

知证卫生决策工具之八 ——确定系统评价的可信度*

SUPPORT Tools for evidence-informed health Policymaking (STP) 8: Deciding how much confidence to place in a systematic review*

Simon Lewin¹, Andrew D Oxman², John N Lavis³, Atle Fretheim⁴

1. Norwegian Knowledge Centre for the Health Services, P.O. Box 7004, St. Olavs plass, N-0130 Oslo, Norway; Health Systems Research Unit, Medical Research Council of South Africa;

2. Norwegian Knowledge Centre for the Health Services, P.O. Box 7004, St. Olavs plass, N-0130 Oslo, Norway;

3. Centre for Health Economics and Policy Analysis, Department of Clinical Epidemiology and Biostatistics, and Department of Political Science, McMaster University, 1200 Main St. West, HSC-2D3, Hamilton, ON, Canada, L8N 3Z5;

4. Norwegian Knowledge Centre for the Health Services, P.O. Box 7004, St. Olavs plass, N-0130 Oslo, Norway; Section for International Health, Institute of General Practice and Community Medicine, Faculty of Medicine, University of Oslo, Norway

摘要 健康干预效果系统评价的可靠性参差不齐。因此,政策制定者及支持者需评价此类证据的可信度。采用系统透明的评价过程可避免判断时出现误差和偏倚。在确定干预效果系统评价结果的可信度时,本文建议考虑以下五个问题:①这篇评价是否明确提出了一个恰当的决策或管理问题?②是否使用了恰当的标准去评价系统评价中的研究?③对相关研究的检索是否详尽且较全面?④对纳入研究与评价主题相关性及其偏倚风险所展开的评估是否具有可重复性?⑤不同研究之间的结果是否相似?

1 关于STP

本文是为卫生政策和卫生项目决策者及其决策支持者所写“知证卫生决策支持工具”系列文章中的一篇。该系列文章旨在帮助他们确保其所做的决策基于当前可及的最佳研究证据。SUPPORT工具及使用方法在本系列文章的导论中已有详述。每篇文后均附有本系列文章的所有术语(见附件1)。SUPPORT网站上可找到本系列论文西班牙文、葡萄牙文、法文和中文翻译的链接(www.support-collaboration.org)。欢迎通过邮箱 STP@nokc.no 提出修改意见

2 案例情景

情景1 你是一名高级公务员,将向部长递交一份关于为解决优选卫生问题而拟定的不同政策和项目方案以支持证据的提案。你很关心每一方案支持证据的系统评价的可信度,并想确保你的下级已对其进行了合理评价。

情景2 你在卫生部工作,正在草拟一份关于优

选卫生问题解决方案的文件。已找出数篇关于方案效果的系统评价,要求评价每篇系统评价的可信度。

情景3 假如你在支持卫生部知证决策的独立机构工作。正在为卫生部准备一份关于高优选问题的解决方案可能带来影响的文件,你想在评价每一方案影响效果的系统评价可信度上得到指导。

3 背景

对于决策者(情景1),本文列出了一些他们可能会要求下级在确定卫生保健干预效果系统评价结果可信度时应考虑的问题。

对于决策支持者(情景2和3),本文推荐了一些问题以指导其严格评价系统评价结果。

多个随机对照试验(RCT)的系统评价因其提供了卫生保健干预措施效果最可靠的证据而被广泛接受^[2,3]。其特点是识别、筛选、评价相关研究,纳入研究收集和分析数据的过程系统、明晰^[3]。系统评价也正越来越多地被用于识别、评价及合成干预措施经济学结果的证据^[4],如提高新生儿病房婴儿母乳喂养的成本效果^[5]或不同指南传播与实施策略的成本^[6]。系统评价也可用于从定性研究中归纳证据,如卫生服务接受者和提供者对卫生干预措施的观点^[7-10]。本文重点关注卫生保健政策或项目效果的系统评价,包括对卫生服务提供措施的系统评价,

* 本文英文原文发表于 *Health Research Policy and Systems*, 2009, 7(Suppl 1):S8 doi:10.1186/1478-4505-7-S1-S8

作者简介: Simon Lewin, Email: simon.lewin@nokc.no

如在初级医疗保健中用护士取代医生的效果^[11],及改革策略的效果,如为卫生工作者举办继续教育会议的效果^[12]。

系统评价采用系统而明晰的方法旨在降低因偶然因素导致的偏倚和误差风险,有助于严格评价这些合成证据^[13,14]。但生产系统评价时的严谨程度不一,因此并非所有系统评价都同等可靠——即,各系统评价结果我们赋予的可信度不同。在利用研究结果支持决策时,仅仅依赖被本文称为“系统评价”(或 Meta 分析)的评价是不够的。

当利用有关效果的系统评价结果支持决策时,决策者及其他人需要判断这些证据的可信度。系统、透明的评价证据的方法可避免在判断中引入误差和偏倚。系统、透明的评价证据方法还可使其他利益相关者,包括公众,理解和评价这些判断。其在这种评价对临床干预或服务推荐和决策的影响方面^[15],或实施或停止项目或政策的决策方面尤其重要。图 1 是查找和评价系统评价支持知证决策的步骤。

系统评价结果的可信度可能会受限于很多因素,包括:

- 实施评价前未(如在公开发表的系统评价计划书上)明确评价的问题和方法,
- 未明确研究的纳入和排除标准
- 未充分描述纳入研究
- 未评价纳入研究的偏倚风险
- 未评价发表偏倚风险,如一些研究,尤其是阳性有统计学意义结果的研究,比其它研究更易被发表从而增加了其被纳入评价的可能性
- 未使用恰当的方法综合相关研究的结果 (Meta 分析)
- 未充分检查纳入研究之间结果的差异(如结果的“异质性”)
- 未基于纳入数据作出评价结论

系统评价潜在的其他局限性包括利益冲突(利益冲突能够以上述任一种方式影响评价的可信度)和过时的评价。

如某研究从方法学和报道内容方面比较 Cochrane 系统评价与纸质版杂志上发表的系统评价,发现两者可信度不同。Cochrane 系统评价的报告内容更能降低偏倚风险,得益于其清楚地描述了纳入和排除标准及严格评价每个纳入研究的偏倚风险^[16]。类似的,另一个研究比较了关于同一药物的 Cochrane 系统评价和药厂资助系统评价的方法学质量和结论。Cochrane 系统评价质量评分更高。这是因为与药厂资助的系统评价相比, Cochrane 系

统评价更多考虑潜在偏倚。而药厂资助的系统评价更倾向于尚未确定的研究结果^[17]。其他一些对系统评价的研究也报告两者在质量和结论方面存在的差异^[18-21]。

现已设计出许多用于评价系统评价质量的工具,包括 AMSTAR (A Measurement Tool to Assess Reviews)^[22]、CASP (Critical Appraisal Skills Programme)^[23] 和由 Oxman 和 Guyatt 研发的评价工具^[24-26],但都包含了相似的标准(AMSTAR 工具介绍见表 1)。一些工具还采用评分量表对系统评价可信度打分。如今越来越多的系统评价包含了这种可信度的评价。一般来说,等级评分高代表系统评价结果的可信度更高。相反,等级评分低代表系统评价结果的可信度更低,此时应仔细分析该系统评价以确定其主要局限性。但需强调三点:① 总分或总级别不一定表明系统评价在某些方面的可信度就高——一些方面比其它方面可信度更高;② 评分过程本身也涉及给评价工具中的不同项目分配权重。很难说明哪一项目应被分配更大权重^[27]。③ 分级工具仅能评价报告内容的可信度。若未报告系统评价使用方法的关键信息,则不清楚做了什么或做到哪种程度,这将是一个重大的局限性。

评价系统评价结果可信度应与对系统评价自身结果的认识区分开来。表 2 提供了有关效果的系统评价结果应看何种指南。表 3 提供评价定性研究的系统评价和经济学研究系统评价可信度的指南。

对系统评价结果可信度的评价还应与任何与特定政策问题相关的评价区分开来。如对相关性问题的考虑包括:该系统评价是否提供了不同决策或项目效果的证据,该系统评价结果是否适用于拟实施决策的环境。系统评价结果适用性的评价过程将在该系列文章的文 9 中进一步讨论^[28]。

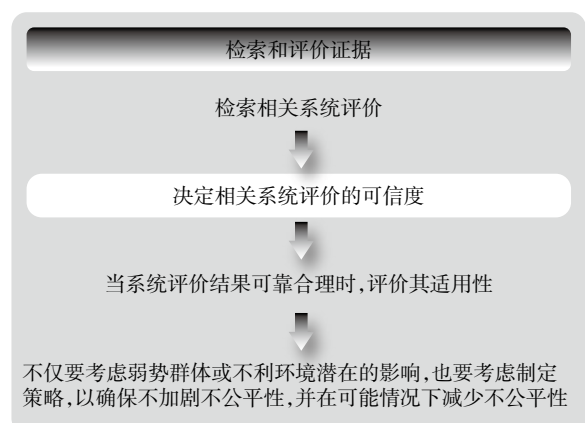


图 1 查找和评价系统评价支持政策和项目方案的决策

表 1 AMSTAR-评价系统评价的测量工具^[22]

1. 是否提供“先验”假设? 研究问题纳入标准应在系统评价前提出	<input type="checkbox"/> 是 <input type="checkbox"/> 否 <input type="checkbox"/> 不能回答 <input type="checkbox"/> 不适用
2. 是否进行双重数据筛选和提取? 至少应有两名独立的数据提取者,并说明在出现分歧时达成共识的方法	<input type="checkbox"/> 是 <input type="checkbox"/> 否 <input type="checkbox"/> 不能回答 <input type="checkbox"/> 不适用
3. 文献检索是否全面? 至少检索两个电子资源。至少应报告检索年限与检索数据库(如 Central,EMBASE 及 MEDLINE)。必须说明检索的关键词和主题词,若可行应提供检索策略。同时应补充检索当前的内容、综述、教科书、专题数据库、特定研究领域的专家意见及追溯已获研究的参考文献	<input type="checkbox"/> 是 <input type="checkbox"/> 否 <input type="checkbox"/> 不能回答 <input type="checkbox"/> 不适用
4. 研究纳入标准是否考虑了发表类型(如灰色文献)? 作者应声明检索研究不受其发表类型的限制。作者应声明是否因发表情况、语言等(从系统评价中)排除任何研究。	<input type="checkbox"/> 是 <input type="checkbox"/> 否 <input type="checkbox"/> 不能回答 <input type="checkbox"/> 不适用
5. 是否提供了纳入和排除的研究清单? 应该提供纳入和排除研究的列表	<input type="checkbox"/> 是 <input type="checkbox"/> 否 <input type="checkbox"/> 不能回答 <input type="checkbox"/> 不适用
6. 是否提供了纳入研究的特征? 应以汇总形式,如表格形式,列出原始研究的数据如受试者、干预措施类型及结果指标。应报告所有的分析研究特征范围,包括年龄、种族、性别、相关社会经济数据、疾病状况、持续时间、严重性及其他疾病	<input type="checkbox"/> 是 <input type="checkbox"/> 否 <input type="checkbox"/> 不能回答 <input type="checkbox"/> 不适用
7. 是否评价、记录了纳入研究的科学质量? 应提供“先验”评价方法(如作者仅选择随机双盲安慰剂对照或分配隐藏作为纳入标准)。对其他类型的研究,则应考虑其它相应的评价方法	<input type="checkbox"/> 是 <input type="checkbox"/> 否 <input type="checkbox"/> 不能回答 <input type="checkbox"/> 不适用
8. 作结论时是否恰当地使用了纳入研究的科学质量? 当撰写推荐意见时,在系统评价的分析和结论中应明确阐述研究方法的严谨程度与科学质量	<input type="checkbox"/> 是 <input type="checkbox"/> 否 <input type="checkbox"/> 不能回答 <input type="checkbox"/> 不适用
9. 综合研究方法的方法是否恰当? 对于综合结果,应有检验结果保证研究结果之间可以综合并评价其同质性(如采用同质性 χ^2 检验, I^2)。如果存在异质性,应考虑随机效应模型和/或临床合理性综合结果(综合结果是否合适?)。	<input type="checkbox"/> 是 <input type="checkbox"/> 否 <input type="checkbox"/> 不能回答 <input type="checkbox"/> 不适用
10. 是否评价发表偏倚的可能性? 综合采用图形辅助(如漏斗图及其他可行的方法)和/或统计学检验(Egger 回归)评价发表偏倚	<input type="checkbox"/> 是 <input type="checkbox"/> 否 <input type="checkbox"/> 不能回答 <input type="checkbox"/> 不适用
11. 是否声明了利益冲突? 应清楚说明系统评价和纳入研究的中潜在资助者	<input type="checkbox"/> 是 <input type="checkbox"/> 否 <input type="checkbox"/> 不能回答 <input type="checkbox"/> 不适用

表 2 解释关于效果的系统评价结果

以下问题有助于指导决策者解释关于效果的系统评价结果^[33,47,48]:

- 效应的估计值大小? 许多系统评价提供纳入研究的平均效应值。通常以 OR、RR 及 SMD 表示
- 纳入研究的平均效应估计值是否合理? 系统评价通过统计学方法总结、综合纳入研究的结果数据。为保证综合结果数据的合理性,需要考虑纳入研究在人群、干预措施、对照方法及结局测量方面是否有足够的相似性。若不可能得到平均效应估计值时,应对所获数据作描述性评价
- 是否报告了效应估计值的可信度? 系统评价应报告平均效应估计值的可信区间。可信区间越宽,我们能确信的效应的真实有效性越差
- 如果报告了亚组分析结果,其是否合理? 一篇系统评价可能会报道不同受试者亚组和不同研究亚组的结果^[49]。如一项对降低 5 岁以下儿童腹泻的干预措施的系统评价,可能也会考虑 1 岁以下儿童的干预效果。同样,系统评价可能包括判断为低偏倚风险研究的亚组分析。亚组分析应有助于阐明整个系统评价的问题和提供有关可能影响或调整干预措施效果等因素的信息。如更强的干预措施可能会产生更大的效应。应在进行系统评价前设计亚组分析,且亚组分析结果的可信度较低。这是因为其可信度低于基于所有纳入研究的分析,且多重统计分析会产生由机遇所致的阳性结果
- 是否将“有关效果缺乏证据”解释为“证明无效的证据”? “缺乏效果有关证据”不等于“证明无效的证据”。前者指尚无足够的可获证据得出干预措施有无效果的结论。后者指从纳入研究中得到清楚的证据证明干预措施无期望的效果^[50]。
- 系统评价的结论和推荐意见是否从最初的问题和系统评价证据得出? 考虑结论是否由评价者直接基于数据得出而不是超越此证据得出非常重要
- 该证据是否适用于正在考虑的决策问题? 卫生系统之间的差别意味着在某环境实施良好的项目和干预措施不一定在其他环境中同样有效。决策者需要评价来自系统评价的研究证据是否适用于他们的条件。相关指南见文 9^[28]

* 这里列出的问题与系统评价可信度评价有一些重复。因为在评价和理解系统评价结果时可信度是非常重要的一个部分

表 3 评价定性研究和经济学研究系统评价的可信度

定性研究的系统评价与日俱增。其使用的方法很多,包括描述性证据合成、Meta 人类学分析及实况研究。定性研究的系统评价不但可提供重要信息,还可为有关效果的系统评价提供线索和作其补充^[51,52]。但对读者而言,对这些系统评价的可信度评价非常重要。但迄今却鲜有用于此特定目的评价工具。许多用于指导决策者评价效果系统评价可信度的问题对评价定性研究的可信度同样有用。这些问题包括:

1. 该系统评价是否提出了一个恰当的决策或管理问题? 系统评价的问题应该可以通过定性数据得到解决且这些问题应与决策者相关。定性研究的系统评价能提供利益相关者对健康和卫生保健方面的观点与经验,这将有助于明确问题^[39]。定性研究的系统评价也能提供方案如何和为何起作用(如通过对实施中的政策和项目作实时过程评估)及利益相关者对方案的看法及其相关经验^[40,53]。
2. 用于筛选研究的标准是否恰当? 对怎样筛选纳入研究的描述应与研究的问题合理相关。
3. 是否对使用的检索方法给出了明晰和恰当的解释? 有些定性研究的系统评价进行了全面的文献检索,而另一些可能使用抽样方法。应清楚描述并评价所选方法。
4. 评价纳入研究可信度的方法是否恰当? 应描述如何考虑纳入研究的可信度。
5. 分析纳入研究结果的方法是否恰当? 应使用可接受的证据合成方法并描述选择该方法的合理性。

当评价经济学研究系统评价的可信度时应考虑以下问题^[54]:

1. 重要的相关研究是否有遗漏?
2. 所选文章的纳入标准是否恰当?
3. 对纳入研究的系统评价是否可重复?
4. 纳入研究的设计方法和主题大体上可比吗?
5. 所有结果的可重复性如何?
6. 结果是否有助于卫生保健资源配置?

在评价有关方案效果的系统评价结果的可信度时,我们建议考虑以下五个问题:

4 思考问题

以下问题可在决策者确定方案效果系统评价结果的可信度时为其提供指导:

- (1) 该评价是否明确提出了一个恰当的决策或管理问题?
- (2) 该系统评价纳入原始研究时,是否使用了恰当的标准?
- (3) 对相关研究的检索是否详尽且较全面?
- (4) 对纳入研究与系统评价主题的相关性及其偏倚风险的评价是否可重复?
- (5) 不同研究间的结果是否相似?

4.1 该评价是否明确提出了一个恰当的决策或管理问题?

评价系统评价结果可信度的关键的第一步是检查所提出的问题。一个设计和制作都堪称一流的系统评价,如果系统评价结果不能解决一个敏感、恰当并与决策者考虑的问题相关的政策和管理问题,则其不可能有助于决策。

一个恰当的政策或管理问题应该:

- 明确: 即应对问题进行详细阐述而不是暗含在呈递的材料里。如果没有明确或清晰地阐述系统评价的问题,则很难充分评价系统评价的制作过程。这是因为制作系统评价必须至少部分与其提出的问题相关^[29]。如评价一篇系统评价的研究纳入标准是否恰当,需要评价纳入研究是否与回答欲解决的问题有

关。一个清楚的问题还可帮助读者去判断该系统评价是否与自己的工作有关^[29]。

- 事先: 即应在进行系统评价之前确定。在进行系统评价甚至撰写系统评价计划书之前明确提出问题很重要。如要制作 Cochrane 系统评价都必须事先发表系统评价计划书。这些计划书的例子可参见 Cochrane 图书馆 (<http://www3.interscience.wiley.com/cgi-bin/mrwhome/106568753/HOME>)。如果在开始进行系统评价前未能使评价问题具体化,则将面临如下风险: 根据已获证据来改动问题继而降低系统评价结果的可信度。
- 解决与决策和管理相关的问题: 这需在特定环境下,且基于问题的范围(在特定时间和权限内的重要问题)进行评价,否则与系统评价的问题不相关:
 - 范围太窄: 如某个系统评价仅考虑在特定环境下,对特定年龄组受试者的项目效果,或仅局限于得到有限的项目结果。这种情况下,该项目的结果将不能外推至其他人群、环境或结果。
 - 范围太宽: 如一篇系统评价可能基于一个很广泛的实践范围确定一个项目,而这些实践并非都与确定项目相关,或可能提出一个很宽泛但对决策者无用的问题。如“护士能否有效提供健康促进项目?”这样的问题,是否特定的骨干护士群体(如注册护士)能有效针对一个特定卫生问题,如 HIV/ 艾滋病预防。

- 没有设立恰当的对照组: 如一个项目应该与未实施该项目比较, 而不是与当今最好的处理方案比较。

构建一个好的系统评价问题需考虑以下几点: 该评价应覆盖的人群与环境的类型(居住在疟疾流行地区的1月~6岁儿童); 项目设计与对照的类型[定期给予抗疟药治疗(干预措施)与安慰剂或非药物治疗比较(对照组)]; 关注的结局类型(临床疟疾和恶性贫血)^[30,31]。有时可用缩略词PICO (Population, Intervention, Comparison, Outcome)来总结问题的4个关键部分。

显然, 一个系统评价问题需要很好地构建, 但仍有许多叙述性系统评价没做到这点。一篇对发表在主要医学期刊上的此类系统评价的抽样调查显示: 20% 未能清楚陈述系统评价的目的。

4.2 该系统评价纳入原始研究时, 是否使用了恰当的标准?

一篇系统评价的纳入和排除标准应详细列出该系统评价关注的人群类型、干预措施、对照及结局。这些标准特别列于系统评价的计划书中, 将决定该系统评价纳入哪些研究, 它们将强烈影响评价结果。纳入/排除标准对评价问题至关重要。

当考虑一篇评价中使用的纳入/排除标准是否恰当时, 需要核实以下问题:

- 该系统评价有无明确的纳入和排除标准? 这些标准对避免在系统评价中与纳入研究相关的偏倚非常重要。如最近一篇对外科系统评价方法学的质量评价发现, 仅70%的系统评价报道了其纳入研究的标准^[18]
- 纳入和排除标准是否明确与以下几方面相关: 目标人群、干预措施和对照及结局类型
- 纳入和排除标准是否与研究问题吻合^[33]? 如一篇系统评价旨在评价抗疟药物预防与间歇治疗对疟疾流行地区儿童疟疾的作用, 纳入标准是否明确要求纳入研究中儿童来自相似的环境, 是否明确规定预防与治疗的形式^[31]? 同样, 如果系统评价旨在关注某干预措施能否提高农村或低服务水平地区专业卫生人员的比例, 纳入标准中是否明确指出了专业卫生人员纳入的范围及其教育或经济干预措施的类型^[34]?

4.3 对相关研究的检索是否详尽且较全面?

系统评价的一个重要方面是对符合纳入标准的研究展开全面检索且该检索具有可重复性。这也是系统评价与一般描述性综述的不同之处。系统检索

可确保已纳入所有相关证据而将偏倚降至最小。因而有助于评价决策或项目的效果^[35]。

发表偏倚即根据研究结果指向和强度选择性发表文章^[36], 是偏倚被引入系统评价的途径之一。最近1篇系统评价检验了随机试验文章的发表受是否得出阳性结果及当前可认识到的试验结果的重要性的影响程度。结果显示, 阳性结果的试验比阴性结果试验更易发表^[37]。该系统评价和其他研究还发现报告阳性结果比报告阴性结果的试验能更快地发表^[38]。因此, 除非对已发表和未发表的研究都尝试进行查找, 否则系统评价可能会高估干预措施的阳性效果。

系统评价检索的全面程度不一。如对已发表有关哮喘治疗系统评价进行再评价, 结果显示: 纳入的33篇系统评价中, 仅52%对疗效相关证据进行了合理全面的检索^[20]。因此审查对相关研究的检索情况非常重要。

当考虑对相关研究的检索是否详尽且合理全面时, 需审查以下几个方面:

- 该系统评价是否详细描述了检索相关研究的策略? 包括: ① 检索了哪些资源, ② 检索这些资源时使用的关键词(何处使用), ③ 各资源的检索年限。表4列出了cochrane图书馆系统评价的检索资源范围
- 检索策略是否包括了发表研究的电子数据库? 可获得的发表研究的电子数据库很多, 其中一些数据库还可免费或低价检索。主要的数据库包括PubMed/MEDLINE(美国国立医学图书馆建立), the Cochrane Central Register of Controlled Trials (CENTRAL-Cochrane 协作网建立), 地区性数据库如LILACS(拉丁美和加勒比海卫生科学)。本系列文4^[39]和文5^[40]提供更多查找相关研究文献的信息
- 除电子数据库外, 是否进行了补充检索? 这些补充可以是追溯相关研究的参考文献、与该领域的作者和专家联系及查询与系统评价主题领域相关研究的专题注册库。这种补充检索有助于找到更多已发表和未发表的研究(可能包含从“灰色”文献中获取的研究, 即来源于非收录和经同行评审杂志的文献资源)
- 检索是否更新? 该评价是否注明了检索覆盖的时间范围, 是最近的检索吗? 一篇已发表的与政策问题相关的系统评价已经过时。因此, 如果一个系统评价未能纳入当前所有最

表 4 系统评价检索资源举例

系统评价类型	检索资源
卫生系统系统评价举例: 初级医疗机构和社区卫生保健中非专业卫生人员干预的系统评价 ^[44]	<ol style="list-style-type: none"> 发表研究的电子数据库: <ul style="list-style-type: none"> MEDLINE Cochrane Central Register of Controlled Trials (CENTRAL) and specialised cochrane Register (EPOC and Consumers and Communication Review Groups) Science Citations EMBASE CINAHL (Cumulative Index to Nursing and Allied Health Literature) Healthstar AMED (Allied and Complementary Medicine Database) Leeds Health Education Effectiveness Database 被纳入评价的研究数据库书目 请所有联系的作者提供研究细节
公共卫生系统评价举例: 男性包皮环切术预防向异性传染HIV的系统评价 ^[55]	<ol style="list-style-type: none"> 发表研究的电子数据库 <ul style="list-style-type: none"> MEDLINE EMBASE Cochrane Central Register of Controlled Trials (CENTRAL) 会议摘要电子数据库 <ul style="list-style-type: none"> AIDSearch Conference databases 在研试验电子数据库 <ul style="list-style-type: none"> ClinicalTrials.gov Current Controlled Trails 联系该领域的研究者和相关组织 查阅通过上述方法获得研究的参考文献, 核查在检索过程中检出的每一篇系统评价、Meta 分析及预防指南
临床系统评价举例: 他汀类药物预防老年痴呆的系统评价	<ol style="list-style-type: none"> 电子数据库 <ul style="list-style-type: none"> The specialized Register of the Cochrane Demential and Cognitive Improvement Group Cochrane Central Register of Controlled Trials (CENTRAL) MEDLINE EMBASE PsycINFO (心理学文献数据库) CINAHL SIGLE (欧洲灰色文献) LILACS (Latin American and Database Health Science Literature) 会议摘要电子数据库 <ul style="list-style-type: none"> ISTP (Index to Scientific and Technical Proceedings) INSIDE (British Library Database of Conference Proceedings and Journals) 专题数据库 <ul style="list-style-type: none"> Index to Theses (formerly ASLIB) (United Kingdom and Ireland theses) Australian Digital Theses Program Canadian Theses and Dissertations DATAD-Database of African Theses and Dissertations Dissertation Abstract Online (USA) 在研试验数据库: 大范围检索相关数据库

相关的证据,它对政策和项目方案实施效果的评价将不可靠

4.4 对纳入研究与系统评价主题的相关性及其偏倚风险的评价是否可重复?

系统评价作者需要对每个可能纳入的原始研究作两个判断。首先,该研究是否满足本系统评价的纳入标准,即是否与该系统评价主题相关?其次,该研究结果的偏倚风险有多大?偏倚风险指“结果或结论中存在系统误差或偏离真值的风险”^[27]。同样涉及该研究结果的准确性问题^[27]。因为这些判断将影响系统评价的结果,因此以透明、可重复的方式呈现这些信息非常重要。他人需要了解这些判断是如何做出,且可以重复这些评价。

综上所述,系统评价需要制定明确的纳入和排除标准,以避免纳入研究过程中导致偏倚。这些标

准和判断通过影响纳入研究的筛选而必然影响到系统评价的结果。通过以下方法可以使这些判断的偏倚和误差最小化:第一,应由两名评价员独立筛选纳入研究,如意见不一致时,可通过与其他评价员讨论解决分歧。第二,纳入研究的理由(和排除看似相关研究的理由)应在发表的系统评价里注明。这样做可让读者对研究的纳入或排除研究与否做出自己的判断,也为该系统评价提供透明的“可寻之迹”,以保证过程可以重复。

系统评价得出关于政策或项目实施效果的结论还受从每个纳入研究获得数据真实性的影响。如果纳入的每个研究数据的真实性很低,汇总研究结果或小结将可能得出误导性结果。因此,评价纳入研究结果的偏倚风险是系统评价的重要部分。对偏倚风险的评价应写入系统评价的解释和结果中^[27]。

对随机试验已有许多不同评价偏倚风险或质量的方法^[27,41,42],我们在此不讨论这些不同的评价方法。值得注意的是,系统评价应该非常清楚可用的方法,并一丝不苟地按此方法作评价。

当评价纳入研究与系统评价主题相关性及潜在的偏倚风险时,需考虑以下问题:

- 是否使用明确而透明的方法评价纳入研究与系统评价主题的相关性? 系统评价应阐述怎样评价相关性,并提供纳入和排除研究的清单
- 是否使用明确而透明的方法评价纳入研究的偏倚风险? 系统评价应报道用于评价偏倚风险的工具,评价如何进行及评价结果
- 偏倚风险评价结果是否用于解释系统评价结果? 如当纳入研究的偏倚风险很大时系统评价结果的可信度较低。

4.5 不同研究间的结果是否相似?

针对一个项目的某个特殊结局,系统评价中纳入的研究结论可能很相似,也可能不同。系统评价纳入研究的变异通常被称为“异质性”^[27]。系统评价纳入研究间的变异程度部分取决于该系统评价的范围。当范围较宽时,预期纳入研究的范围可能较广,并因此导致其变异程度可能较大。相反,系统评价范围较窄时,纳入研究之间可能更相似。

如果一篇系统评价纳入研究的受试者、干预措施及结局变异很大,而干预效果又受这些因素的影响,则可能导致变异或异质性。因为这些研究间真实的干预效果不同,这种情况下平均效果则无用。

根据变异水平,系统评价可用不同方法总结纳入研究的信息,如:

- 计算研究间的平均或汇总效果:当纳入研究间的变异很低时,用此方法有用。如关于“早出院与家庭医院的项目对传统需要急诊住院护理的健康问题,代之以医疗机构在病人家里提供治疗)”发现,系统评价纳入研究足够相似,故可估算项目的平均效果。这篇系统评价结果发现没有足够证据证明“早出院与家庭医院”有经济或健康优势。
- 计算系统评价纳入研究亚组的平均效果:当纳入研究间的变异很高,不能用于亚组内研究变异性较小的研究。如一个初级和社区卫生保健中非专业卫生人员的系统评价依据非专业卫生人员所解决的卫生问题对纳入研究进行分组。有些组如促进计划免疫和母乳喂养组,可对相关研究计算平均效果。该系统

评价结果表明非专业卫生人员能改进计划免疫和母乳喂养^[44]。

- 描述效应量的范围:当研究之间不足够相似,不能计算平均效果时,仍可能描述研究间效果的范围。如评价监督与反馈医疗机构服务效果的系统评价结果表明:与预期值符合的范围从下降 16% 到增加 70%,中位数为 5%,提示监督与反馈可使医疗服务更有效,但其效果为低 - 中程度^[45]。
- 针对特定问题的干预类型分类:有些系统评价的范围很广,因此纳入研究的变异性很大,即定量综合纳入研究的结果甚至描述其效果的范围都不敏感。这时可用描述性评价。如针对降低卫生领域的不公平性,评价卫生服务干预效果的系统评价,纳入研究的项目可能仅在卫生系统实施,也可能与其他机构一起实施。纳入研究的范围很广,从改善血压控制到健康促进干预不可能进行统计学汇总分析^[46]。

当研究间结果不同时,需考虑以下问题:

- 对发现的各研究结果间的不同是否有令人信服的解释? 包括纳入研究的受试者、干预措施、对照组、结果、环境及试验时期不同。如有些研究纳入的受试者年龄范围更宽或试验前身体状况不同。
- 如何进行综合数据的定量评估,是否可能更能说明问题? 如果纳入的研究变异大,采用综合定量评估意义不大。进一步探索通过亚组分析作数据分析,但其结果可能不可靠。

随着可获系统评价数量增加,针对一个特殊决策问题找到不止 1 篇系统评价越来越常见。有时这些系统评价的结果或结论不同。表 5 向决策者提供了处理各种问题的指南。

5 结论

进行有关效果的系统评价的严格程度存在明显差异。因此,为了判断这些系统评价证据的可信度,在进行决策时评价系统评价的可信度非常重要。为此,应使用系统而透明的评价方法,且许多专门的评价工具也已开发出来。但这些工具只能用于评价系统评价报告的内容。这就是为什么任何使用这些评价工具进行的评价必须三思和谨慎。

当一篇系统评价的可信度很低时,决策者应降低对结果的可信度,决策时应谨慎使用这些结果(见图 2 的总结)。当根据一篇系统评价的证据做决定

表 5 当针对相同问题的系统评价得出不同结果时, 决策者该如何处理?

当为支持一项决策而寻找证据时, 找到不止一项相关系统评价并不少见。有时这些系统评价的结果可能不同, 这可能导致作者对同一干预效果得出不同的结论。此情况不同于两个或更多研究结果相同, 只是研究者或其他人对结果的解释不同的情况^[19]。不同系统评价结果不同的原因很多。其包括: 系统评价解决的问题; 所用的纳入/排除标准; 从纳入研究中提取的哪些; 如何评价研究质量; 所选的数据统计分析方法^[19]

由Jadad及其同事设计的下列问题有助于找出和解决不一致性^[19]:

- 不同的系统评价是否提出了同样的问题? 否则, 被选中的这篇系统评价提出的问题应与决策需要的证据最接近, 或者其评价的结局与决策问题最相关
- 如果不同系统评价解决相同的问题, 这些系统评价是否纳入了相同的试验或原始研究? 如果没有纳入相同的试验, 则应考虑选择纳入与决策问题最相关的研究的系统评价
- 如果不同系统评价纳入了相同的研究, 其研究的质量是否相同? 如果不同, 应选择高质量的系统评价。当两个系统评价都相关时, 如它们是解决同一问题的不同方面, 则二者都可用于取证

问题	可信度较低的原因
这篇评价是否明确提出了一个恰当的决策或管理问题?	<ul style="list-style-type: none"> - 系统评价提出的问题不明确……很可能回答不清楚 - 系统评价提出的问题与寻求的决策或管理问题不同……可能有误导作用
是否使用了恰当的标准去评价系统评价中的研究	<ul style="list-style-type: none"> - 模糊的标准……可能导致纳入研究的选择偏倚 - 标准与系统评价提出的问题不一致……很可能纳入不适合于回答该问题的研究
对相关研究的检索是否详尽且合理全面?	<ul style="list-style-type: none"> - 没有设计好(描述)检索策略……存在找不到重要相关研究证据的风险 - 检索策略不合理全面……存在偏倚风险(如更可能找到“阳性结果”的研究)且可能找不到重要的相关证据
研究系统评价主题的相关性和偏倚风险的评价是否可重复?	<ul style="list-style-type: none"> - 如果系统评价中未使用明确的标准或仅由1人作所有需要的判断……这些判断可能无重复性且存在偏倚
不同研究之间的结果是否相似?	<ul style="list-style-type: none"> - 如果研究结果间存在重要差异, 且对这些差异无令人信服的解释……很难知道真实结果是什么, 或当你实施一项相似决策时, 不知道期望值的大小

图 2 系统评价结果不可行和产生误导的原因

时, 决策者除需考虑该系统评价与决策问题的相关性和当地条件的证据外, 还应考虑系统评价的可信度评价结果。

资源

有用的文献及其他补充书目

- Higgins JPT, Altman DF: Chapter 8: Assessing risk of bias in included studies. In Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.1 (updated September 2008). Edited by Higgins JPT, Green S. The Cochrane Collaboration; 2008. Available at:<http://www.cochranehandbook.org>.
- Counsell C: Formulating Questions and Locating Primary Studies for Inclusion in Systematic Reviews. Ann Intern Med 1997, 127: 380-387.
- Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C et al.: Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol 2007, 7: 10. Available at:

<http://www.biomedcentral.com/1471-2288/7/10>

网站链接

- The Rx for Change database:<http://www.cadth.cinx.php/en/compus/optimal-ther-resources/interventions> - This summarises current research evidence about the effects of strategies to improve drug prescribing practice and drug use. This database includes summaries, including reliability assessments, of systematic reviews that evaluate the effects of strategies targeting professionals, the organisation of healthcare, and consumers.
- Cochrane Effective Practice and Organisation of Care (EPOC) Review Group:<http://www.epoc.cochrane.org/en/index.html> - The Review Group provides guidance on assessing the reliability of different types of studies of effectiveness.
- The SUPPORT (SUPporting POLicy relevant Reviews and Trials) Collaboration:<http://www.support-collaboration.org/index.htm> - This project produces summaries of high

priority reviews for low- and middle-income countries. These include assessments of reliability.

附件

术语: <http://www.biomedcentral.com/content/supplementary/1478-4505-7-S1-S8-S1.doc>

利益冲突、作者贡献及致谢

见原文。

参 考 文 献

- Lavis JN, Oxman AD, Lewin S, et al. SUPPORT Tools for evidence-informed health Policymaking (STP). Introduction. *Health Res Policy Syst*, 2009, 7(Suppl 1): I1.
- Lavis JN, Posada FB, Haines A, et al. Use of research to inform public policymaking. *Lancet*, 2004, 364: 1615-1621.
- Oxman AD, Lavis JN, Lewin S, et al. SUPPORT Tools for evidence-informed health Policymaking (STP). 1. What is evidence-informed policymaking. *Health Res Policy Syst*, 2009, 7(Suppl 1): S1.
- Oxman AD, Fretheim A, Lavis JN, et al. SUPPORT Tools for evidence-informed health Policymaking (STP). 12. Finding and using research evidence about resource use and costs. *Health Res Policy Syst*, 2009, 7(Suppl 1): S12.
- Renfrew MJ, Craig D, Dyson L, et al. Breastfeeding promotion for infants in neonatal units: a systematic review and economic analysis. *Health Technol Assess*, 2009, 13: 1-iv.
- Grimshaw JM, Thomas RE, MacLennan G, et al. Effectiveness and efficiency of guideline dissemination and implementation strategies. *Health Technol Assess*, 2004, 8: iii-72.
- Carlsen B, Glenton C, Pope C. Thou shalt versus thou shalt not: a meta-synthesis of GPs' attitudes to clinical practice guidelines. *Br J Gen Pract*, 2007, 57: 971-978.
- Mays N, Pope C, Popay J. Systematically reviewing qualitative and quantitative evidence to inform management and policy-making in the health field. *J Health Serv Res Policy*, 2005, 10(Suppl 1):6-20.
- Munro SA, Lewin SA, Smith HJ, et al. Patient adherence to tuberculosis treatment: a systematic review of qualitative research. *PLoS Med*, 2007, 4: e238.
- Pound P, Britten N, Morgan M, et al. Resisting medicines: a synthesis of qualitative studies of medicine taking. *Soc Sci Med*, 2005, 61: 133-155.
- Laurant M, Reeves D, Hermens R, et al. Substitution of doctors by nurses in primary care. *Cochrane Database Syst Rev*, 2005, 2: CD001271.
- Forsetlund L, Bjorndal A, Rashidian A, et al. Continuing education meetings and workshops: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev*, 2009, 2: CD003030.
- Mulrow CD. Rationale for systematic reviews. *BMJ*, 1994, 309: 597-599.
- Oxman AD, Schunemann HJ, Fretheim A. Improving the use of research evidence in guideline development: 8. Synthesis and presentation of evidence. *Health Res Policy Syst*, 2006, 4: 20.
- Oxman AD, Glasziou P, Williams JW Jr. What should clinicians do when faced with conflicting recommendations? *BMJ*, 2008, 337: a2530.
- Jadad AR, Cook DJ, Jones A, et al. Methodology and reports of systematic reviews and metaanalyses: a comparison of Cochrane reviews with articles published in paper-based journals. *JAMA*, 1998, 280: 278-280.
- Jorgensen AW, Hilden J, Gotzsche PC. Cochrane reviews compared with industry supported meta-analyses and other meta-analyses of the same drugs: systematic review. *BMJ*, 2006, 333: 782.
- Dixon E, Hameed M, Sutherland F, et al. Evaluating meta-analyses in the general surgical literature: a critical appraisal. *Ann Surg*, 2005, 241: 450-459.
- Jadad AR, Cook DJ, Browman GP. A guide to interpreting discordant systematic reviews. *CMAJ*, 1997, 156: 1411-1416.
- Jadad AR, Moher M, Browman GP, et al. Systematic reviews and meta-analyses on treatment of asthma: critical evaluation. *BMJ*, 2000, 320: 537-540.
- Linde K, Willich SN. How objective are systematic reviews? Differences between reviews on complementary medicine. *J R Soc Med*, 2003, 96: 17-22.
- 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.
- Critical Appraisal Skills Programme: 10 questions to help you make sense of reviews 2006 [http://www.phru.nhs.uk/Doc_Links/S.Reviews%20Appraisal%20Tool.pdf]. United Kingdom, Public Health Resource Unit.
- Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. *J Clin Epidemiol*, 1991, 44: 1271-1278.
- Canadian Coordinating Office for Health Technology Assessment: Proposed Evaluation Tools for COMPUS 2005 [http://www.cadth.ca/media/compus/pdf/COMPUS_Evaluation_Methodology_final_e.pdf]. Ottawa, Ottawa: Canadian Coordinating Office for Health Technology Assessment.
- West S, King V, Carey TS, et al. Systems to rate the strength of scientific evidence [Evidence report/technology assessment no 47] 2002 [<http://www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=hserta&part=A73054>]. Publication No. 02-E016. Rockville, MD, USA, Agency for Healthcare Research and Quality.
- Higgins JPT, Altman DF. Chapter 8: Assessing risk of bias in included studies. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.1 (updated September 2008)* 2008 [<http://www.cochrane-handbook.org/>]. The Cochrane Collaboration.
- Lavis JN, Oxman AD, Souza NM, et al. SUPPORT Tools for evidence-informed health Policymaking (STP). 9. Assessing the applicability of the findings of a systematic review. *Health Res Policy Syst*, 2009, 7(Suppl 1): S9.
- Counsell C. Formulating Questions and Locating Primary Studies for Inclusion in Systematic Reviews. *Ann Intern Med*, 1997, 127: 380-387.
- Higgins JPT, Green S. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.1 [updated September 2008]* 2008 [<http://www.cochrane-handbook.org/>]. The Cochrane Collaboration.
- Meremikwu MM, Donegan S, Esu E. Chemoprophylaxis and intermittent treatment for preventing malaria in children. *Cochrane Database Syst Rev*, 2008, 2: CD003756.
- Mulrow CD. The medical review article: state of the science. *Ann Intern Med*, 1987, 106: 485-488.
- Oxman AD. Checklists for review articles. *BMJ*, 1994, 309: 648-651.
- Grobler LA, Marais BJ, Mabunda S, et al. Interventions for increasing the proportion of health professionals practicing in underserved communities. *Cochrane Database Syst Rev*, 2009, 1: CD005314.
- Lefebvre C, Manheimer E, Glanville J, on behalf of the Cochrane

- Information Retrieval Methods Group. Searching for studies. Cochrane Handbook for systematic reviews of interventions. Version 5.0.1 [updated September 2008] 2008 [http://www.cochrane-handbook.org/]. The Cochrane Collaboration.
- 36 Dickersin K, Min YI. Publication bias: the problem that won't go away. *Ann N Y Acad Sci*, 1993, 703: 135-146.
 - 37 Hopewell S, Loudon K, Clarke MJ, et al. Publication bias in clinical trials due to statistical significance or direction of trial results. *Cochrane Database Syst Rev*, 2009, 21(1): MR000006.
 - 38 Hopewell S, Clarke M, Stewart L, et al. Time to publication for results of clinical trials. *Cochrane Database Syst Rev*, 2007, 2: MR000011.
 - 39 Lavis JN, Wilson M, Oxman AD, et al. SUPPORT Tools for evidence-informed health Policymaking (STP). 4. Using research evidence to clarify a problem. *Health Res Policy Syst*, 2009, 7(Suppl 1): S4.
 - 40 Lavis JN, Wilson MG, Oxman AD, et al. SUPPORT Tools for evidence-informed health Policymaking (STP). 5. Using research evidence to frame options to address a problem. *Health Res Policy Syst*, 2009, 7(Suppl 1): S5.
 - 41 Katrak P, Bialocerkowski A, Massy-Westropp N, et al. A systematic review of the content of critical appraisal tools. *BMC Medical Research Methodology*, 2004, 4: 22.
 - 42 Moher D, Jadad AR, Nichol G, et al. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Control Clin Trials*, 1995, 16: 62-73.
 - 43 Shepperd S, Doll H, Broad J, Gladman J, et al. Early discharge hospital at home. *Cochrane Database Syst Rev*, 2009, 1: CD000356.
 - 44 Lewin SA, Dick J, Pond P, et al. Lay health workers in primary and community health care. *Cochrane Database Syst Rev*, 2005, 1: CD004015.
 - 45 Jamtvedt G, Young JM, Kristoffersen DT, et al. Audit and feedback: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev*, 2006, 2: CD000259.
 - 46 Arblaster L, Lambert M, Entwistle V, et al. A systematic review of the effectiveness of health service interventions aimed at reducing inequalities in health. *J Health Serv Res Policy*, 1996, 1: 93-103.
 - 47 Moher D, Jadad AR, Klassen TP. Guides for reading and interpreting systematic reviews: III. How did the authors synthesize the data and make their conclusions? *Arch Pediatr Adolesc Med*, 1998, 152(9): 915-20.
 - 48 Oxman AD, Cook DJ, Guyatt GH. Users' guides to the medical literature. VI. How to use an overview. Evidence-Based Medicine Working Group. *JAMA*, 1994, 272: 1367-1371.
 - 49 Oxman AD, Lavis JN, Lewin S, et al. SUPPORT Tools for evidence-informed health Policymaking (STP). 10. Taking equity into consideration when assessing the findings of a systematic review. *Health Res Policy Syst*, 2009, 7(Suppl 1): S10.
 - 50 Oxman AD, Lavis JN, Fretheim A, et al. SUPPORT Tools for evidence-informed health Policymaking (STP). 17. Dealing with insufficient research evidence. *Health Res Policy Syst*, 2009, 7(Suppl 1): S17.
 - 51 Dixon-Woods M, Agarwal S, Jones D, et al. Synthesising qualitative and quantitative evidence: a review of possible methods. *J Health Serv Res Policy*, 2005, 10: 45-53.
 - 52 Noyes J, Popay J, Pearson A, et al. Chapter 20: Qualitative research and Cochrane reviews. Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.1 (updated September 2008) 2008 [http://www.cochrane-handbook.org/]. The Cochrane Collaboration.
 - 53 Lavis JN. Supporting the Use of Systematic Reviews in Policymaking. *PLoS Med*, 6(11): e1000141.
 - 54 Jefferson T, Demicheli V, Vale L. Quality of systematic reviews of economic evaluations in health care. *JAMA*, 2002, 287: 2809-2812.
 - 55 Siegfried N, Muller M, Deeks JJ, et al. Male circumcision for prevention of heterosexual acquisition of HIV in men. *Cochrane Database Syst Rev*, 2009, 2: CD003362.
 - 56 McGuinness B, Craig D, Bullock R, et al. Statins for the prevention of dementia. *Cochrane Database Syst Rev*, 2009, 2: CD003160.

李幼平 总审校
王莉 陈尹 陈群飞 译
蒋兰慧 杜亮 审校

本文编辑: 蔡羽嘉