.0017242.
- [23] Graham N, Haines T, Goldsmith CH, et al. Reliability of 3 assessment tools used to evaluate randomized controlled trials for treatment of neck pain [J]. Spine, 2012, 37 (6) : 515-522. DOI: 10.1097/BRS.0b013e31822671eb.
- [24] Armijo-Olivo S, Ospina M, da Costa BR, et al. Poor reliability between Cochrane reviewers and blinded external reviewers when applying the Cochrane Risk of Bias Tool in physical therapy trials [J]. PLoS One, 2014, 9 (5) : e96920. DOI: 10.1371/journal.pone.0096920.
- [25] Berger VW, Mickenautsch S. On the need for objective measures of risk of bias [J]. Contemp Clin Trials, 2015, 41: 202-203. DOI: 10.1016/j.cct.2015.02.001.
- [26] 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. DOI: 10.1186/1471-2288-4-22.
- [27] Seehra J, Pandis N, Koletsi D, et al. Use of quality assessment tools in systematic reviews was varied and inconsistent [J]. J Clin Epidemiol, 2016, 69: 179-184. DOI: 10.1016/j.jclinepi.2015.06.023.
- [28] Jordan VM, Lensen SF, Farquhar CM. There were large discrepancies in risk of bias tool judgments when a randomized controlled trial appeared in more than one systematic review [J]. J Clin Epidemiol, 2016, 81: 72-76. DOI: 10.1016/j.jclinepi.2016.08.012.
- [29] Jørgensen L, Paludan-Müller AS, Laursen DRT, et al. Evaluation of the Cochrane tool for assessing risk of bias in randomized clinical trials: overview of published comments and analysis of user practice in Cochrane and non-Cochrane reviews [J]. Syst Rev, 2016, 5: 80. DOI: 10.1186/s13643-016-0259-8.
- [30] Faggion CM Jr. The rationale for rating risk of bias should be fully reported [J]. J Clin Epidemiol, 2016, 76: 238-238. DOI: 10.1016/j.jclinepi.2017.03.007.
- [31] Booth A, Clarke M, Dooley G, et al. The nuts and bolts of PROSPERO: an international prospective register of systematic reviews [J]. Syst Rev, 2012, 1: 2. DOI: 10.1186/2046-4053-1-2.
- [32] Liu YL, Yang SP, Dai JJ, et al. Risk of bias tool in systematic reviews/meta-analyses of acupuncture in Chinese journals [J]. Plos One, 2011, 6 (12) : e28130. DOI: 10.1371/journal.pone.0028130.
- [33] Zhang J, Wang J, Han L, et al. Tools to assess risk of bias in systematic reviews of nursing intervention in China: global implications of the findings [J]. Nurs Outlook, 2016. DOI: 10.1016/j.outlook.2016.11.004.
- [34] Hopewell S, Boutron I, Altman DG, et al. Incorporation of assessments of risk of bias of primary studies in systematic reviews of randomised trials: a cross-sectional study [J]. BMJ Open, 2013, 3 (8) : e003342. DOI: 10.1136/bmjopen-2013-003342.
- [35] Katikireddi SV, Egan M, Petticrew M. How do systematic reviews incorporate risk of bias assessments into the synthesis of evidence? A methodological study [J]. J Epidemiol Community Health, 2015, 69 (2) : 189-195. DOI: 10.1136/jech-2014-204711.
- [36] Savović J, Weeks L, Sterne JA, et al. Evaluation of the Cochrane Collaboration's tool for assessing the risk of bias in randomized trials: focus groups, online survey, proposed recommendations and their implementation [J]. Syst Rev, 2014, 3: 37. DOI: 10.1186/2046-4053-3-37.

(收稿日期:2017-01-17)

(本文编辑:王岚)