

预后因素研究系统评价/Meta 分析的方法学质量评价工具 (AMSTAR-PF) 解读



张亚玲^{1, 2}, 张 强^{2, 3, 4}, 邓秀秀⁵, 郭雯燕⁶, 李东东⁷, 赵信科⁵, 卢存存⁸

1. 河南中医药大学第一附属医院血液肿瘤科 (郑州 450000)
2. 中西医防治重大疾病河南省协同创新中心 (郑州 450000)
3. 河南中医药大学第一附属医院脾胃肝胆科 (郑州 450000)
4. 河南省中医药科学院 (郑州 451163)
5. 甘肃中医药大学中西医结合研究所 (兰州 730101)
6. 甘肃中医药大学第一临床医学院 (兰州 730101)
7. 河南省中医院肿瘤科 (郑州 450003)
8. 川北医学院护理学院 (四川南充 637000)

【摘要】预后因素研究的系统评价/Meta 分析是整合疾病预后因素证据、支撑临床决策的关键方法,然而其方法学质量参差不齐,且现有通用质量评价工具未能充分涵盖此类研究的方法学特征。为此,国际学者近期在 *BMJ* 发表了首个专门用于评价预后因素研究的系统评价/Meta 分析方法学质量工具—AMSTAR-PF。本文系统介绍了该工具的制定背景与过程,并结合实例对其条目进行了解读,以期国内研究人员开展高质量预后因素研究的系统评价/Meta 分析提供方法学参考,促进该领域证据的规范生成与有效转化。

【关键词】预后因素;系统评价;Meta 分析;方法学质量;质量评价工具

【中图分类号】 R195.1 **【文献标识码】** A

Interpretation of a measurement tool to assess systematic reviews/Meta-analyses of prognostic factor studies (AMSTAR-PF)

ZHANG Yaling^{1,2}, ZHANG Qiang^{2,3,4}, DENG Xiuxiu⁵, GUO Wenyan⁶, LI Dongdong⁷, ZHAO Xinke⁵, LU Cuncun⁸

1. Department of Hematology and Oncology, The First Affiliated Hospital of Henan University of CM, Zhengzhou 450000, China
2. Collaborative Innovation Center of Prevention and Treatment of Major Diseases by Chinese and Western Medicine, Zhengzhou 450000, China
3. Department of Gastroenterology and Hepatology, The First Affiliated Hospital of Henan University of CM, Zhengzhou 450000, China
4. Henan Academy of Chinese Medical Sciences, Zhengzhou 451163, China
5. Institute of Traditional Chinese and Western Medicine, Gansu University of Chinese Medicine, Lanzhou 730101, China
6. The First School of Clinical Medicine, Gansu University of Chinese Medicine, Lanzhou 730101, China
7. Department of Oncology, Henan Province Hospital of Traditional Chinese Medicine, Zhengzhou 450003, China

DOI: 10.12173/j.issn.1004-5511.202604065

基金项目: 河南省科技研发计划联合基金 (优势学科培育类) 项目 (242301420112); 国家中医药传承创新中心联合共建科研专项 (2024ZXZ1176)

通信作者: 卢存存, 博士, Email: lu17metrics@163.com

张强, 博士, Email: zhangqiangq7@126.com

8. School of Nursing, North Sichuan Medical College, Nanchong 637000, Sichuan Province, China

Corresponding authors: LU Cuncun, Email: lu17metrics@163.com; ZHANG Qiang, Email: zhangqiangq7@126.com

【Abstract】 Systematic reviews and Meta-analyses of prognostic factor studies are critical for synthesizing evidence on disease prognostic factors and informing clinical decision-making. However, their methodological quality varies widely, and existing generic quality assessment tools fail to adequately cover the specific methodological features of this type of research. In response, an international team recently developed and published AMSTAR-PF in the *BMJ*—the first tool dedicated to evaluating the methodological quality of systematic reviews and Meta-analyses of prognostic factor studies. This article systematically outlines the development background and process of AMSTAR-PF, interprets its items with illustrative examples, and aims to offer methodological guidance for Chinese researchers conducting high-quality systematic reviews and Meta-analyses in this field, thereby facilitating the standardized generation and effective translation of prognostic evidence.

【Keywords】 Prognostic study; Systematic review; Meta-analysis; Methodological quality; Quality assessment tool

预后研究是循证医学的重要研究类型之一，为临床风险评估、结局预测及分层决策提供了关键的证据支持^[1-2]。预后研究策略（Prognosis Research Strategy, PROGRESS）框架将预后研究系统划分为基础预后研究、预后因素研究、预后模型研究以及分层医疗研究4个相互衔接的模块^[2-3]。其中，预后因素研究用于识别与疾病未来结局相关的临床特征、生物学指标及心理社会因素等，是个体化预后沟通、构建精准预后模型以及制定靶向干预策略的重要证据基础^[3-5]。近年来，随着电子病历数据库、注册登记平台的大规模应用，以及多组学检测技术和人工智能（如机器学习、深度学习）方法的快速发展，潜在预后因素的探索维度不断扩展，相关原始研究及系统评价/Meta分析的数量也持续增长^[6-8]。然而，相关证据合成研究在起点定义、暴露与结局测量、混杂因素控制等方面仍存在较大异质性，不仅影响研究结果的稳定性，甚至可能导致结论不一致，从而影响预后证据向临床实践的转化^[4-5, 9]。

系统评价/Meta分析作为整合预后因素原始研究证据的重要方法，其“高产出”并不等同于“高质量”^[4]。*Journal of Clinical Epidemiology*发表的一项大型调查研究显示，医学领域系统评价/Meta分析发表数量增长迅速，但同一主题常出现研究重复、方法学异质性及结论不一致等问题^[8]，无疑削弱了证据对临床指南与卫生决策的支持力度^[10-11]。目前，PRISMA（Preferred Reporting Items for Systematic Reviews and Meta-Analyses）及其系列

扩展版报告规范已在提升系统评价/Meta分析报告的透明度方面发挥了重要作用^[12-13]；在偏倚风险与质量评价方面，QUIPS（Quality in Prognosis Studies）用于评价预后因素原始研究的偏倚风险^[9]，ROBIS（Risk of Bias in Systematic Reviews）用于评价系统评价/Meta分析的偏倚风险^[14]，AMSTAR（A Measurement Tool to Assess Systematic Reviews）及其更新版AMSTAR-2则提供了系统评价/Meta分析方法学质量的通用评价框架^[13, 15]。然而，上述工具均未全面涵盖预后因素系统评价/Meta分析的关键方法学环节，如预后因素与结局的分类、校正因素、效应量的计算等^[4-5]。基于此，Henry等学者在AMSTAR-2基础上，于2025年在*BMJ*上发表首个专门用于评价预后因素研究的系统评价/Meta分析方法学质量工具—AMSTAR-PF（A Measurement Tool to Assess Systematic Reviews of Prognostic Factor Studies）^[5]。本文旨在系统介绍该工具的制定过程，并结合已发表实例对其条目进行逐条解读，以期国内学者开展相关研究提供方法学借鉴，推动预后因素高质量证据的规范生成与临床转化。

1 AMSTAR-PF 制定过程简介

AMSTAR-PF以AMSTAR-2为框架基础制定而成，研制流程主要包括以下3个阶段^[5]。

1.1 核心框架构建与内部开发

研究于2020年7月启动，组建了在AMSTAR、AMSTAR-2、ROBIS等多个国际方法学评价工具

或规范研制方面具有丰富经验的核心团队。经论证,团队一致决定以AMSTAR-2为基础框架,主要因其结构完整、用户友好,且侧重于全面的质量评价。通过将AMSTAR-2中干预性研究相关条目替换并调整为预后因素研究特有的内容,经多轮线上会议与邮件协作修订,于2021年7月形成AMSTAR-PF 1.0版及配套指南。

1.2 外部评审与反馈完善

工具先后历经2轮外部评审。首轮面向Cochrane预后方法学组23名核心成员征求意见,于2021年10月25日至11月27日收到9份有效回复,核心研究小组据此修订形成2.0版;次轮将2.0版及相同匿名问卷发给该组全体成员(207人),于2022年4月29日至5月27日收到21份有效回复。团队通过匿名问卷收集关于条目冗余、遗漏、相关性及实用性的反馈,据此修订工具与指南,并同步开发了可自动汇总评价答案的Excel辅助文件,形成3.0版本。

1.3 试点测试与评价者间一致性测试

为检验工具的可用性、评价者间一致性及时间成本,研究开展了3轮试点测试,共26名评价者参与。初步测试由2名博士生完成,其反馈意见形成4.0版;随后10名评价者使用该版本对23篇系统评价/Meta分析进行评价,进一步优化形成5.0版。最终测试采用5.0版,由14名未参与工具开发及前期测试的评价者对8篇系统评价/Meta分析进行独立评价,并正式计算Gwet一致性系数。根据测试反馈与一致性分析结果,对指南文件、信号条目及问题表述进行微调,最终形成AMSTAR-PF正式版。计算结果显示,多数条目的评价者间一致性达到“中度”至“几乎完全一致”水平,单篇评价平均耗时约39 min,达成共识平均约需11 min^[5],表明工具具有良好的可操作性与时间可接受性。

2 AMSTAR-PF 条目解读

AMSTAR-PF适用于预后因素系统评价/Meta分析的方法学质量评价,是该领域首个专用工具。该工具共包含14个领域、19个条目,部分条目设有评价要点以辅助判断,见附件表1。回答选项采用五级制,包括“是(Y, Yes)”“可能是(PY, Probably yes)”“可能否(PN, Probably no)”“否(N, No)”及部分条目涉及

“不适用(N/A, Not applicable)”。其中,“是”表示方法学质量较高,“否”则表示较低;若信息不足,可选择“可能是/可能否”以记录不确定性。

基于对全部19个条目的回答,可将单篇系统评价的整体方法学质量划分为以下4个等级之一^[5]:①高(high):所有问题均回答为“是”“可能是”或“不适用”;②中(moderate):出现不超过2个“可能否”或“否”的回答,且评价者判断这些条目对系统评价/Meta分析的结果与结论影响有限;③低(low):出现不超过2个“可能否”或“否”的回答,但评价者判定这些条目对系统评价的结果与结论具有显著影响;④极低(critically low):“可能否”或“否”的回答总数超过2个。需要注意的是,在正式应用前,建议评价团队对判断标准通过校准达成统一;在实际使用中,建议评价团队综合考量各条目识别出的所有局限性,审慎判断其对系统评价结果与结论的可能影响,并据此确定其整体方法学质量的最终评价等级,同时提供明确的判断依据。

为便于读者进一步理解AMSTAR-PF的具体应用方法,本文以2019年发表于BMJ的《与院内心脏骤停后生存相关的骤停前及骤停中预后因素:系统评价与Meta分析》^[16]一文为例,对各条目进行逐项解读。

2.1 领域1 研究问题

条目1 该系统评价是否清晰定义了研究问题,包括PICOTS相关组成部分?

一项高质量且具可信度的预后因素研究系统评价,其核心基础在于构建预先界定、表述清晰且信息维度充分的研究问题。AMSTAR-PF明确强调应采用PICOTS框架构建此类问题,该框架通过明确目标人群(population, P)、待评价预后因素(index prognostic factor, I)、对照因素或其他预后因素(comparator/other prognostic factors, C)、结局指标(outcome, O)、时间节点(timing, T;含预后评估起始时点与结局发生/判定的终止时点)及研究场景(setting, S;场景或情境)6大核心维度,确保系统评价的可重复性,而清晰界定这些要素不仅能为文献纳入与排除标准的制定提供逻辑依据,更有助于读者精准判断评价结果在自身临床实践或研究中的适用性,因此采用PICOTS框架规范呈现研究问题是体现预后因素研究系统评价方法学严谨性的关键。

实例研究问题中虽明确界定了目标人群（院内心脏骤停成年患者）、结局指标（住院期间、28 d 或 30 d 死亡率）及研究场景（急诊科、普通病房或重症监护室），但对预后因素的描述仅停留在“骤停前及骤停中预后因素”的宽泛层面，未进一步细化区分待评价预后因素与对照/其他预后因素，故该研究在条目 1 可评价为“可能是”。

2.2 领域 2 预注册方案

条目 2a 该系统评价是否声明在开展研究前已注册了方案，且注册信息是否公开可获取？

条目 2b 该系统评价是否说明了任何对方案的偏离？

AMSTAR-PF 将研究方案的可追溯性视为确保预后证据可信度的基石。该工具要求，系统评价应在实施前撰写研究方案，并在相关平台^[17]（如 PROSPERO、INPLASY、OSF）进行注册，以确保方案可公开获取。后续任何对已注册方案的修改，都必须在系统评价中明确记录并合理解释，从而最大限度地降低因“根据结果调整方法”而引入的偏倚风险。为此，AMSTAR-PF 将该内容拆分为条目 2a 和 2b 两个独立条目进行评价。

实例方法学部分明确报告了其研究方案已在 PROSPERO 平台预先注册并提供了注册号；同时指出，在实施阶段依据预后因素系统评价方法学指南的建议，将原计划使用的质量评价工具调整为 QUIPS 工具，故该研究在条目 2a 和 2b 均可评价为“是”。

2.3 领域 3 纳入研究的设计与类型

条目 3 该系统评价是否报告了符合纳入标准的预后因素研究类型，如哪些原始研究设计是合格的？

纳入研究设计的选择会直接影响预后因素系统评价证据的偏倚风险与方法学异质性，因此 AMSTAR-PF 要求系统评价作者在方案中明确说明拟纳入的研究设计类型，例如前瞻性或回顾性队列研究、病例对照研究、注册登记研究、随机对照试验等。其中，前瞻性队列设计通常被视为预后因素研究的理想设计，因其能更好地控制暴露与结局的时间顺序、测量时点及方法，从而有效降低偏倚、提高研究效度。除专门探讨预后因素的研究外，预后模型研究或因果推断研究也可提供相关预后因素信息。但需

注意，纳入不同研究目的或设计类型的研究可能引入异质性，因此在进行效应量合并分析时，需结合实际情况考虑采用分层分析或单独合成的策略。

实例文献筛选部分明确报告纳入前瞻性与回顾性观察性研究、随机对照试验及准随机试验，并通过清晰的排除标准界定了需排除的不符合目标研究设计或研究人群的文献，故该研究在条目 3 可评价为“是”。

2.4 领域 4 检索策略

条目 4 该系统评价是否使用了全面的检索策略收集相关研究？

综合、系统的检索是保证预后因素研究系统评价证据完整性的关键环节^[18]。AMSTAR-PF 强调，检索策略必须足够全面，以确保能检索到所有相关研究，从而实现对现有证据的全面分析与可靠总结。若检索不充分，可能导致重要研究被遗漏，进而使系统评价对证据现状的判断失真、结论可信度降低。全面检索的具体内涵需结合具体学科领域、所关注的预后因素及结局类型来灵活界定。评价者应据此判断所选数据库（如电子数据库、研究注册平台等）、构建的检索式（如主题词与自由词组合）以及对灰色文献与补充检索（如追溯参考文献）的利用，是否与研究问题相匹配。此外，检索结果的时效性（通常要求检索日期应在系统评价发表前的 12 个月内）也是重要的考量因素。

实例接收于 2019 年 10 月 29 日，并在 12 月 4 日被正式发表。其在方法学部分明确报告检索了 PubMed、Embase、Scopus、Web of Science 及 The Cochrane Library 等多个数据库，检索时限为建库至 2019 年 2 月 4 日，并在附件中提供了详细的检索策略。此外，该研究还提及由专业图书馆员参与检索策略制定，并进行了引文追溯。然而，其未描述灰色文献的检索情况，也未明确说明检索时是否有发表语言或日期限制，故该研究在条目 4 可评价为“可能是”。

2.5 领域 5 评价与纳入研究的过程

条目 5 在评价研究是否纳入系统评价时，是否遵循了严格的过程？

为确保研究筛选过程的客观性与可重复性，AMSTAR-PF 要求必须由至少 2 名评价者独立完成文献的标题/摘要筛选及全文复筛，并预先确

立分歧解决机制（如讨论达成共识或第三方仲裁）。

实例报告使用了 Covidence 软件进行文献管理，由 2 名研究者独立完成所有检索文献的标题/摘要筛选及后续的全文复筛。所有意见分歧均通过双方讨论达成一致，故该研究在条目 5 可评价为“是”。

2.6 领域 6 排除的研究

条目 6 该系统评价是否列出了在经过全文筛选后被排除的所有研究，并说明了排除原因？

在全文筛选阶段，所有被排除的研究均应以清单形式逐一列出，并明确说明每项研究的排除理由。该要求旨在确保筛选过程具有充分的透明度与可追溯性，使读者能够评估排除决策是否合理、是否存在潜在的筛选偏倚（如“按结果排除”的风险），从而判断系统评价的可靠性及其结果的适用性。

实例虽提供了 PRISMA 流程图并说明了全文筛选的数量与主要排除原因，但未提供被排除研究的详细清单及其对应的具体排除理由，故该研究在条目 6 可评价为“可能否”。

2.7 领域 7 数据提取

条目 7a 在对每项纳入研究进行数据提取时，是否遵循了严格的过程？

条目 7b 该系统评价是否充分描述了纳入研究的特征？

条目 7c 在预后因素的效应估计值及其精确度未在原始研究中完全报告的情况下，该系统评价是否使用了适当的技术来计算或获取它们？

提取数据的准确性直接影响系统评价结果的可靠性^[19]。AMSTAR-PF 将其视为预后因素系统评价中易引入偏倚的关键环节，要求至少 2 名评价者独立提取，并规范、完整地报告提取过程与数据细节。该领域重点关注 3 个方面：①提取流程是否规范（双人独立提取、使用预设提取表、具备分歧解决机制）；②提取信息是否充分（详细描述研究特征，明确校正/未校正效应量及协变量）；③效应量转换计算是否合理（必要时采用恰当方法推算并清晰说明）。若仅合并未校正效应量，还需说明其适用性与局限性。

实例明确报告由 2 名研究者使用预先设计的表格独立提取数据，内容涵盖作者、发表年份、生存相关的未校正或校正比值比（odds ratio，

OR）等；同时说明排除了未报告 OR 及其置信区间且无法通过计数数据计算 OR 的文献。此外，该研究在正文与附件中详细呈现了纳入研究特征及完整的数据提取表，故该研究在条目 7a、7b 和 7c 均可评价为“是”。

2.8 领域 8 评价偏倚风险

条目 8a 在评价每项纳入研究的偏倚风险时，是否遵循了严格的过程？

条目 8b 该系统评价是否使用了适当的技术来评价纳入的单项研究的偏倚风险？

预后因素证据的内部效度高度依赖于偏倚风险评价的准确性。AMSTAR-PF 明确要求，至少由 2 名评价者独立完成偏倚风险评价，并预先制定分歧解决方案（如讨论达成共识或第三方仲裁）。在评价工具方面，应优先选用专为预后因素研究设计的工具（如 QUIPS），以保证评价的针对性与适用性。该领域重点关注评价流程是否规范，以及是否采用了恰当的评价方法。

实例明确报告由 2 名研究者使用预后研究专用工具 QUIPS 独立开展偏倚风险评价，分歧通过共同讨论解决；同时文中也指出该选择符合相关方法学指南的推荐要求，故该研究在条目 8a 与 8b 均可评价为“是”。

2.9 领域 9 数据合成

条目 9a 若进行了数据合成，采用的方法是否确保了结果的可解释性？

条目 9b 若进行了 Meta 分析，该系统评价是否使用了适当的分析方法？

在预后因素研究的系统评价中，数据合成可采用定性或定量方法，其中定量合成即 Meta 分析。AMSTAR-PF 强调，无论采用何种合成方式，均应明确阐述数据合成原则，包括效应量的选择与合并依据、重复队列或多次报告的处理方式、多模型或多时间点数据的整合策略等。若进行 Meta 分析，通常推荐使用随机效应模型以处理预后研究中普遍存在的异质性^[20]，并对异质性进行量化与探索。

实例在数据合成部分明确报告了对来自同一大型注册数据库的重复报告，优先纳入样本量最大的研究；主要分析采用随机效应模型合并校正后的 OR，并通过 I^2 、 τ^2 及森林图评估异质性；同时以未校正 OR 的合并结果作为次要分析。该研究虽报告了异质性指标并采用随机效应模型，但

未预先设定分层合并或综合探索异质性来源的分析方案,故该研究在条目 9a 和 9b 均可评价为“可能是”。

2.10 领域 10 小样本效应

条目 10 若该系统评价进行了定量数据合成,是否对小样本效应进行了充分的调查?

AMSTAR-PF 强调“小样本效应”并不等同于发表偏倚^[5],要求作者关注 Meta 分析是否采用漏斗图等合适方法对其进行检验,并在结果解释时避免将不对称性简单归因于发表偏倚,因为异质性等其他因素同样可能导致该现象^[5, 21],从而提高证据解读的严谨性。

实例在正文及附件中均未报告对小样本效应的评估(如漏斗图或相关统计检验),也未就此展开讨论,故该研究在条目 10 可评价为“否”。

2.11 领域 11 讨论偏倚风险

条目 11 该系统评价在解释/讨论结果时,是否考虑了单项研究的偏倚风险?

系统评价中不仅需要进行偏倚风险评价,更应在结果解释中深入探讨偏倚风险对效应量方向与大小的潜在影响。若纳入研究质量参差不齐(这在预后因素研究的系统评价中较为常见),可通过敏感性分析、亚组分析或 Meta 回归等方法,检验结论在不同偏倚风险情境下的稳健性,避免仅停留于形式化的评价。

实例明确报告其所纳入研究的总体偏倚风险较低,且无研究在 QUIPS 的任一领域被评为高风险;同时,在附件中完整呈现了各研究在所有 QUIPS 领域的详细评价结果,故该研究在条目 11 可评价为“是”。

2.12 领域 12 讨论异质性

条目 12 该系统评价是否讨论了其结果中观察到的任何异质性?

在预后因素研究的系统评价中,研究间常存在较高的异质性。为此,AMSTAR-PF 要求系统评价作者必须全面分析并阐明异质性的潜在来源,包括临床异质性(如人群特征、预后因素与结局的定义及测量、随访时间等)和方法学异质性(如研究设计、统计模型、变量调整策略等)。若进行了 Meta 分析,还应量化报告异质性(如通过 τ^2 和预测区间),以帮助读者判断结果的科学性与外推性。

实例不仅讨论了统计学异质性(如较高的 I^2

值),还详细分析了临床异质性的来源(如不同研究中“监测”定义差异),并在局限性部分阐述了这些异质性对结果解读的潜在影响,故该研究在条目 12 可评价为“是”。

2.13 领域 13 利益冲突

条目 13 该系统评价是否报告了纳入的单项研究以及系统评价作者团队中任何潜在利益冲突来源,包括所获资助?

存在利益冲突的基金资助往往会影响科学研究结果的可信度^[22]。因此 AMSTAR-PF 要求对潜在的利益冲突进行双重声明:一是系统评价作者自身的资金来源及其他利益关系;二是所纳入原始预后因素研究的资金来源及可能存在的利益冲突。这一要求旨在增强证据透明度,帮助读者识别可能影响系统评价结果客观性的潜在因素。

实例虽在文末提供了系统评价作者层面的利益冲突与资助声明,但未系统报告所纳入各项原始研究的相应信息。由于利益冲突的存在可能直接影响系统评价结论的可信度,故将该研究在条目 13 评价为“可能否”。

2.14 领域 14 结果的确信度

条目 14 该系统评价是否说明了其关键发现的确信度,并使用适当的方法对每个结局得出总体确信度判断?

在预后因素研究的系统评价中,AMSTAR-PF 要求对关键结局的证据质量或确定性程度进行正式评价。该过程旨在将方法学质量、偏倚风险、异质性、间接性及不精确性等维度整合起来,形成对证据临床适用性及结论确定性的综合判断,从而避免仅基于统计显著性做出过度推断。GRADE (Grading of Recommendations Assessment, Development and Evaluation) 是目前评价证据确定性的常用框架,但其尚未完全覆盖所有类型的预后因素研究。因此,系统评价作者在应用时可能需要基于现有关于预后因素的 GRADE 指导原则进行适当调整,或参考其核心思想制定适用的评价方法。

实例明确报告由 2 名研究者采用 GRADE 方法对主要结局的证据质量进行了评价,并在附件中提供了证据概要表,故该研究在条目 14 可评价为“是”。

综合所有条目评价结果(条目 6 和 13 为“可能否”,条目 10 为“否”),实例的总体方法学质量可评定为“极低”。

3 讨论

AMSTAR-PF 作为首个针对预后因素研究的系统评价/Meta 分析的方法学质量评价工具,其核心价值在于将质量评价的维度与预后证据特有的推断链条紧密结合。本文概述了 AMSTAR-PF 的制定过程,并结合已发表实例解读了工具条目,可为国内学者开展相关研究提供方法学参考。但需说明的是,文中实例仅用于演示该工具的实操流程,不能代表该领域系统评价的整体方法学质量,同时评价过程中不可避免存在一定的主观偏差。

预后因素研究的核心并非直接验证干预措施的效果,而是识别与结局相关的特征,为风险评估和个体化决策提供依据^[2-3]。因此,其证据的可信度高度依赖于对预后因素的明确定义、对混杂因素和时间因素的恰当处理,以及对效应量的合理解释^[3-4]。值得注意的是,AMSTAR-PF 并非对 AMSTAR-2 的简单修订,而是一次以证据类型为导向的系统性重构。其摒弃了通用工具中可能不适用于预后情境的“关键条目”概念,转而强调所有条目都可能在不同情境下成为影响结论可信度的关键,并要求评价者结合具体研究目的进行综合判断^[5]。与 ROBIS 主要关注偏倚风险、AMSTAR-2 聚焦于干预性研究的评价逻辑不同,AMSTAR-PF 通过整合 PICOTS 框架、注重调整后效应量的优先性^[4-5],并全面考察预后因素研究特有的偏倚来源,使其评价结果能更直接地反映预后结论在临床应用中的稳健性。

该工具的应用价值不仅体现在对已完成证据合成研究的质量评价与分级上,更在于为前瞻性证据合成研究的设计与报告提供了方法学规范。清晰、透明的报告是确保研究可重复、可评价和可转化的基石^[23]。AMSTAR-PF 在方法学上的严格要求,与 PRISMA 2020 等报告规范形成了互补与协同体系:PRISMA 系列规范旨在确保研究过程与结果的完整呈现^[24-25],而 AMSTAR-PF 则在于评价这些已报告内容的方法学可靠性。两者结合,能够系统性地提升预后因素研究系统评价/Meta 分析的透明度与研究质量^[25]。在实际应用中,AMSTAR-PF 可作为研究者撰写此类系统评价研究方案和报告时的“自检清单”^[26],引导其在研究设计阶段就主动遵循此类研究的核心

方法学规范,例如明确定义 PICOTS 要素、详尽报告混杂因素的调整策略等。对于证据使用者而言,该工具提供的结构化评价框架有助于在指南制定、证据图谱或同类系统评价^[27-28]中,以透明、可重复的方式识别证据的薄弱环节,并审慎评价其对结论方向与稳健性的潜在影响,从而做出合理的决策。

未来,AMSTAR-PF 的推广与深化应用有望与人工智能技术,特别是与大语言模型的发展深度结合。一方面,基于自然语言处理技术的自动化工具可以辅助完成系统评价中耗时费力的环节,如标题的初筛、核心数据的提取等,从而提升证据合成的效率^[29-30]。另一方面,通过训练大语言模型(如 DeepSeek、ChatGPT)理解和应用 AMSTAR-PF 的评价逻辑,未来或可开发出智能化的初评工具,为研究者提供初步的方法学质量分析,或辅助期刊编辑和评审人进行快速、一致的质量筛查^[31-32]。当然,这种自动化辅助的核心目的在于提升效率与一致性,而非替代研究者的专业判断。AMSTAR-PF 与先进技术、报告规范的结合,将共同推动预后因素研究证据的生产、评价与转化走向更高程度的规范化、透明化与智能化,为临床实践与卫生决策提供更为坚实、可信的证据基础。

附件见《医学新知》官网附录(<https://yxxz.whuznhmedj.com/futureApi/storage/appendix/202604065.pdf>)

伦理声明:不适用

作者贡献:研究选题与指导:卢存存;论文撰写、经费支持:张亚玲、张强;论文审查与修订:邓秀秀、郭雯燕、李东东、赵信科、卢存存

数据获取:不适用

利益冲突声明:无

致谢:不适用

参考文献

- 王勤荣. 疾病预后评价——基层医疗中不可忽视的问题[J]. 全科医学临床与教育, 2004, (3): 147-149. [Wang QR. The assessment of disease prognosis—a noticeable problem for primary care[J]. Clinical Education of General Practice, 2004, (3): 147-149.]
- Hemingway H, Croft P, Perel P, et al. Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes[J]. BMJ, 2013, 346: e5595.
- Riley RD, Hayden JA, Steyerberg EW, et al. Prognosis Research Strategy (PROGRESS) 2: prognostic factor research[J]. PLoS Med,

- 2013, 10(2): e1001380.
- 4 Riley RD, Moons KGM, Snell KIE, et al. A guide to systematic review and Meta-analysis of prognostic factor studies[J]. *BMJ*, 2019, 364: k4597.
- 5 Henry ML, O'Connell NE, Riley RD, et al. AMSTAR-PF: a critical appraisal tool for systematic reviews of prognostic factor studies[J]. *BMJ*, 2025, 391: e085718.
- 6 Shen Y, Yu J, Zhou J, et al. Twenty-five years of evolution and hurdles in electronic health records and interoperability in medical research: comprehensive review[J]. *J Med Internet Res*, 2025, 27: e59024.
- 7 王道洋, 胡建颖, 宁雪丽, 等. 大模型驱动的多组学队列数据整合与疾病预测[J]. *生命科学*, 2025, 37(12): 1634-1644. [Wang DY, Hu JY, Ning XL, et al. Large model-driven integration of multi-omics cohort data and disease prediction[J]. *Chinese Bulletin of Life Sciences*, 2025, 37(12): 1634-1644.]
- 8 Hoffmann F, Allers K, Rombey T, et al. Nearly 80 systematic reviews were published each day: observational study on trends in epidemiology and reporting over the years 2000-2019[J]. *J Clin Epidemiol*, 2021, 138: 1-11.
- 9 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.
- 10 Ou S, Luo J, Jiang Q. Overlapping systematic reviews on the same topic: a systematic literature review of quantitative research[J]. *J Eval Clin Pract*, 2025, 31(4): e70148.
- 11 Kwok W, Dallant T, Martin G, et al. Systematic reviews on the same topic are common but often fail to meet key methodological standards: a research-on-research study[J]. *J Clin Epidemiol*, 2026, 189: 112018.
- 12 Helbach J, Hoffmann F, Pieper D, et al. Reporting according to the preferred reporting items for systematic reviews and Meta-analyses for abstracts (PRISMA-A) depends on abstract length[J]. *J Clin Epidemiol*, 2023, 154: 167-177.
- 13 王子怡, 卢存存, 黄家艺, 等. 中文期刊发表的预测模型系统评价文献调查与评价: 方法学质量和报告质量[J]. *协和医学杂志*, 2024, 15(4): 927-935. [Wang ZY, Lu CC, Huang JY, et al. Investigation and evaluation of systematic reviews of prediction models published in Chinese journals: methodological and reporting quality[J]. *Medical Journal of Peking Union Medical College Hospital*, 2024, 15(4): 927-935.]
- 14 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.
- 15 Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both[J]. *BMJ*, 2017, 358: j4008.
- 16 Fernando SM, Tran A, Cheng W, et al. Pre-arrest and intra-arrest prognostic factors associated with survival after in-hospital cardiac arrest: systematic review and Meta-analysis[J]. *BMJ*, 2019, 367: i6373.
- 17 Ferreira RT, Reis MQ, Lange MHP, et al. What evidence syntheses reveal about PROSPERO, INPLASY, OSF, the Research Registry, and protocols.io: a Meta-research study[J]. *Front Res Metr Anal*, 2026, 11: 1738112.
- 18 Hartling L, Featherstone R, Nuspl M, et al. Grey literature in systematic reviews: a cross-sectional study of the contribution of non-English reports, unpublished studies and dissertations to the results of Meta-analyses in child-relevant reviews[J]. *BMC Med Res Methodol*, 2017, 17(1): 64.
- 19 Xu C, Yu T, Furuya-Kanamori L, et al. Validity of data extraction in evidence synthesis practice of adverse events: reproducibility study[J]. *BMJ*, 2022, 377: e069155.
- 20 Nikolakopoulou A, Mavridis D, Salanti G. How to interpret Meta-analysis models: fixed effect and random effects Meta-analyses[J]. *Evid Based Ment Health*, 2014, 17(2): 64.
- 21 Hong C, Salanti G, Morton SC, et al. Testing small study effects in multivariate Meta-analysis[J]. *Biometrics*, 2020, 76(4): 1240-1250.
- 22 Hakoum MB, Anouti S, Al-Gibbawi M, et al. Reporting of financial and non-financial conflicts of interest by authors of systematic reviews: a methodological survey[J]. *BMJ Open*, 2016, 6(8): e011997.
- 23 Hansford HJ, Richards GC, Page MJ, et al. Reporting health and medical research[J]. *BMJ Evid Based Med*, 2024, 29(6): 358-362.
- 24 张强, 赵文霞, 王子怡, 等. 结局测量工具系统评价的报告规范 (PRISMA-COSMIN for OMI 2024) 解读[J]. *中国全科医学*, 2026, 29(13): 1766-1775. [Zhang Q, Zhao WX, Wang ZY, et al. Interpretation of the reporting guideline for systematic reviews of outcome measurement instruments (PRISMA-COSMIN for OMI 2024) [J]. *Chinese General Practice*, 2026, 29(13): 1766-1775.]
- 25 Chin JM, Grown B, Sebastian J, et al. The transparency and reproducibility of systematic reviews in forensic science[J]. *Forensic Sci Int*, 2022, 340: 111472.
- 26 Martinez J, Manjaly C, Manjaly P, et al. Janus kinase inhibitors and adverse events of acne: a systematic review and Meta-analysis[J]. *JAMA Dermatol*, 2023, 159(12): 1339-1345.
- 27 石舒原, 周庆欣, 孙凤, 等. 糖尿病肾病可变风险因素证据图谱构建[J]. *中华流行病学杂志*, 2025, 46(11): 2055-2064. [Shi SY, Zhou QX, Sun F, et al. Construction of evidence graph of modifiable risk factors for diabetic nephropathy[J]. *Chinese Journal of Epidemiology*, 2025, 46(11): 2055-2064.]
- 28 Kim JH, Kim JY, Lee S, et al. Environmental risk factors, protective factors, and biomarkers for postpartum depressive symptoms: an umbrella review[J]. *Neurosci Biobehav Rev*, 2022, 140: 104761.
- 29 Chai KEK, Lines RLJ, Gucciardi DF, et al. Research Screener: a machine learning tool to semi-automate abstract screening for systematic reviews[J]. *Syst Rev*, 2021, 10(1): 93.
- 30 Courvoisier DS, Buitrago-Garcia D, Buclin CP, et al. Beyond human gold standards: a multimodel framework for automated abstract classification and information extraction[J]. *Res Synth Methods*, 2026, 17(2): 365-377.
- 31 Marques-Cruz M, Pinto F, Vieira RJ, et al. Use of artificial intelligence to support the assessment of the methodological quality of systematic reviews[J]. *J Clin Epidemiol*, 2025, 187: 111944.
- 32 Xia D, Lai H, Zhao W, et al. Assessing risk of bias of cohort studies with large language models[J]. *Res Synth Methods*, 2025, 16(6): 990-1004.

收稿日期: 2026 年 04 月 14 日 修回日期: 2026 年 07 月 20 日
本文编辑: 杨室淞 曹 越

引用本文: 张亚玲, 张强, 邓秀秀, 等. 预后因素研究系统评价/Meta 分析的方法学质量评价工具 (AMSTAR-PF) 解读[J]. 医学新知, 2026, 36(9): 963-970. DOI: 10.12173/j.issn.1004-5511.202604065.

Zhang YL, Zhang Q, Deng XX, et al. Interpretation of a measurement tool to assess systematic reviews/Meta-analyses of prognostic factor studies (AMSTAR-PF)[J]. *Yixue Xinzhi Zazhi*, 2026, 36(9): 963-970. DOI: 10.12173/j.issn.1004-5511.202604065.