

系统评价/Meta分析方法学质量的评价工具 AMSTAR

熊俊^{1,2} 陈日新^{1*}

1. 江西中医学院附属医院(南昌 330006); 2. 广州中医药大学(广州 510405)

摘要 AMSTAR 是国外最新研发的用于评价系统评价/Meta 分析方法学质量的测量工具, 具有良好的效度、信度和反应度, 得到了较为广泛的运用。本文就 AMSTAR 的研发过程、评价条目、运用现状等问题逐一介绍给国内研究者和使用者。

关键词 循证医学; 系统评价; Meta 分析; 方法学; 质量评价; AMSTAR

An Introduction to a Measurement Tool to Assess the Methodological Quality of Systematic Reviews/ Meta-analysis: AMSTAR

XIONG Jun^{1,2}, CHEN Ri-xin^{1*}

1. The Affiliated Hospital of Jiangxi University of Traditional Chinese Medicine, Nanchang 330006, China;

2. Guangzhou University of Traditional Chinese Medicine, Guangzhou 510405, China

Abstract AMSTAR (Assessment of Multiple Systematic Reviews) is currently developed as a measurement tool with extensive application to assess the methodological quality of systematic review/meta-analysis. It has good reliability, validity, and responsibility, and has been widely applied. This paper introduces AMSTAR to researchers and users in China, in view of development procedure, assessment items, and application status.

Key words Evidence-based medicine; Systematic review; Meta-analysis; Methodology; Quality assessment; AMSTAR

系统评价是循证医学重要的研究方法和最佳证据的重要来源, 被公认为是评价临床疗效、制订临床指南和规范的基石^[1], 尤其是国际 Cochrane 协作网生产的系统评价, 以其严格的方法学而著称于世。第 5 版《流行病学词典》对系统评价赋予了准确的定义: 系统评价是运用减少偏倚的策略, 严格评价和综合针对某一具体问题的所有相关研究, Meta 分析可能但不一定是这个过程的一部分^[2,3]。可见, 减少研究偏倚, 确保研究结果的真实性和完整性是系统评价的灵魂。但随着系统评价/Meta 分析数量的快速增长, 其研究质量的参差不齐也愈来愈引起国内外学者的关注和担忧^[4-7]。尽管系统评价/Meta 分析是临床最佳证据的来源之一, 但只有高质量的系统评价/Meta 分析才能为临床医师、患者及其他决策者提供科学的依据^[8]。反之, 很可能误导决策者。所以,

基金项目: 国家重点基础研究发展计划(“973”计划)资助项目(编号: 2009CB522902); 国家自然科学基金项目(编号: 30760320); 江西省重大科技支撑计划项目

作者简介: 熊俊, 男(1983年~), 博士研究生, 住院医师, 助教。研究方向: 艾灸效应及其机制研究, 循证医学。

Email: xiongjun196071@yahoo.com.cn

* 通讯作者, Email: chenrixin123@yahoo.com.cn

进行有效的质量评价是正确使用系统评价/Meta 分析, 以及谨慎对待其研究结果的重要环节。

目前, 有关系统评价/Meta 分析质量评价的工具主要包括两类, 一类为方法学质量评估工具, 包括 OQAQ (Oxman-Guyatt Overview Quality Assessment Questionnaire)^[10] 和 SQAC (Sack's Quality Assessment Checklist)^[11] 等, 另一类为报告质量评估工具, 包括 QUOROM (the quality of reporting of meta-analyses) 声明^[11] 等, 上述评价工具在国内均有相应的中文版^[4,13]。但这些量表或清单在使用过程中遇到一些亟待解决的问题, 特别是评价方法学质量的测量工具能否全面有效涵盖偏倚来源引起了研究者的争论^[14]。

鉴于此, 本文期望通过对 AMSTAR 量表的研发背景、评价条目、运用现状以及今后的发展趋势的介绍, 为国内研究者和使用者提供一个更为实用、有效、全面的质量评价工具。

1 背景

2007 年, 来自荷兰 VU 大学(Vrije Universiteit University) 医学研究中心和加拿大渥太华大学的

临床流行病学专家们在英国医学委员会期刊《医学研究方法学》上发表了名为“Development of AMSTAR: A Measurement Tool to Assess Systematic Reviews”的专论, 标志着 AMSTAR 的正式形成^[15]。研发小组认为, 系统评价在跟进医学专业最新知识和信息、形成有价值的临床证据以及影响医疗卫生决策方面扮演着难以替代的重要角色。但系统评价制作过程是一项复杂、需要谨慎进行的工作, 研究者和使用者都必须检查其真实性。研究如何评价系统评价/ Meta 分析的质量已逾 10 年, 如 Oxman、Guyatt 以及 Sacks 研制的测量工具。已有超过 24 种测量工具被报道使用过。然而, 它们绝大多数运用不广泛, 有些测量工具要么条目非常冗长, 要么使用过程很复杂、不便于把握。尤其是在不断的运用过程中, 许多实证性证据发现了系统评价诸多偏倚来源。比如, 最近的方法学研究强调发表文种和出版偏倚是系统评价的重要潜在偏倚^[16-18]。所以, 研发小组基于以上认识, 把有参考价值的代表性评价工具、长期使用过程中形成的实证性证据以及专家共识作为研发 AMSTAR 的科学来源。

2 评价条目

AMSTAR 是用于衡量系统评价/ Meta 分析的避免或减少偏倚的程度, 即方法学质量的一种量表。它的条目形成基础有 OQAQ 的 10 个条目、SQAC 的 24 个条目以及另外 3 个考虑文种偏倚、发表偏倚和灰色文献的条目。研发组采用探索性因素分析和名义群体技术保证了量表的表面效度和内容效度^[15]。在后续研究中, 该研发组进一步考验了 AMSTAR 的信度、结构效度和实用性^[19]。还专门委托加拿大药物卫生技术评估中心评估其科学性, 评估结果也十分令人满意^[20]。在上述工作的基础上, 研发小组正式提出了 AMSTAR 的标准条目, 共 11 个条款, 其英文版可从 <http://www.biomedcentral.com/content/supplementary/1471-2288-7-10-S1.doc> 上免费获取, 开放使用。本文把 AMSTAR 用中文表述成表 1。每个条款的评语选项有“是”、“否”、“不清楚”以及“未采用”。

3 条目解读

偏倚是影响系统评价/ Meta 分析质量的主要因素。只有采用必要措施避免或消除偏倚来源, 减小偏倚影响程度, 研究者才能确保研究结果真实性。而方法学质量即是考察系统评价/ Meta 分析在研究设

表 1 AMSTAR 评价清单及说明

条目	描述及说明
1	是否提供了前期设计方案? • 在系统评价开展以前, 应该确定研究问题及纳入排除标准
2	纳入研究的选择和数据提取是否具有可重复性? • 至少要有两名独立的数据提取员, 而且采用合理的不同意见达成一致的方法过程
3	是否实施广泛全面的文献检索? • 至少检索 2 种电子数据库。检索报告必须包括年份以及数据库, 如 Central、EMbase 和 MEDLINE。必须说明采用的关键词/ 主题词, 如果可能应提供检索策略 • 应咨询最新信息的目录、综述、教科书、专业注册库, 或特定领域的专家, 进行额外检索, 同时还可检索文献后的参考文献
4	发表情况是否已考虑在纳入标准中, 如灰色文献? • 应该说明评价者的检索是不受发表类型的限制 • 应该说明评价者是否根据文献的发表情况排除文献, 如语言
5	是否提供了纳入和排除的研究文献清单? • 应该提供纳入和排除的研究文献清单
6	是否描述纳入研究的特征? • 原始研究提取的数据应包括受试者、干预措施和结局指标等信息, 并以诸如表格的形成进行总结 • 应该报告纳入研究的一系列特征, 如年龄、种族、性别、相关社会经济学数据、疾病情况、病程、严重程度等
7	是否评价和报道纳入研究的科学性? • 应提供预先设计的评价方法, 如治疗性研究, 评价者是否把随机、双盲、安慰剂对照、分配隐藏作为评价标准, 其它类型研究的相关标准条目一样要交代
8	纳入研究的科学性是否恰当地运用在结论的推导上? • 在分析结果和推导结论中, 应考虑方法学的严格性和科学性。在形成推荐意见时, 同样需要明确说明
9	合成纳入研究方法是否恰当? • 对于合成结果, 应采用一定的统计检验方法确定纳入研究是可合并的, 以及评估它们的异质性(如 Chi-squared test)。如果存在异质性, 应采用随机效应模型, 和 / 或考虑合成结果的临床适宜程度, 如合并结果是否敏感?
10	是否评估了发表偏倚的可能性? • 发表偏倚评估应含有某一种图表的辅助, 如漏斗图, 以及其他可行的检测方法和 / 或统计学检验方法, 如 Egger 回归
11	是否说明相关利益冲突? • 应清楚交待系统评价及纳入研究中潜在的资助来源

计、实施过程中对偏倚来源的控制情况。AMSTAR 依据 11 个条目检测系统评价 /Meta 分析控制偏倚的情况。为了提高本标准的可操作性, 本文根据 AMSTAR 研发小组的发表论文进行解读如下。

条目 1 指出一个合格的系统评价 /Meta 分析必须预先制定前期设计方案。这是区别于传统综述的重要方面。这样, 系统评价 /Meta 分析的方法得到预先确定, 使得其研究虽然是回顾性质的, 其纳入的研究大多是已完成并公开发表的研究, 但其研究过程可以尽可能严格和精细, 从而保持前瞻性。本文建议研究者应该尽可能在 Cochrane 协作网注册相应的研究计划书。在发表研究结果时, 文中就可以标注研究方案的注册号。这样能极大方便证据使用者判断研究者是否提供了前期设计方案。AMSTAR 进一步解释, 前期的设计方案应当有确定的研究问题及纳入排除标准。研究问题可以按照循证医学的 PICO 原则进行构建, 包括患者 / 人群、干预措施 / 暴露因素, 对照以及结果四个方面。一项完整的纳入排除标准清单同样至少包括四个方面的内容: 研究类型、研究对象、干预措施、结局指标, 以及独立或分列上述要素的排除标准。需要指出的是纳入标准和排除标准并不是互补关系, 排除标准是在符合纳入标准的文献中进一步筛除掉影响研究偏倚的文献。

条目 2 主要考察选择研究对象和提取数据的可重复性。目前常用的控制措施有: 采用多人或盲法选择, 也可采用专业与非专业人员相结合的共同选择, 对选择文献过程中存在的意见分歧可通过讨论或第三人的方法解决, 如果可能, 最好进行预试验, 以摸索经验, 标化和统一选择文献。文献资料的选择应分三步进行: ① 初筛, 根据检出引文信息, 如题目、摘要筛除明显不合格的文献, 对肯定或不能肯定的文献应查出全文再进行筛选。② 阅读全文, 对可能合格的文献资料, 应该逐一阅读和分析, 以确定是否合格。③ 最好与作者联系, 一旦被排除的文献将不再录用, 因此, 如果文中提供的信息不全面而不能确定, 或者有疑问和有分歧的文献应先纳入, 通过作者联系获得有关信息后再决定取舍或在以后的选择过程中进一步评价。同样提取数据也是直接影响结果的准确性。为了保证数据提取的准确性, 要求至少两名评价人员各自独立提取资料, 然后相互复核, 准确无误和意见统一后才能进行下一步。因此, 该条目要求研究者交代清楚保证选择研究对象和提取数据的可重复性的措施。

条目 3 用于衡量制定的检索策略是否全面。这里实际包括两个方面的内容, 一是检索策略的要素

要齐全。一项完整的检索策略是指在解析相关问题的基础上, 明确检索目的和信息需求, 选择检索系统, 确定检索词, 构造检索式, 从而制定出较为完善的检索计划或方案。一般在 Cochrane 系统评价中, 每个数据库的检索策略均详细给出。而国内期刊由于版面和作者的考量, 很少有全部检索策略的描述, 这可以理解。但是检索词、简练的检索式和日期应该有所体现, 最好能详细描述一个有代表性的数据库的检索策略, 这样既能方便证据使用者评价研究的质量, 也能提示研究者重视检索策略的制定。二是检索范围要广。原则上, 一个系统评价 /Meta 分析必须检索世界上可能得到的所有文献来源, 但在实际操作中, 由于信息资源的限制, 完美的全面检索是很困难的。因此, AMSTAR 提出了最低要求: 至少两种电子资源。对于国内的研究者, 本文建议必须要检索至少 4 个数据库: MEDLINE、EMbase、CBM、Cochrane Library。MEDLINE 是目前第一大医学文献数据库, 是每一个系统评价 /Meta 分析研究者必须检索的, 而 EMbase 比前者收录更多的药学文献, 更新更频繁, 基本涵盖欧洲大部分出版的医学期刊。CBM 的独特之处在于完善的主题词检索系统, 和美国医学图书馆主题词检索相一致, 检索起来准确性高。Cochrane Library 的优势在于 Cochrane 协作网自行收集全世界的随机对照试验, 依靠遍布全世界的协作中心或机构, 采用专业人员手工采集试验信息, 具有多语种、标引率高的特点。不论怎样, 在一般的操作过程中, 研究者必须至少检索一国内大型数据库, 如 CNKI、VIP、CBM、WF 等, 和至少一个国外大型数据库, 如 MEDLINE、EMbase、SCI 等。本条目还强调了补充检索的重要性。咨询最新信息的目录、综述、教科书、专业注册库, 或特定领域的专家, 进行额外检索, 同时还可检索文献后的参考文献。

条目 4 要求研究者交代发表情况是否已考虑在纳入标准中, 如灰色文献。该条目是考察发表偏倚和查寻偏倚的。一个高质量的系统评价 /Meta 分析应该在可能情况下包括各种类型的文献, 即使是灰色文献。例如, 对作者而言, 多数情况下, 愿意将阳性结果的研究投稿于英文杂志, 而对阴性结果的研究则很可能束之高阁, 或者出现在一些非正规发表的资料中。语言是影响查寻偏倚的一个重要因素。只利用英文文献进行系统评价, 其结果极有可能发生偏倚。例如进行中医药研究的系统评价, 研究者不应把分析对象仅限于中文文献, 日、韩及其它许多国家和地区的相关文献也应纳入分析对象。应尽

可能查找多个国家、不同语种的文献数据库才能最大程度减小语言偏倚、数据库收录偏倚。

条目 5 评价论文是否提供了纳入和排除的研究文献清单。一般来说, 纳入研究文献都会出现在参考文献中, 但排除文献往往不会出现在发表的论文中。Cochrane 系统评价一般在参考文献列表中同时标注纳入和排除标准, 使得证据使用者很容易判断选择文献的质量。而在一般学术刊物上, 尤其是国内的期刊杂志, 由于版面的限制, 编辑不太可能把较为冗长的排除文献列于论文后, 因此, 笔者认为杂志编辑或同行评议者应在论文审稿阶段向作者索要排除文献的清单, 在论文发表前进行细致的评判。

条目 6 要求说明纳入研究的特征。该条目实际是对结果部分的研究特征情况的评价。报告者应当描述每个试验的特征, 如年龄、样本量、干预、剂量、疗程、随访期限。这些信息既是研究者自身判断临床异质性, 梳理受试者信息的重要手段, 也是使用者据此评价证据实用性的重要依据, 而且很多具体的临床指导信息都来源于这些试验特征的详细描述, 如针对什么样的患者, 采取什么样的剂量, 如何控制疗程等方面。这一部分最好用描述性语言和一览表式表格相结合的方法进行展示。

条目 7 指出研究者应当评价和报道纳入研究的科学性。首先要交代运用了怎样的评价方法评估纳入文献的真实性。绝大多数研究采用 Cochrane 评价员手册系列标准和 / 或 Jadad 量表, 也有一部分研究采用其他标准。有研究表明, 迄今为止有 9 种清单、60 余种量表用于评价随机对照试验的质量, 分别有 3 ~ 57 个条目, 需要 10 ~ 45 分钟完成, 但目前尚无评价随机对照试验真实性的“金标准”工具。但是不论采用何种工具, 研究者需要描述评价工具的出处、内容和执行方法。然后要在结果部分交代纳入文献的真实性。研究者可以采用文字叙述纳入文献的真实性, 但是最好结合表格或图的形式。

条目 8 要求纳入研究的科学性应该恰当地运用在结论的推导上。该条目主要运用在讨论或结论部分。循证医学要求客观、谨慎地叙述研究结果并做出适当的推论, 而不是急于做出具有倾向性质的判定。此阶段最重要的是充分考虑纳入研究质量, 指出可能的方法学缺陷, 结合定性 / 定量结果有针对性地阐述重要结局指标的临床意义及其价值。例如在证据强度评价时, 研究者需要考虑纳入研究的质量、效应量大小、各纳入研究之间的效应量是否一致、有无量效关系、有无支持此推论的间接证据、其他混杂因素等。

条目 9 是判定合成纳入研究方法是否恰当。一个完整的数据合并方案应该包括三方面的内容: ① 同质性检验, 这是数据合并的先决条件。它是对不同原始研究之间结果变异程度进行检验。如果检验结果有显著性差异, 应该解释其可能的原因并考虑进行结果合并是否恰当, 具有临床异质性的数据是不适合合成的。② Meta 分析, 应根据资料类型及评价目的选择效应量和统计分析方法。③ 必要时进行敏感性分析, 以判定结果的稳定性和强度。数据合并的正确与否直接关系到研究结果的真实性。系统评价 / Meta 分析的目的就是对研究资料进行综合分析, 得到一个全面综合的结论, 以指导决策或促进临床实践。如果合并方法得当, 数据的合成将为从这些资料中得出有意义的结论提供有力的工具, 而合并错误则只能误导使用者。因此, 研究者应该在这一部分描述主要效应测量指标、合并结果的方法 (统计学检验与可信区间)、缺失资料的处理、异质性评价、敏感性分析、亚组分析等情况。

条目 10 要求研究者在论文中交代是否评估了发表偏倚的可能性。对文献进行系统评价 / Meta 分析的过程中, 由于杂志编辑、作者、研究资金及研究结果等各方面的原因, 系统评价难免受到发表偏倚的影响。众所周知, 并非所有进行的研究都可以发表, 往往那些认为干预措施与传统治疗有明显差别的研究 (阳性结果的研究) 很容易发表, 而阴性结果的研究论文, 即研究得出干预措施与传统治疗差异不大时, 通常会被忽略, 这种由研究结果的方向性和强度来决定研究论文是否被发表, 就会导致发表性偏倚。评估发表偏倚有两个途径, 一个是寻找可能产生发表偏倚的原因。这些原因可以来自于杂志编辑、作者、文化背景和政治因素、研究资金来源、各国文献收录标准的差异等。另一个是系统评价者需要用一定的方法来判断此种偏倚的程度。其方法包括图像法 (漏斗图和修剪填充法) 和分析法 (失安全数、敏感度分析和预先注册程序)。

条目 11 要求评价者说明相关利益冲突, 清楚地交待系统评价及纳入研究中潜在的资助来源。这一点在药物系统评价 / Meta 分析中尤为重要。医药公司资助的研究项目比政府或其它学术团体资助的项目, 其发表阴性结果的百分率要低。资助者的经济利益冲突对医学临床试验可能会产生消极影响。如果评价者交代了此项研究的资助来源, 那么证据使用者就会依次分析资助者在临床试验中的潜在利益冲突, 考虑这些利益冲突是否会对研究的设计、执行和报告产生影响, 以及是否有高估研究结果的

可能。

4 运用现状

AMSTAR 自问世以来, 在国外得到了较为广泛的运用。目前运用较多的领域有同类系统评价研究、测量工具方法学研究、卫生技术评估以及被专业期刊编辑作为推荐工具等。

AMSTAR 的评价目的是评价系统评价 /Meta 分析的方法学质量。正在兴起的同类系统评价 (overviews of reviews 或 umbrella reviews), 或系统评价的系统评价、荟萃分析的荟萃评估是针对同一疾病的治疗或同一健康问题的干预措施, 全面总结当前多个相关系统评价并对其进行综合性评价的一种方法^[21]。简而言之, 它是系统评价的再一次综合和荟萃。完成一个同类系统评价的重要步骤就是对系统评价 /Meta 分析进行合理的方法学质量评价, 这在第 5 版的 Cochrane 系统评价员手册中已有明确阐述。而 AMSTAR 在这类研究中得到了大量运用, 如 2009 年第 4 期发表的一项评估生物制剂对类风湿性关节炎的疗效就采用 AMSTAR 作为方法学评价工具^[22], 而且不少同类系统评价计划书也决定采用该量表。非 Cochrane 同类系统评价也有采用的, 如综合评价预防儿童受虐干预措施情况^[23]、评价心外科领域系统评价 /Meta 分析的方法学质量^[24]、评价先天瘫痪患儿上肢功能治疗方法^[25] 等研究。

为了进一步完善系统评价 /Meta 分析方法学质量的评价方法, 国外有不少研究者采用横向比较不同测量工具的检测效度^[26,27], 研究结果也同样肯定 AMSTAR 的科学性。国外一些医学杂志编辑也指出方法学测量工具的重要性, 如美国《内科学年鉴》的编辑在“同类系统评价: 科学和艺术的精炼”一文中阐述同类系统评价的现实意义和科学价值, 认为 AMSTAR 是有效的测量工具之一^[28]。也有一些卫生技术评估报告采用 AMSTAR 作为评价干预性系统评价 /Meta 分析的方法学工具, 如美国华盛顿卫生局发布的一项评估神经电刺激治疗疼痛的技术报告^[29]。

需要指出的是, AMSTAR 的评价对象是系统评价 /Meta 分析的方法学质量, 这与专门用于评价报告质量的标准 QUOROM 及其更新版 PRISMA (Preferred Reporting Items for Systematic Reviews and Meta Analysis) 声明^[31] 有所不同。虽然报告质量和方法学质量存在一定的联系, 但也有差别。报告质量好的研究不一定方法学正确, 报告质量不好的研究也可能具有较好的真实性, 但是低下的报告质量

将影响结果的实用性。方法学质量越高, 研究的可重复性就越好, 其论证强度越高, 结果也越可靠。

5 研究展望

就目前的运用状况来看, AMSTAR 是具有良好的效度和信度的一种评价量表。当然一份优秀的量表不是短时间能够成熟和完善起来的, 需要在实践中不断提升。正如研发小组的专家所说: 通过广泛的使用, 将进一步考验 AMSTAR 的效度、信度和反应度^[19]。研发组也希望进一步扩展 AMSTAR 的应用范围, 不仅限于干预性系统评价 /Meta 分析的方法学质量评价, 也可考虑在病因学、诊断学以及预后学等方面研究的评价运用。

6 小结

据统计, 全球每年新发表系统评价约 2 500 个^[30]。面对如此庞大的阅读来源, 没有一系列切实高效的批判性选择方法是难以想象的。方法学质量评价工具是确保系统评价 /Meta 分析的生产者严格设计、实施评价过程的重要指南, 也是其使用者谨慎评价研究结果, 科学理性采纳有价值证据信息的重要前提。AMSTAR 是一个值得推荐使用的评价工具, 它不仅继承了早期开发出的具有代表性测量工具的优点, 而且还吸收了方法学研究领域的实证性证据, 通过执行严格的编制和考验程序而形成的。随着系统评价 /Meta 分析不断发展, 该标准也将会定期更新, 希望大家在使用过程中不断提出宝贵意见, 使 AMSTAR 日臻完善。

参 考 文 献

- 1 李静, 李幼平. 不断完善与发展的Cochrane系统评价. 中国循证医学杂志, 2008, 8(9): 742-743.
- 2 Porta M. Chief editor. A dictionary of epidemiology. Fifth Edition. New York: Oxford University Press, 2008: 217.
- 3 蔡羽嘉, 陈耀龙, 王梦书, 等. 循证医学术语介绍-VI. 中国循证医学杂志, 2009, 9(9): 942-945.
- 4 刘建平, 主编. 循证中医药研究方法. 第1版. 北京: 人民卫生出版社, 2009: 298-299.
- 5 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. 中国中西医结合杂志2007, 21(4): 306-310.
- 6 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.
- 7 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.
- 8 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.
- 9 Shea B, Dubé C, Moher D. Assessing the quality of reports of sys-

- tematic reviews: the QUOROM statement compared to other tools. In: *Systematic Reviews in Health Care: Meta-analysis in context*. Edited by: Egger M, Smith GD, Altman DG. London: BMJ books, 2001: 122-139.
- 10 Oxman AD. Checklists for review articles. *BMJ*, 1994, 309(6955): 648-651.
 - 11 Sacks H, Berrier J, Reitman D, *et al*. Meta-analyses of randomized controlled trials. *N Engl J Med*, 1987, 316(8): 450-455.
 - 12 Moher D, Cook DJ, Eastwood S, *et al*. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. *Lancet*, 1999; 354(9193): 1896-1900.
 - 13 张俊华, 商洪才, 张伯礼. 系统评价和Meta分析质量的评价方法. *中西医结合学报*, 2008, 6(4): 337-339.
 - 14 Sterne JAC, Gavaghan D, Egger M. Publication and related bias in meta-analysis power of statistical tests and prevalence in the literature. *Journal of Clinical Epidemiology*, 2000, 53(11): 1119-1129.
 - 15 Shea BJ, Grimshaw JM, Wells GA, *et al*. Development of AMSTAR: A Measurement Tool to Assess Systematic Reviews. *BMC Medical Research Methodology*, 2007, 7(10): 1-10. e7-8.
 - 16 The Cochrane Library. Volume 3. Chichester, UK: John Wiley & Sons Ltd; 2004.
 - 17 Egger M, Zellweger-Zahner T, Schneider M, *et al*. Language bias in randomised controlled trials published in English and German. *Lancet*, 1997, 351(9095): 326-329.
 - 18 Moher D, Pham B, Klassen T, *et al*. What contributions do languages other than English make to the results of meta-analyses? *Journal of Clinical Epidemiology*, 2000, 53(9): 964-972.
 - 19 Shea BJ, Bouter LM, Peterson J, *et al*. External Validation of a Measurement Tool to Assess Systematic Reviews (AMSTAR). *PLoS ONE*, 2007, 2(12): 1-4. e1350.
 - 20 Canadian Agency for Drugs and Technologies in Health. Available at: <http://www.cadth.ca>.
 - 21 Becker LA, Oxman AD. Chapter 22: Overviews of reviews. In: Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.0.1 [updated September 2008]. The Cochrane Collaboration, 2008. Available at: www.cochrane-handbook.org.
 - 22 Singh JA, Christensen R, Wells GA, *et al*. Biologics for rheumatoid arthritis: an overview of Cochrane reviews. *Cochrane Database of Systematic Reviews*, 2009, Issue 4. Art. No.: CD007848.
 - 23 Miktona C, Butchart A. Child maltreatment prevention: a systematic review of reviews. *Bull World Health Organ*, 2009, 87(5): 353-361.
 - 24 Panesar SS, Rao C, Joshua A. *et al*. Development of the Veritas plot and its application in cardiac surgery: an evidence-synthesis graphic tool for the clinician to assess multiple meta-analyses reporting on a common outcome. *Can J Surg*, 2009, 52(5): 137-142.
 - 25 Sakzewski L, Ziviani J, Boyd R. Systematic Review and Meta-analysis of Therapeutic Management of Upper-Limb Dysfunction in Children With Congenital Hemiplegia. *Pediatrics*, 2009, 123(4): 1111-1122.
 - 26 Mookink LB, Terwee CB, Stratford PW, *et al*. Evaluation of the methodological quality of systematic reviews of health status measurement instruments. *Quality of Life Research*, 2009, 18(3): 313-333.
 - 27 Bessa-Nogueira RV, Vasconcelos CE, Niederman R. The methodological quality of systematic reviews comparing temporomandibular joint disorder surgical and non-surgical treatment. *BMC Oral Health*, 2008, 8(10): 27.
 - 28 Cornell JE, Laine C. The Science and Art of Deduction: Complex Systematic Overviews. *Ann Intern Med*, 2008, 148(10): 786-782.
 - 29 HTA final report: Electrical Nerve Stimulation for Treatment of Pain. Octo.7, 2009. Available at: <http://www.hta.hca.wa.gov>.
 - 30 Moher D, Tetzlaff J, Tricco AC, *et al*. Epidemiology and reporting characteristics of systematic reviews. *PLoS Med*, 2007, 4(3): 78.
 - 31 Moher D, Liberat A, Tetzlaff J, 等. 系统综述和荟萃分析优先报告的条目: PRISMA声明. *中西医结合学报*, 2009, 7(9): 889-896.

收稿日期: 2010-03-03 修回日期: 2011-07-02
 本文编辑: 杜亮