

我国医学随机对照试验偏倚风险一致性评估

汤文婷¹ 楼雯^{1,2} 刘彦希¹

(¹ 华东师范大学经济与管理学院信息管理系 上海 200062

² 华东师范大学统计与数据科学前沿理论及应用教育部重点实验室 上海 200062)

〔摘要〕 **目的/意义** 为临床研究和元分析提供更科学统一的评估标准, 提高医学证据质量。**方法/过程** 检索中国知网, 筛选出 390 条重复的偏倚风险评估记录。对其从整体结果一致性、专题和内容一致性、指标一致性 3 方面进行分析。**结果/结论** 超过 80% 的组别对于同一随机对照试验的风险评估不一致。在主题内容方面, 感染性疾病及传染病、内分泌腺及全身性疾病领域研究一致性较高; 中医学的 Kappa 系数分布较广, 风险评估一致性差异较大。在指标方面, 选择性报道、结果数据的完整性和其他偏倚分歧较大。

〔关键词〕 偏倚风险; 一致性; 随机对照试验; 元分析

〔中图分类号〕 R-058 **〔文献标识码〕** A **〔DOI〕** 10.3969/j.issn.1673-6036.2025.08.006

Assessment of Consistency in the Risk of Bias in Medical Randomized Controlled Trials in China

TANG Wenting¹, LOU Wen^{1,2}, LIU Yanxi¹

¹ Department of Information Management, School of Economics and Management, East China Normal University, Shanghai 200062, China; ² Key Laboratory of Advanced Theory and Application in Statistics and Data Science (East China Normal University), Ministry of Education of China, Shanghai 200062, China

〔Abstract〕 **Purpose/Significance** To provide more scientific and unified assessment standards for clinical research and meta-analysis, in order to improve the quality of medical evidence. **Method/Process** 390 duplicate records of bias risk assessment are retrieved and screened out from CNKI. These records are analyzed from three aspects: overall result consistency, thematic and content consistency, and indicator consistency. **Result/Conclusion** Over 80% of the groups have inconsistent risk assessments for the same randomized controlled trial. In terms of subject content, there is a high consistency in research on contagious diseases and infectious diseases, endocrine glands and systemic diseases. The Kappa coefficient of traditional Chinese medicine is widely distributed, and there are significant differences in the consistency of risk assessment. In terms of indicators, there are significant differences in selective reporting, completeness of outcome data, and other biases.

〔Keywords〕 bias risk; consistency; randomized controlled trial (RCT); meta-analysis

〔修回日期〕 2025-04-22

〔作者简介〕 汤文婷, 硕士研究生; 通信作者: 楼雯, 博士, 副教授, 博士生导师。

〔基金项目〕 教育部重点实验室自主科研项目 (项目编号: KLATASDS2406); 上海市哲学社会科学项目 (项目编号: 2024BJC005)。

1 引言

随着医学迅速发展,医学文献爆炸式增长,医疗决策面临前所未有的挑战^[1]。医生和医疗专家不仅要有效筛选信息,还必须依赖高质量且可靠的证据支持临床决策。因此,如何确保证据的科学性和一致性成为亟待解决的重要问题^[2-3]。

循证医学作为现代医学的重要决策工具,通过对现有科学证据的系统汇总与分析,为临床实践提供坚实的理论基础^[4-5]。元分析作为循证医学的核心方法之一,通过对现有科学证据的收集、梳理和深入分析,为临床决策提供强有力的支持^[6-7]。然而,元分析结果并非总是稳定和可靠的,其准确性不仅依赖于严格的分析方法,还受限于原始研究本身的质量与评估一致性^[8-10]。尤其在随机对照试验(randomized controlled trial, RCT)中,偏倚风险评估的准确性直接影响元分析结论,进而影响临床实践决策依据。

尽管当前有多种工具可用于 RCT 偏倚风险评估,如 Cochrane 偏倚风险评估工具和 Jadad 量表^[9-10],但对其在实际应用中是否能保持一致性,仍然缺乏系统研究。在医学领域,偏倚风险评估的不一致可能导致对同一研究的不同解读,从而降低元分析结果的可靠性,甚至误导临床决策。因此,开展偏倚风险评估一致性研究具有极其重要的现实意义。

本研究旨在深入评估我国医学领域研究人员在进行元分析时对同一随机对照试验偏倚风险的判断一致性。通过对国内文献的系统筛选与分析,揭示当前偏倚评估中存在的显著不一致性,特别是在不同主题和指标下,偏倚风险评估一致性的变化,并分析原因,以期为临床研究和元分析提供更为科学和统一的评估标准,进而提高循证医学在临床决策中的应用效果,同时促进医学证据质量的提高。

2 相关研究

2.1 偏倚风险评估工具及其指标设置

2.1.1 Cochrane 及其指标设置 Cochrane 偏倚风险评估工具由国际 Cochrane 协作网制定,致力于

为医疗决策提供最佳证据。其结果可直接用于证据可靠性分级,是循证医学的核心方法学标准。根据《Cochrane 干预措施系统评价手册》(第 5 版)^[11],Cochrane 系统评价的偏倚风险评估表包含 7 个条目:随机序列生成、分配序列隐藏、盲法实施、结果测量者盲法、结果数据的完整性、选择性报告和其他潜在偏倚。(1) 随机序列生成和分配序列隐藏。是影响选择性偏倚的关键因素,评估依据为是否详细描述随机分配方法和隐藏过程。(2) 盲法实施。指受试者、研究人员和结果评估者在试验中不知道干预分配情况,缺乏盲法会夸大干预效应。(3) 结果数据的完整性。失访或数据缺失会导致结果数据不完整偏倚,意向性分析(intention-to-treat, ITT)常用于减小这种偏倚。本研究将“失访”条目和“意向性分析”条目并入“结果数据的完整性”条目。(4) 选择性报告偏倚。指报告和未报告结果之间的系统差异,通常统计学显著的结果更容易报告。(5) 其他潜在偏倚。包括未预见或特定研究特有的问题,如基线不平衡。

2.1.2 Cochrane 和 Jadad 指标相关性 Jadad 量表是另一种评估临床试验质量的工具,由 Jadad A R 等^[12]于 1996 年首次提出。该量表主要从随机化、盲法实施、失访/退出情况 3 个关键方面评价研究。其评分范围为 0~5 分,2 分及以下被视为低质量研究,3 分及以上被视为高质量研究。Cochrane 和 Jadad 在评估临床试验时有相似之处,且存在指标相关性。Hartling L 等^[13]通过分析 163 份 RCT 发现,Cochrane 的随机序列生成条目与 Jadad 中的随机化条目高度相关,盲法条目也高度相关。然而,两者在评估研究结果完整性和分配序列隐藏方面有所不同。Cochrane 通过详细条目评估随机序列生成和分配序列隐藏,而 Jadad 没有明确评估分配隐藏的条目。Julien H 等^[14]研究显示,Cochrane 在不完整结果数据评估上更为全面,而 Jadad 主要关注失访处理方法。本研究充分考虑两者的相关性和差异,以确保后续一致性评估的准确性和可靠性。

2.2 偏倚风险一致性分析

一致性分析数据来源主要为原始随机对照试验

研究。例如，贺海艳等^[15]为评价溃疡性结肠炎治疗方法的有效性和安全性，按照纳入和排除标准剔除 1 257 篇文献，最终纳入 14 篇原始 RCT 研究进行元分析。此类研究通常局限于单一主题，难以全面反映不同研究之间的整体偏倚风险评估状况。通过对多个研究综合评估发现，元分析能更全面地反映各研究在偏倚风险评估上的一致性情况。这种方法不仅能够汇集大量数据，提高统计效能，还能揭示不同研究之间的异同，为评估偏倚风险一致性提供坚实的依据。因此，本研究收集元分析文献，而非单一随机对照试验数据。

既往研究主要关注试验对象之间的异质性，较少探讨成对重复 RCT 的作者间评价一致性。相关研究大多使用卡方检验和 I^2 统计量检验判断异质性^[16]。只有极少数探讨不同作者对重复 RCT 评价的一致性，且通常聚焦于单一研究主题的指标差异。例如，Jordan V M B^[17]通过统计“低风险”“不确定”和“高风险”比例发现，月经失调和生育能力低下主题下，随机序列生成和不完整结果数据表现出较高一致性。但上述研究只局限于特定主题，缺乏对不同主题下偏倚风险评估一致性的全面分析。

本研究通过检索中国知网中所有医学领域重复 RCT，旨在通过统计学检验全面评估当前中文文献中偏倚风险评估的一致性，识别研究者在不同主题下评价一致性较高的指标；使用元分析数据，提供更具普遍性和代表性的研究视角，深入探讨不同主题和内容下的评估一致性差异。

3 研究设计

基于中国知网筛选医药卫生领域元分析文献，提取每个 RCT 的质量评价和偏倚风险判断。经过多次查重和交叉核对，最终确定 370 条重复 RCT 的风险评估记录，并整合同一 RCT 的评估结果，为一致性分析奠定基础。

3.1 文献检索

在中国知网进行高级检索，时限设定为建库至 2023 年。为了最大程度地检索全部医药卫生领域

RCT 元分析相关文献，确保覆盖所有相关临床研究和系统评价，设定检索策略为：篇文摘 = （“元分析” + “meta 分析” + “系统评价”）AND 中图分类号 = “R”。同时为了保证研究对象具有较高质量，在来源类别处仅保留《中文核心期刊要目总览》（简称北大核心）和中文社会科学引文数据库（Chinese social sciences citation index, CSSCI）中的学术期刊，共检索出 14 669 篇文献。

3.2 纳入与排除标准

纳入标准：中英文临床随机对照试验；采用系统评价、元分析作为研究方法。排除标准：重复发表文献；与作者联系后仍无法获取全文；综述、学位论文类文献；无法提取详细偏倚风险评估结果；未含有重复的 RCT 系统评价、元分析文献。

3.3 数据提取

通过标题及摘要排除重复、非 RCT、综述等文献 6 318 篇；通过全文复筛，排除无法提取详细偏倚评估结果的文献 6 957 篇；阅读剩余的 1 394 篇全文，提取 RCT 偏倚风险评估记录 10 633 条。两名研究人员独立将记录录入表格，去除未重复评估记录 10 263 条。多次查重后，最终纳入重复 RCT 风险评估记录 370 条。检索及数据提取流程，见图 1。提取每个 RCT 的基本信息，包括所纳入元分析中对每个 RCT 的质量评价/偏倚风险的判断和依据。两名研究人员各自独立完成信息提取，整理后再行交叉核对，如果有争议则及时讨论，协商解决。

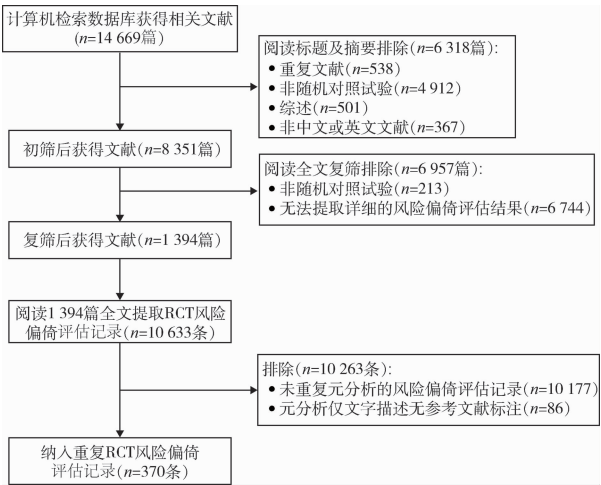


图 1 检索及数据提取流程

4 分析与结果

从整体结果一致性、专题和内容一致性、指标一致性 3 个方面进行分析。首先通过 Kappa 检验评估偏倚风险整体一致性，其次分析各医学专题和关键词在一致性上的分布特征，最后针对偏倚风险各指标进行一致性检验和统计分析。

4.1 整体结果一致性分析

将 370 条重复 RCT 偏倚风险评估记录按同一 RCT 分组，共形成 129 组。采用 Kappa 检验评估一致性，以 Kappa 值划分 5 级标准（ $Kappa < 0.2$ 为差、 $0.2 \leq Kappa < 0.4$ 为一般、 $0.4 \leq Kappa < 0.6$ 为中等、 $0.6 \leq Kappa < 0.8$ 为较强、 $Kappa \geq 0.8$ 为强），并以 $P < 0.05$ 为显著性判断依据。以 Kappa 系数为横轴，P 值为纵轴，绘制 Kappa 检验散点图，见图 2。点的颜色越深表明相应研究组数越多，粉色带状区间为置信区间。红色虚线表示 $P = 0.05$ 。在纳入分析的 129 组重复 RCT 偏倚风险评估记录中，106 组 $P \geq 0.05$ ，说明超过 80% 的组别对同一 RCT 偏倚风险评估的记录一致性并不具有统计学意义。这意味着多数研究对同一 RCT 的偏倚风险评估结论存在显

著分歧，严重削弱了研究结果的可靠性和可信度，见表 1。进一步分类统计显示，仅 13 组在 $P < 0.05$ 时 $Kappa \geq 0.8$ ，说明 90% 的研究组别对于元分析风险评估存在显著差异。值得关注的是，部分 Kappa 值为负数，意味着评估一致性低于随机猜测水平，反映出研究组别对同一 RCT 的评估分歧已超过随机预期，进一步突显了元分析在偏倚风险评估环节可能面临的深层次问题。

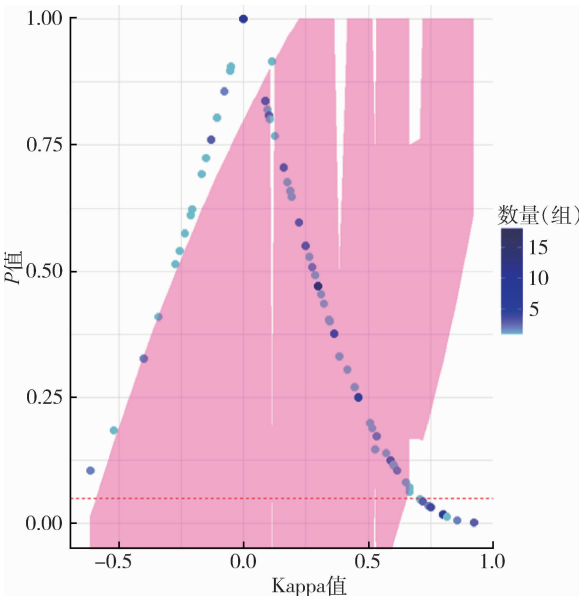


图 2 Kappa 检验散点

表 1 P-Kappa 值对应分类统计（组）

P	Kappa < 0	0 ≤ Kappa < 0.2	0.2 ≤ Kappa < 0.4	0.4 ≤ Kappa < 0.6	0.6 ≤ Kappa < 0.8	0.8 ≤ Kappa ≤ 1	合计
P ≥ 0.05	21	25	37	16	7	0	106
P < 0.05	0	0	0	0	10	13	23
合计	21	25	37	16	17	13	129

4.2 专题和内容一致性分析

4.2.1 专题一致性分析 根据中国知网所列出的文献所属专题，统计上述 370 条重复 RCT 偏倚风险评估记录的医学研究主题，占比最高的 6 类专题分别为：中医学、感染性疾病及传染病、内分泌腺及全身性疾病、临床医学、妇产科学、肿瘤学，将除此之外的专题归为“其他”，绘制桑基图，见图 3。2011—2013 年，中医学、内分泌腺及全身性疾病研

究占比较大；2014 年，感染性疾病及传染病受关注；2015 年妇产科学研究占比达到峰值（50%）；2016—2017 年妇产科学研究占比缩小，中医学、临床医学、肿瘤学研究占比回升，其他专题呈多样化发展；2018 年感染性疾病及传染病研究占比显著增加，可能与全球公共卫生事件相关；2019—2022 年，研究集中于内分泌腺及全身性疾病、中医学专题领域。

安全性时则意见分歧较大。这一观察为未来处理偏倚风险评估以及设计研究方案提供了重要启示。

4.3 指标一致性分析

纳入重复 RCT 偏倚风险评估的结果分布，见图 7。其中，随机序列生成偏倚、结果数据不完整偏倚、选择性报告偏倚中“低风险”占比最高，分别为 56%、64% 和 50%，分配隐藏、对研究者和受试者施盲、对结果测量者施盲和其他潜在偏倚中“不清楚”判断最多，分别占比 53%、44%、48% 和 53%。各指标风险统计，见表 2。

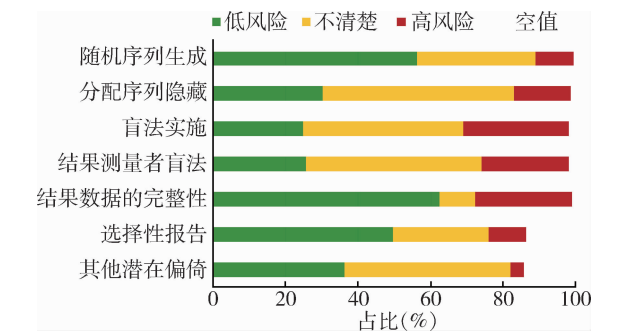


图 7 各指标高、低、不清楚风险占比

表 2 各指标风险统计 (组)

条目名称	不一致/非空组	一致/非空组
随机序列生成	28/88	60/88
分配序列隐藏	29/86	57/86
盲法实施	28/85	57/85
结果测量者盲法	28/85	57/85
结果数据的完整性	36/87	51/87
选择性报告	26/66	40/66
其他潜在偏倚	36/67	31/67

4.3.1 选择偏倚 选择偏倚包括随机序列生成和分配隐藏。(1) 随机序列生成。所有 176 组重复研究均对随机序列生成偏倚作出了判断。68% 的评价结果一致，其中 63% 被评为“低风险”。(2) 分配隐藏。173 组包含研究设计是否分配隐藏的判断。66% 的系统评价对重复 RCT 的评价结果一致，其中 69 组被评定为“不清楚”，65 组被评定为“高风险”。

4.3.2 盲法 (1) 对研究者和受试者施盲。171 组重复研究均对实施偏倚作出了判断。该偏倚风险下 66% 意见一致。在不一致中，有 46 组被归类为

“不清楚”或“高风险”。(2) 对结果测量者施盲。171 组包含研究设计是否对结果测量者施盲的判断，其中 72 组存在判断不一致，且 32 组研究的歧义在“高风险”与“不清楚”之间。

4.3.3 结果数据的完整性、选择性报告、其他潜在偏倚 (1) 结果数据的完整性。174 组重复研究均对结果数据的完整性作出了判断。其中 72 组对失访或退出的判断结果不一致。在不一致判断中，42 组的研究人员判断不同，且均在“低风险”和“高风险”间出现判断不一致。(2) 选择性报告。有 133 组重复研究对此作出明确判断，其中 52 组研究的判断结果不一致。在不一致判断中，“低风险”与“不清楚”的情况有 25 组，“高风险”与“不清楚”的情况有 23 组。(3) 其他潜在偏倚。135 组做出的判断中，有 73 组重复研究存在判断不同且均在“低风险”和“不清楚”间出现判断不一致。剩余 62 组判断结果一致的情况下有 34 组为“不清楚”判断。

5 讨论

5.1 专题和内容相关讨论

不同医学主题研究在一致性上存在明显差异。总体来看，感染性疾病及传染病、内分泌及全身性疾病研究一致性较高，而中医学研究一致性较低。在具体疾病疗法研究中，研究人员对风险的评估较一致，但在治疗有效性和安全性评估上则存在显著分歧。

中医学研究一致性较低的原因可能是其常涉及多种疗法和个性化治疗，对研究方法判断不统一。例如，蒲圣雄等^[18]认为 Diener H C 等^[19]的 RCT 分为针刺组和药物对照组，但由于针刺和药物的治疗形式完全不同，无法对患者及实施者给予盲法，因此对研究者和受试者施盲方面评价为“高风险”，但另一名学者对此并没有提出异议，判定为“低风险”。此外，中医学领域缺乏国际通用标准和指南，导致不同研究之间的比较和一致性评估较困难。例如，两个不同偏倚风险评估记录对同一中医学 RCT 随机序列生成条目均判断为“随机字样”，一个评价为“虽提到随机字样但方法不明”，判为“不明

风险”，另一个评价为“低风险”。

纳入文献质量不高在一定程度上导致了偏倚风险评估一致性下降。多篇文献均提到纳入文献整体质量不高，例如，有研究^[20]纳入 36 个 RCT，其中只有两个提及使用双盲法，9 个有病例退出或随访失败且报告了具体原因，但均未进行意向性分析，纳入的所有研究均未提及分配隐藏。研究发现，约 70% 的纳入研究未能提供足够的盲法实施细节，尤其在随机序列生成评估中，35% 的 RCT 未能详细说明随机化方法，表明缺乏关键细节报告显著影响了评估一致性，评估人员缺少足够信息以致无法作出判断，从而在“高风险”和“不清楚”之间摇摆。

5.2 指标相关讨论

对随机序列生成、分配序列隐藏和盲法实施的评价展现出较高的一致性，其中随机序列生成一致性最高，这与既往研究结果相符。随机序列生成的高度一致性原因可能有以下两点。一是《Cochrane 干预措施系统评价手册》中详细列举了具体随机方法归属，例如随机数字表法、抽签法属于低偏倚风险，而根据就诊号码、患者意愿分配则属于高偏倚风险。二是 Cochrane 在某些专题下有原则限制。如不允许妇科和生育组专题下的研究纳入准随机研究，如果纳入将被判定为高风险^[11]。

进一步分析对研究者和受试者施盲、分配序列隐藏、对结果测量者施盲的判断一致性情况，发现这 3 个指标与随机序列生成指标结果存在显著差异，56% 的随机序列生成评估结果为“低风险”，表明在多数纳入的研究团队看来，同一 RCT 随机序列的生成方法得当，能够较好地控制偏倚风险，而超过 60% 的分配序列隐藏、对研究者和受试者施盲、对结果测量者施盲结果则为“不清楚”，这更像是由于在原随机对照试验中未详细提及该指标内容，研究团队缺少足够信息而无法做出判断的权衡之举。

6 结论

本研究基于 1 394 篇文献提取 RCT 偏倚风险评估结果记录 10 633 条，经过多次查重，最终确定纳

入重复 RCT 的风险评估记录 370 条。进行 Kappa 一致性分析，并结合文献专题、关键词共现情况以及偏倚风险的 7 个指标进行深入探讨。研究结果显示，超过 80% 的组别在偏倚风险评估记录的一致性上未呈现出统计学意义，不同医学主题研究和不同偏倚风险指标的一致性水平存在显著差异，这些发现对元分析的可靠性和临床决策的科学性提出挑战，提高偏倚风险评估的一致性对于提升元分析质量至关重要。

因此提出以下建议。一是强化随机对照试验报告的规范化与透明化。原始研究报告细节的缺失，特别是在盲法实施方面，导致评估分歧。SHARE 报告清单^[21]和 ACURATE 指南^[22]增设了盲法评价相关条目，要求详细报告盲法的实施对象、方法、时间和结果。清晰完整的报告可以减少主观偏差，提高偏倚风险评估的一致性，增强元分析的证据基础。二是推动 Cochrane 评估工具的精细化与领域适配。针对中医学等评估分歧较大的领域，《Cochrane 干预措施系统评价手册》编制组应制定差异化规范，如为中医针灸和中药复方研究提供专属评估指南，明确个性化干预中随机序列生成和盲法可行性等指标的判断标准，并设计标准化评估模板，这将有助于提升复杂干预研究的评估可靠性。三是提升研究人员对评估工具的应用能力。研究者应系统学习 Cochrane 等工具的核心原则，通过培训和指南解读确保统一理解关键指标，如分配序列隐藏和选择性报告。当原始文献存在信息缺失时，应主动联系作者获取详细信息，避免数据不全导致的评估偏差。

本研究尚存在一定局限性。一是研究数据来源仅限于中国知网，未能全面反映国际随机对照试验风险评估的一致性。未来研究可涵盖更多国际数据库，进行跨国对比研究，以便为全球范围内的循证医学提供更具代表性和广泛适用性的结论。二是本研究采用 Kappa 值衡量评估一致性，但对研究人员在偏倚评估过程中可能存在的主观性偏差分析较为有限。未来研究可以结合定性分析，进一步探讨研究人员主观判断对偏倚评估一致性的影响。

未来研究方向首先应扩大研究范围和样本量，涵盖更多领域和不同地区的随机对照试验，并进行

跨国比较。其次可以系统比较多种偏倚风险评估工具,探讨其在不同领域的应用效果和一致性,特别是在复杂干预和多元研究设计中的适应性。最后可进一步探讨偏倚风险评估与临床结局之间的关系,分析偏倚评估一致性对元分析结论的可靠性和临床决策的影响。通过这些深入研究,进一步提高偏倚评估的一致性,增强元分析的科学性和临床应用价值,从而为循证医学的发展提供更为可靠的支持。

作者贡献: 汤文婷负责数据收集与分析、论文撰写;楼雯负责研究设计、论文审核与修订;刘彦希负责数据整理。

利益声明: 所有作者均声明不存在利益冲突。

参考文献

- 1 KALKUM E, KLOTZ R, SEIDE S, et al. Systematic reviews in surgery – recommendations from the study center of the German society of surgery [J]. *Langenbecks archives of surgery*, 2021, 406 (6): 1723 – 1731.
- 2 DENG H, XU J. Evidence – based medicine meta – evidence collecting and generating method, involves analyzing to – be – audited information, auditing information, and storing information into meta – evidence database to generate meta – evidence [R]. Shanghai: Shanghai University of Traditional Chinese Medicine, 2019.
- 3 MULROW C D. Systematic reviews: critical links in the great chain of evidence [J]. *Annals of internal medicine*, 1997, 126 (5): 389.
- 4 毕桢干, 黄桥, 阎思宇, 等. 系统评价与 Meta 分析的结果解读与可视化呈现——决策者和患者的视角 [J]. *中国循证医学杂志*, 2024, 24 (6): 715 – 721.
- 5 MURAD M H, MONTORI V M, IOANNIDIS J P A, et al. How to read a systematic review and meta – analysis and apply the results to patient care [J]. *JAMA*, 2014, 312 (2): 171 – 179.
- 6 曾宪涛, 包翠萍, 曹世义, 等. Meta 分析系列之三: 随机对照试验的质量评价工具 [J]. *中国循证心血管医学杂志*, 2012, 4 (3): 183 – 185.
- 7 DA COSTA B R, BECKETT B. Effect of standardized training on the reliability of the Cochrane risk of bias assessment tool: a prospective study [J]. *Systematic reviews*, 2017, 6 (1): 44.
- 8 MARTSEVICH S Y, KUTISHENKO N P. Clinical trials evaluating drug therapy: types, reliability of results, place in modern evidence – based medicine [J]. *Rational pharmacotherapy in cardiology*, 2021, 17 (2): 249 – 255.
- 9 吴婷婷, 刘丹璐, 黄娇, 等. 风险偏倚评估工具在针刺 Cochrane 系统评价中的应用 [J]. *中国循证医学杂志*, 2014, 14 (3): 361 – 364.
- 10 YALI L, SHENGPING Y, JUNJIE D, et al. Risk of bias tool in systematic reviews/meta – analyses of acupuncture in chinese journals [J]. *Plos one*, 2011, 6 (12): e28130.
- 11 HIGGINS P T, GREEN S, HIGGINS J, et al. *Cochrane handbook for systematic reviews of interventions* [M]. New York: John Wiley and Sons Ltd, 2009.
- 12 JADAD A R, MOORE R A, CARROLL D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary [J]. *Controlled clinical trials*, 1996, 17 (1): 1 – 12.
- 13 HARTLING L, OSPINA M, LIANG Y, et al. Risk of bias versus quality assessment instruments: a comparison of instruments [J]. *Systematic reviews*, 2013, 2 (1): 76.
- 14 JULIEN H, PIEDBOIS P, CHWALOW S, et al. Evaluation of the quality of methodological indicators in randomized controlled trials in surgery: a prospective, international, multicenter study [J]. *Annals of surgery*, 2015, 262 (2): 207 – 216.
- 15 贺海艳, 何舰, 王馨祎, 等. 肠道菌群移植治疗溃疡性结肠炎的有效性和安全性: 系统评价和 Meta 分析 [J]. *胃肠病学和肝病学杂志*, 2024, 33 (10): 1335 – 1341.
- 16 HARTLING L, HAMM M P. Testing the risk of bias tool showed low reliability between individual reviewers and across consensus assessments of reviewer pairs [J]. *Journal of clinical epidemiology*, 2013, 66 (9): 973 – 981.
- 17 JORDAN V M B, LENSEN S F. There were large discrepancies in risk of bias tool judgments when a randomized controlled trial appeared in more than one systematic review [J]. *Journal of clinical epidemiology*, 2017, 81 (9): 72 – 76.
- 18 蒲圣雄, 谭戈, 邓勇, 等. 针刺对比药物预防治疗偏头痛有效性及安全性的 Meta 分析 [J]. *重庆医科大学学报*, 2017, 42 (10): 1293 – 1298.
- 19 DIENER H C, KRONFELD K, BOEWING G, et al. Efficacy of acupuncture for the prophylaxis of migraine: a multi-centre randomised controlled clinical trial [J]. *Lancet neurol*, 2006, 5 (4): 310 – 316.
- 20 武士青, 方秦, 李秀秀, 等. 中国莫西沙星治疗耐多药肺结核疗效及安全性的 Meta 分析 [J]. *华中科技大学学报 (医学版)*, 2018, 47 (3): 354 – 360.
- 21 MA P, LIU X, LIU Z, et al. The SHARE: sham acupuncture reporting guidelines and a checklist in clinical trials [J]. *Journal of evidence – based medicine*, 2023, 16 (4): 428 – 431.
- 22 LEE Y S, KIM S Y, LEE H, et al. ACURATE: a guide for reporting sham controls in trials using acupuncture [J]. *Journal of evidence – based medicine*, 2023, 16 (4): 82 – 90.