

STAR 指南术语体系 (二)：临床问题确定、证据检索与评价、推荐意见形成

朱迪¹，李昊东¹，王子君²，史乾灵³，刘辉²，秦艺珊²，姚媛媛²，
李哲玮¹，贺洪峰¹，田金徽²，葛龙^{1,4}，陈耀龙^{1,2,4,5,6,7,8}，
STAR 工作组方法与技术专业委员会

兰州大学¹ 公共卫生学院² 基础医学院循证医学中心³ 第一临床医学院

⁵ 健康数据科学研究院，兰州 730000

⁴ 甘肃省循证医学重点实验室，兰州 730000

⁶ 中国医学科学院循证评价与指南研究创新单元，兰州 730000

⁷ 中华医学杂志社指南与标准研究中心，兰州 730000

⁸ 世界卫生组织指南实施与知识转化合作中心，兰州 730000

通信作者：葛龙，E-mail：gelong2009@163.com

陈耀龙，E-mail：chevidence@lzu.edu.cn

【摘要】目的 介绍并分析指南制订过程中临床问题确定、证据检索与评价、推荐意见形成阶段的相关指南术语。**方法** 系统检索指南制订手册及相关方法学文献，检索时限为建库至 2024 年 10 月 25 日。针对纳入文献，提取上述三个阶段相关指南术语并进行标准化处理，通过共识会议进行讨论和评估，最终确定指南术语清单及其释义。**结果** 共纳入 30 部指南制订手册和 15 篇方法学文献，对指南制订过程中临床问题确定、证据检索与评价、推荐意见形成阶段术语进行系统梳理，共确定了 23 条核心术语。**结论** 建议建立名称统一、定义明确、符合我国本土语言环境和使用习惯的规范化、科学化指南术语体系，同时基于该体系加强对指南制订者和使用者的术语培训，以加深其对指南术语的正确理解和恰当应用。

【关键词】 指南；术语；标准化；STAR

【中图分类号】 R181.2；C04 **【文献标志码】** A **【文章编号】** 1674-9081(2025)03-0756-09

DOI: 10.12290/xhyxzz.2025-0169

STAR Guideline Terminology(II) : Clinical Question Formulation, Evidence Retrieval and Appraisal, and Recommendation Development

ZHU Di¹, LI Haodong¹, WANG Zijun², SHI Qianling³, LIU Hui², QIN Yishan², YAO Yuanyuan²,
LI Zhewei¹, HE Hongfeng¹, TIAN Jinhui², GE Long^{1,4}, CHEN Yaolong^{1,2,4,5,6,7,8},
Methodology and Technology Specialisation Committee of the STAR Working Group

¹School of Public Health, ²Evidence-based Medicine Center, School of Basic Medical Sciences,

³The First School of Clinical Medicine, ⁵Institute of Health Data Science, Lanzhou University, Lanzhou 730000, China

⁴Key Laboratory of Evidence-based Medicine of Gansu Province, Lanzhou 730000, China

⁶Research Unit of Evidence-based Evaluation and Guidelines, Chinese Academy of Medical Sciences, Lanzhou 730000, China

基金项目：中国医学科学院医学与健康科技创新工程-循证评价与指南研究创新单元（2021RU017）；王宽诚教育基金会

引用本文：朱迪，李昊东，王子君，等. STAR 指南术语体系 (二)：临床问题确定、证据检索与评价、推荐意见形成 [J]. 协和医学杂志, 2025, 16 (3): 756-764. doi: 10.12290/xhyxzz.2025-0169.

⁷Guidelines and Standards Research Center, Chinese Medical Association Publishing House, Lanzhou 730000, China

⁸WHO Collaborating Center for Guideline Implementation and Knowledge Translation, Lanzhou 730000, China

Corresponding authors: GE Long, E-mail: gelong2009@163.com

CHEN Yaolong, E-mail: chevidence@lzu.edu.cn

【Abstract】Objective To introduce and analyze guideline terminology related to clinical question formulation, evidence retrieval and appraisal, and recommendation development. **Methods** A systematic search was conducted in guideline development manuals and relevant methodological literature, covering publications up to October 25, 2024. Terminology related to the three aforementioned stages of related to guideline development was extracted from the included literature, standardized, and refined through consensus meetings to finalize a comprehensive terminology list and definitions. **Results** A total of 30 guideline development manuals and 15 methodological articles were included, and 23 core terms were identified. **Conclusions** It is recommended to develop a standardized and scientifically sound guideline terminology system with unified naming, clear definitions, and alignment with the linguistic environment and usage habits in China. At the same time, it is essential to strengthen terminology training for both guideline developers and users based on this system, in order to deepen their correct understanding and proper application of guideline terminology.

【Key words】 guidelines; terminology; standardization; Scientific, Transparent and Applicable Rankings

Funding: CAMS Innovation Fund for Medical Sciences-Research Unit of Evidence-based Evaluation and Guidelines (2021RU017); K. C. Wong Education Foundation, Hong Kong, China

Med J PUMCH, 2025, 16(3):756–764

为进一步规范临床实践指南（下文简称“指南”）术语的使用，构建完备的指南术语体系，指南科学性、透明性和适用性评级工具（Scientific, Transparent and Applicable Rankings, STAR）工作组方法与技术专业委员会对当前已有的指南制订手册和指南相关方法学文献进行系统梳理^[1]，撰写了STAR指南术语系列文章，首篇梳理并报告了指南规划与启动阶段的27条核心术语^[2]，本文作为该系列第二篇文章，将承接前文，重点围绕指南临床问题确定、证据检索与评价、推荐意见形成阶段的相关指南术语进行系统归纳与阐释。

在指南制订过程中，临床问题确定、证据检索与评价和推荐意见形成三个阶段是构建指南主体内容的关键步骤^[3]。临床问题确定基于一线临床医务工作者与患者当前最优先的需求，明确指南制订的方向和重点，提出指南能够解决临床实践中的核心问题；证据检索与评价通过系统检索、筛选并汇总与临床问题相关的当前最佳证据，为推荐意见的形成提供可信的科学依据；推荐意见形成则是在充分考虑证据质量、患者价值观与偏好、资源利用、利弊平衡等相关因素后，将复杂的证据转化为简洁明确的推荐意见，为医务工作者提供具体可行的实践指导，帮助其作出当前最佳临床决策。

1 资料与方法

1.1 检索策略

1.1.1 指南制订手册的检索、术语的收集与遴选、术语定义的确定

关于“指南制订手册的检索”“术语的收集与遴选”“术语定义的确定”过程与原则，详细内容可参见《STAR 指南术语体系 (一): 规划与启动》^[2] 方法学部分。

1.1.2 指南制订方法学文献的检索

关于指南临床问题确定、证据检索与评价、推荐意见形成阶段的术语，STAR 工作组系统检索中国知网、万方数据知识服务平台和 PubMed 数据库，检索时限为建库至 2024 年 10 月 25 日。检索过程采取主题词和自由词相结合的方式，中文检索词包括“指南”“标准”“临床问题”“证据”“推荐意见”等，英文检索词包括“Guideline *”“Method *”“Standard *”“Clinical question *”“Evidence”“Recommendation *”等。以 PubMed 为例，检索策略详见图 1。

1.2 文献纳入与排除标准

本研究纳入指南制订手册和相关方法学文献中包含临床问题确定、证据检索与评价、推荐意见形

| | |
|--|-------------------------------------|
| #1 "Practice Guidelines as Topic" [MeSH] | #8 "Manual" [Title] |
| #2 "Guideline * " [Title] | #9 "Clinical question * " [Title] |
| #3 #1 OR #2 | #10 "Evidence" [Title] |
| #4 "Method * " [Title] | #11 "Recommendation * " [Title] |
| #5 "Standard * " [Title] | #12 "Synthesis" [Title] |
| #6 "Principle" [Title] | #13 #4-#12/OR |
| #7 "Handbook" [Title] | #14 #3 AND #13 |

图 1 PubMed 检索策略

Fig. 1 Search strategy of PubMed

成阶段的相关术语，发表语言限制为中文或英文；排除同一机构重复发表的文献及无法获取全文的文献。

2 结果

2.1 检索结果与术语纳入情况

基于获取的 90 部指南制订手册及 4573 篇指南制订方法学文献，依据文献纳入与排除标准及术语收集与遴选原则，本研究重点围绕临床问题确定、证据检索与评价、推荐意见形成相关指南术语进行梳理与汇总，共纳入 30 部指南制订手册和 15 篇方法学文献，并对相关指南术语进行提取、去重与标准化处理，共获得 54 条术语。结合术语使用频次与 STAR 工作组实践经验，经 STAR 工作组方法与技术专业委员会审议，最终纳入 23 条核心术语，并按照指南的标准制订流程顺序呈现术语，详见表 1。

2.2 临床问题确定

2.2.1 人群、干预、对照和结局框架

中文别称：PICO 框架

英文名称：Population, Intervention, Comparison and Outcome (PICO) framework

释义：PICO 框架是循证医学中用于构建临床问题和指导医学研究的一种有效方式，能够帮助指南制订者将复杂模糊的临床问题转化为要点明确的结构性

问题，便于后续系统检索、证据纳入等指南制订步骤的推进。其具体含义如下^[3-4]：

P (Patient/Population)：指患者人群或研究对象，应明确研究的目标人群特征，包括年龄、性别、疾病类型、病情严重程度等。

I (Intervention/Exposure)：指干预措施或暴露因素，是研究中对目标人群实施的诊断、治疗、预防等措施，或某种暴露情况，如药物治疗、手术操作、康复训练、生活方式干预等。

C (Comparator/Control)：指比较措施或对照措施，是与干预措施或暴露因素相对应的干预或状态。对照措施可以是标准治疗、安慰剂、无干预、其他治疗方法等。

O (Outcome)：指在研究对象个体水平上观察、测量的结果或结局，用于评价 I 与 C 的差异，从而反映干预措施或暴露因素对研究对象的影响。结局可以是疾病治愈、患者死亡、症状改善程度、生活质量改变水平等。

2.2.2 临床问题

英文名称：clinical question

释义：临床问题是指临床实践中与具体疾病或患者健康状况、诊断、治疗、预防及预后等相关的问题，主要来源于对指南使用者（特别是一线临床医务人员）的调研，也可从当前文献（相关指南、系统评价或临床研究）获得，根据其含义可分为背景问题和前景问题^[3,5]。

2.2.3 背景问题

英文名称：background question

释义：背景问题是与特定疾病、患者群体、干预措施或指南主题相关的一般性知识问题。背景问题一般情况下可为指南制订提供必要的背景信息支持，通常涉及疾病的定义、病因、病理生理机制、流行病学特征等^[6-7]。

表 1 临床问题确定、证据检索与评价、推荐意见形成阶段相关术语

Tab. 1 Terms associated with the clinical question formulation, evidence retrieval and appraisal and recommendation development of guideline

| 阶段 | 术语名称（频次） |
|---------|--|
| 临床问题确定 | 人群、干预、对照和结局框架 (6)；临床问题 (5)；背景问题 (4)；前景问题 (10) |
| 证据检索与评价 | 证据 (1)；文献检索 (5)；纳入与排除标准 (3)；系统评价 (6)；证据质量 (6)；证据等级 (1)；牛津大学循证医学中心证据分级系统 [*] ；推荐分级的评估、制订与评价 (6)；偏倚风险评估 (4)；结果总结表 (2)；GRADE 证据概要表 (2) |
| 推荐意见形成 | 推荐意见 (3)；推荐意见强度 (11)；强推荐 [*] ；弱推荐 (2)；良好实践声明 (3)；从证据到决策框架 (3)；患者价值观与偏好 (6)；共识方法 (1) |

^{*}指南制订手册中未提及，由 STAR 工作组补充的术语

2.2.4 前景问题

中文别称: 关键问题

英文名称: foreground question; key question; key issue

释义: 前景问题与指南主题紧密相关, 是指南制订过程中应解决的核心问题, 需要基于当前最佳证据形成推荐意见。通常采用 PICO 框架明确其涵盖的关键要素, 以便后续证据检索与系统评价过程的开展^[7-9]。

2.3 证据检索与评价

2.3.1 证据

英文名称: evidence

释义: 指南背景下的证据是指经系统收集和评估后为推荐意见形成提供支持的信息, 包括科学研究(如系统评价、随机对照试验等)和专家证据(特定领域知识渊博或技术熟练的人员所提供的观察结果和经验反馈)^[5,10-11]。

2.3.2 证据检索

2.3.2.1 文献检索

英文名称: literature search

释义: 文献检索是指为回答具体临床问题(主要回答 PICO 问题), 依据预设的系统化检索策略, 从多个数据库和信息源中识别并获取相关证据的过程。这一过程通常以对书目数据库(部分情况下也包括全文数据库)的检索为主, 以对搜索引擎(如谷歌学术、百度学术等)和灰色文献(如未发表的研究、会议记录、政府报告等)的检索为补充, 同时对纳入文献的参考文献再次进行引文检索, 以确保全面识别所有相关研究^[12]。区别于系统评价的文献检索, 指南的文献检索范畴更广泛, 不仅包括原始研究, 还包括系统评价和其他相关指南、注册数据库等资料。

2.3.2.2 纳入与排除标准

英文名称: inclusion and exclusion criteria

释义: 纳入与排除标准是用于筛选和确定检索出的文献是否可作为证据支持推荐意见形成的标准^[7,10]。其可从以下方面进行设置: (1) 研究关注的结局指标; (2) 研究类型(如仅纳入系统评价和随机对照试验或允许纳入多种类型的研究等); (3) 具体研究环境; (4) 研究对象或亚组人群; (5) 研究发表的国家/地区; (6) 研究发表的时间范围; (7) 出版物语言等。

2.3.2.3 系统评价

英文名称: systematic review

释义: 系统评价是指针对某一具体问题, 系统、

全面收集已发表或未发表的相关研究, 采用严格评价文献的原则和方法, 筛选出符合质量标准的文献, 对其进行定性或定量合成, 得出当前最佳综合结论的文献综合评价办法。系统评价可以是定性的(定性系统评价, qualitative systematic review), 也可以是定量的(定量系统评价, quantitative systematic review), 即包含荟萃分析(Meta-analysis, Meta 分析)过程^[5,12]。

指南中的推荐意见基于系统评价形成, 其制订过程中涉及的系统评价包括两种: 一种是指经系统检索与文献筛选后纳入能够回答临床问题的具体文献类型, 即系统评价研究; 另一种是指经系统检索与文献筛选后未获得可用的系统评价研究, 此时需对当前已有相关研究进行文献综合与评价, 以制作新的系统评价回答临床问题, 即系统评价过程。

2.3.3 证据评价

2.3.3.1 证据质量

英文名称: evidence quality; quality of evidence

释义: 证据质量是指在回答指南临床问题(即做出决策或提出建议)时, 综合考虑证据对于某个关键结局指标的总体可信度^[13-14]。

2.3.3.2 证据质量分级

中文别称: 证据等级

英文名称: level of evidence quality

释义: 证据质量分级是指将证据根据其质量高低进行分级的结果。在指南制订过程中, 常见的证据质量分级体系包括: 推荐分级的评估、制订与评价(Grading of Recommendations Assessment, Development and Evaluation, GRADE)分级系统和英国牛津大学循证医学中心(Oxford Centre for Evidence-based Medicine, OCEBM)证据分级系统^[15-16]。

2.3.3.3 牛津大学循证医学中心证据分级系统

英文名称: Oxford Centre for Evidence-based Medicine (OCEBM) Levels of Evidence

释义: OCEBM 证据分级系统是用于评估和分类医学研究证据质量的体系, 该系统针对不同的研究类型进行证据等级划分。其中, 2009 版 OCEBM 证据分级系统对不同的证据等级匹配了相应的推荐强度, 2011 版仅对证据等级的划分标准进行了更新^[17]。

2.3.3.4 推荐分级的评估、制订与评价

英文名称: Grading of Recommendations Assessment, Development and Evaluation (GRADE)

释义: GRADE 是当前证据质量和推荐强度分级的国际标准之一, 为系统评价和指南提供证据质量评价体系的同时, 也为指南中推荐强度的评级提供了系

统方法。GRADE 通过系统化方法对健康干预措施的证据体质量、利弊平衡、患者价值观与偏好以及资源和成本影响进行综合评估，并据此对推荐意见的证据质量和推荐强度进行分级^[13,18]。GRADE 证据质量分级包括 5 个降级因素（偏倚风险、不一致性、间接性、不精确性、发表偏倚）和 3 个升级因素（大效应量、剂量效应关系、负偏倚）。

2.3.3.5 偏倚风险评估

别称：方法学质量评价

英文名称：risk of bias assessment

释义：偏倚风险评估是指对科学的研究设计和实施进行系统分析，以识别造成系统误差（即偏倚）使效应评估不够精确的因素^[19-20]。指南制订过程中常用偏倚风险评估工具汇总详见表 2。

表 2 常用偏倚风险评估工具汇总表^[21]

Tab. 2 Summary of commonly used risk of bias assessment tools

| 研究类型 | 偏倚风险评估工具 |
|---------------|---|
| 系统评价和 Meta 分析 | AMSTAR、AMSTAR2、ROBIS 工具、SIGN 方法学清单、JBI 清单 |
| 随机对照试验研究 | RoB 工具、RoB 2.0 工具、JBI 清单、SIGN 方法学清单 |
| 非随机干预性研究 | ROBINS-I 工具、JBI 清单 |
| 诊断准确性试验研究 | QUADAS 工具、QUADAS-2 工具、JBI 清单、SIGN 方法学清单 |
| 队列研究、病例对照研究 | NOS、JBI 清单、NIH 工具 |
| 横断面研究 | AHRQ 清单、JBI 清单、NIH 工具 |

AMSTAR (A Measurement Tool to Assess Systematic Reviews)：一种评估系统评价的测量工具；ROBIS (Risk of Bias in Systematic Review)：系统评价偏倚风险；SIGN (Scottish Intercollegiate Guidelines Network)：苏格兰校际指南网络；JBI (Joanna Briggs Institute)：乔安娜·布里格斯学院；RoB (Risk of Bias)：Cochrane 偏倚风险评价工具；ROBINS-I (Risk Of Bias In Non-randomised Studies of Interventions)：非随机干预性研究偏倚风险；QUADAS (Quality Assessment of Diagnostic Accuracy Studies)：诊断准确性研究的质量评价；NOS (Newcastle-Ottawa Scale)：纽卡斯尔-渥太华量表；NIH (National Institutes of Health)：美国国立卫生研究院；AHRQ (Agency for Healthcare Research and Quality)：美国卫生保健和质量研究机构

2.3.3.6 结果总结表

中文别称：SoF 表；结果摘要表

英文名称：summary of findings (SoF) table

释义：结果总结表是 GRADE 工作组研发的一种以简洁、易懂的表格形式呈现系统评价或指南中关键结局指标的总结工具。其提供了关于证据质量、干预措施效应大小以及主要结局数据的关键信

息，旨在为决策者和指南使用者提供支持推荐意见的关键信息的简洁总结。结果总结表通常按结局指标排列，对每种结局及其效应估计进行总结，并简要说明证据质量的影响因素^[7,12-13]。结果总结表的示例详见表 3^[22]。

2.3.3.7 GRADE 证据概要表

中文别称：EP 表；GRADE 概况表

英文名称：evidence profiles (EP)；GRADE profiles

释义：证据概要表同样是 GRADE 工作组研发的一种用于展示现有证据质量、影响证据等级的判断依据，以及不同管理策略对关键结局指标影响的工具。其为指南制订者和使用者提供了简洁且透明的证据总结，以支持相关临床决策。证据概要表除具有结果总结表的内容（即对每个结局的结果总结）外，还包含对证据质量每个决定因素的清晰评价^[13,18]。证据概要表的示例详见表 4^[22]。

2.4 推荐意见形成

2.4.1 推荐意见

英文名称：recommendation

释义：推荐意见是针对临床问题形成的可操作性指导建议，是指南的主体内容。推荐意见的形成既要基于当前可获得的最佳研究证据（系统评价），也要综合考虑资源利用、患者价值观与偏好、公平性和可及性等多方面因素^[3]。部分指南或共识的名称有时用推荐意见替代，但其本质仍是综合性指导文件^[23-24]。

2.4.2 推荐意见强度

中文别称：推荐意见等级

英文名称：strength of recommendation；grade of recommendation

释义：推荐意见强度是指遵守推荐意见对目标人群产生的利与弊的把握度，反映了指南推荐意见在临床实践中的适用性和优先级，是基于对证据质量、干预措施的利弊权衡、患者价值观与偏好、资源可及性等因素的综合评估。根据 GRADE 分级系统，推荐意见强度分为“强”和“弱”两个等级^[4-5,13]。

2.4.3 强推荐

英文名称：strong recommendation

释义：强推荐是指指南制订者确信某种干预措施的实施效果利大于弊或弊大于利时，对该干预措施或临床决策给出的推荐意见等级。强推荐意味着指南制订者对该推荐意见的实施具有较强信心，认为其利弊权衡明确^[13,25]。

表 3 结果总结表示例^[22]
Tab. 3 An example of summary of findings table^[22]

临床问题: 甲硫哒嗪对比非典型抗精神病药物治疗精神分裂症

患者或人群: 精神分裂症患者

实施机构: 3 篇研究在非医院场所开展, 其他研究在医院场所开展; 5 篇研究在中国开展, 其他研究在北美和欧洲开展

干预措施: 甲硫哒嗪

对照措施: 非典型抗精神病药物

| 结局指标 | 绝对效应 (95% CI) | | 相对效应 (95% CI) | 样本量 (研究数量) | (GRADE) 证据质量 | 备注 |
|----------------------|-------------------------|--------------------|---------------------|---------------|-------------------------|----|
| | 非典型抗精神病药物 组风险 (每千人)* | 甲硫哒嗪组风险 (每千人) | | | | |
| 自杀死亡 (6 周随访) | 14/1000 | 5/1000 (0~113) | RR 0.34 (0.01~8.27) | 144 (1RCT) | ⊕⊕○○低 ^{a,b} | |
| 症状未改善或加重 (短期随访) | 549/1000 | 549/1000 (445~686) | RR 1.00 (0.81~1.25) | 203 (3RCTs) | ⊕○○○极低 ^{a,b,c} | |
| 不良反应: 低血压 (1~3 个月随访) | 159/1000 | 250/1000 (133~468) | RR 1.58 (0.84~2.95) | 162 (2RCTs) | ⊕⊕○○低 ^{a,b} | |
| 不良反应: 运动障碍 (3 个月随访) | 227/1000 | 66/1000 (27~155) | RR 0.29 (0.12~0.68) | 180 (3RCTs) | ⊕⊕⊕○中 ^a | |
| 不良反应: 嗜睡和困倦 (3 个月随访) | 778/1000 | 443/1000 (194~996) | RR 0.57 (0.25~1.28) | 18 (1RCT) | ⊕⊕○○低 ^{a,b} | |
| 不良反应: 觉醒和失眠 (3 个月随访) | 172/1000 | 269/1000 (100~717) | RR 1.56 (0.58~4.16) | 59 (2RCTs) | ⊕⊕○○低 ^{a,b} | |

GRADE 证据质量分级:

高质量: 非常有把握观察值接近真实值

中等质量: 对观察值有中等把握; 观察值有可能接近真实值, 但也有可能差别很大

低质量: 对观察值的把握有限; 观察值可能与真实值有很大差别

极低质量: 对观察值几乎没有把握; 观察值与真实值可能有极大差别

CI (confidence interval): 置信区间; RR (relative risk): 相对风险; GRADE (Grading of Recommendations Assessment, Development and Evaluation): 推荐分级的评估、制订与评价; RCT (randomized controlled trial): 随机对照试验; * 干预组的风险和 95% CI 是基于对照组的基线风险和干预措施的相对风险计算所得; ^a 偏倚风险降一级: 纳入研究未报告随机和分配隐藏; ^b 不精确性降一级: 纳入研究的事件发生数较少, 置信区间宽, 跨越无效线; ^c 不一致性降一级: $I^2 > 50\%$

表 4 GRADE 证据概要表示例^[22]
Tab. 4 An example of GRADE evidence profile

临床问题: 甲硫哒嗪对比非典型抗精神病药物治疗精神分裂症

实施机构: 3 篇研究在非医院场所开展, 其他研究在医院场所开展; 5 篇研究在中国开展, 其他研究在北美和欧洲开展

文献来源: Fenton M, Rathbone J, Reilly J, et al. Thoridazine for schizophrenia [J]. Cochrane Database Syst Rev, 2007 (3): CD001944.

| 研究数 | 证据质量评价 | | | | | | 事件发生数/样本量 (%) | | 效应量 | | 证据质量 |
|----------------------|--------|------------------|------------------|-----|------------------|------|----------------|----------------|---------------------|---------------------------|-------------------------|
| | 研究设计 | 偏倚风险 | 不一致性 | 间接性 | 不精确性 | 其他偏倚 | 甲硫哒嗪组 | 非典型抗精神病药物组 | 相对效应 (95% CI) | 绝对效应 (95% CI) | |
| 自杀死亡 (6 周随访) | | | | | | | | | | | |
| 1 | RCT | 降一级 ^a | 不降级 | 不降级 | 降一级 ^b | 不降级 | 0/71 (0) | 1/73 (1.4%) | RR 0.34 (0.01~8.27) | <9/1000 人 (<14 到 >100) | ⊕⊕○○低 ^{a,b} |
| 症状未改善或加重 (短期随访) | | | | | | | | | | | |
| 3 | RCT | 降一级 ^a | 降一级 ^c | 不降级 | 降一级 ^b | 不降级 | 55/101 (54.5%) | 56/102 (54.9%) | RR 1.00 (0.81~1.25) | 0/1000 人 (<104 到 >137) | ⊕○○○极低 ^{a,b,c} |
| 不良反应: 低血压 (1~3 个月随访) | | | | | | | | | | | |
| 2 | RCT | 降一级 ^a | 不降级 | 不降级 | 降一级 ^b | 不降级 | 20/80 (25.0%) | 13/82 (15.9%) | RR 1.58 (0.84~2.95) | >92/1000 人 (<25 到 >09) | ⊕⊕○○低 ^{a,b} |
| 不良反应: 运动障碍 (3 个月随访) | | | | | | | | | | | |
| 3 | RCT | 降一级 ^a | 不降级 | 不降级 | 不降级 | 不降级 | 6/92 (6.5%) | 20/88 (22.7%) | RR 0.29 (0.12~0.68) | <161/1000 人 (<73 到 <200) | ⊕⊕⊕○中 ^a |
| 不良反应: 嗜睡和困倦 (3 个月随访) | | | | | | | | | | | |
| 1 | RCT | 降一级 ^a | 不降级 | 不降级 | 降一级 ^b | 不降级 | 4/9 (44.4%) | 7/9 (77.8%) | RR 0.57 (0.25~1.28) | <334/1000 人 (>218 到 <583) | ⊕⊕○○低 ^{a,b} |
| 不良反应: 觉醒和失眠 (3 个月随访) | | | | | | | | | | | |
| 2 | RCT | 降一级 ^a | 不降级 | 不降级 | 降一级 ^b | 不降级 | 8/30 (26.7%) | 5/29 (17.2%) | RR 1.56 (0.58~4.16) | >97/1000 人 (<72 到 >545) | ⊕⊕○○低 ^{a,b} |

CI、RR、RCT、GRADE: 同表 3; a、b、c: 同表 3

2.4.4 弱推荐

中文别称：有条件的推荐意见

英文名称：weak recommendation; conditional recommendation

释义：弱推荐是指推荐的干预措施利弊不确定，或无论证据质量高低均显示利弊相当的情况下，形成的推荐意见等级。弱推荐意味获益与风险和负担之间的平衡存在较大的不确定性。在此情况下，医护人员的临床决策需根据患者的具体情况而定，在确保向患者清晰、全面地解释潜在获益和风险的前提下，考虑患者的价值观与偏好、资源可及性、公平性等因素后，进行更慎重地评估，并与患者共同作出决策^[13,25]。

2.4.5 良好实践声明

中文别称：良好实践要点

英文名称：good practice statements; good practice points

释义：良好实践声明是指由于支持证据的间接性或其他原因，无法使用传统 GRADE 分级系统进行证据质量评级而采用的一种推荐意见类型。良好实践声明被认为是必要的实践指导，虽然缺乏直接的支持证据，但其实施的净效益是明确且具有高度确定性的^[26-28]。

2.4.6 从证据到决策框架

中文别称：EtD 框架

英文名称：Evidence to Decision (EtD) framework

释义：从证据到决策框架是由 GRADE 工作组研发的一种系统化和透明化的决策支持工具，旨在帮助决策者（如指南制订者、政策制定者等）在临床推荐、医保覆盖决策以及卫生系统或公共卫生建议和决策背景下，以结构化方式使用证据指导决策过程^[29-30]。

2.4.7 患者价值观与偏好

中文别称：患者偏好与价值观

英文名称：patients' values and preferences

释义：患者价值观与偏好是指患者及其照护者对于健康相关结局相对重要性的认知，以及他们对不同健康状态和干预措施的个人喜好，包括对健康问题的看法、在医疗保健过程中的倾向、需求、信念、期望和目标。在指南制订过程中，患者价值观与偏好是需要考虑的重要因素^[7,31-32]。

2.4.8 共识方法

英文名称：consensus methods

释义：共识方法是指南制订过程中，专家组为达

成统一决策、形成推荐意见而采取的方法或途径^[10,33]。经典的共识方法包括：共识会议法、名义群体法、德尔菲专家函询法及改良德尔菲法等^[34]。指南制订过程中的共识方法主要用于达成指南推荐意见的共识，此外也常用于确定临床问题，以及对缺乏证据的临床问题基于专家建议形成推荐意见。

3 讨论

对于临床问题确定、证据检索与评价和推荐意见形成三个阶段，当前指南术语主要存在以下问题：（1）部分使用频率较高的术语在指南背景下缺乏明确、统一的定义，如“证据”和“推荐意见”等；（2）某些术语在指南背景下具有更宽泛的含义，如“文献检索”“纳入与排除标准”，以及“系统评价”；（3）已有定义的相似指南术语易于混淆，如“临床问题”“前景问题”与“背景问题”。指南制订过程中收集的“临床问题”来源于指南使用者的调研和文献梳理，可按照其内涵分为“背景问题”与“前景问题”，“背景问题”作为一般性知识问题通常不是指南关注的重点问题，可在指南背景、推荐原理中进行说明，而“前景问题”则是最终纳入指南需要解决的核心问题，应基于当前最佳证据形成推荐意见。例如，关于“某疾病的诊断标准是什么”这一临床问题，若该疾病的诊断标准是具有明确临床共识的“金标准”，此临床问题则为“背景问题”；若该疾病诊断标准尚未达成统一共识，需要基于现有证据形成指南推荐意见，那么该临床问题为“前景问题”。此外，在指南制订与评价过程中，对于上述三个阶段的指南术语，相当一部分指南制订者和使用者虽然知晓，但并不能正确理解术语含义并在实践中恰当应用。如针对“弱推荐”，其核心含义是“谨慎推荐”而非“不推荐”，弱推荐承认并呈现了推荐意见利弊平衡的“不确定性”，更强调个体化决策和患者参与，能够指导证据不足情况下的临床决策，促进个体化决策和共享决策，并为临床实践提供灵活性。但部分指南制订者和使用者将“弱”理解为“不重要”或“可选”，从而在指南实践过程中更倾向于制订和使用“强推荐”，导致“弱推荐”的不确定性，反而阻碍了指南制订者作出类似推荐，同时也影响了临床决策者应用此类推荐。因此，笔者建议在中文翻译体系中，将“强推荐”翻译为“推荐使用”，将“弱推荐”翻译为“考虑使用”；在指南推荐意见的具体表述中，如果某项干预措施是弱推荐，则可使用

措辞“考虑使用某药物或某手术”。

针对上述情况, STAR 工作组方法与技术专业委员会认为应建立名称统一、定义明确、符合我国本土语言环境和使用习惯的规范化、标准化指南术语体系。同时, 基于该体系加强对指南制订者和使用者的术语培训, 以加深其对指南术语的正确理解和恰当应用。

作者贡献: 朱迪负责论文设计和初稿撰写; 李昊东负责论文初稿撰写; 王子君、史乾灵、刘辉、秦艺珊、姚媛媛、李哲玮、贺洪峰参与论文修订; 田金徽、葛龙、陈耀龙负责研究设计、论文指导和审校。

利益冲突: 所有作者均声明不存在利益冲突

志谢: 感谢香港浸会大学中国 EQUATOR 中心玉炫、兰州大学基础医学院循证医学中心周奇对于本文撰写提供的帮助。

参 考 文 献

- [1] Yang N, Liu H, Zhao W, et al. Development of the scientific, transparent and applicable rankings (STAR) tool for clinical practice guidelines [J]. Chin Med J (Engl), 2023, 136 (12): 1430-1438.
- [2] 李哲玮, 史乾灵, 刘辉, 等. STAR 指南术语体系 (一): 规划与启动 [J]. 协和医学杂志, 2025, 16 (1): 216-223.
- [3] Li Z W, Shi Q L, Liu H, et al. STAR guideline terminology (I): planning and launching [J]. Med J PUMCH, 2025, 16 (1): 216-223.
- [4] 陈耀龙, 杨克虎, 王小钦, 等. 中国制订/修订临床诊疗指南的指导原则 (2022 版) [J]. 中华医学杂志, 2022, 102 (10): 697-703.
- Chen Y L, Yang K H, Wang X Q, et al. Guiding principles for the development/revision of clinical diagnosis and treatment guidelines in China (2022 edition) [J]. Natl Med J China, 2022, 102 (10): 697-703.
- [5] Dumonceau J M, Hassan C, Riphauw A, et al. European society of gastrointestinal endoscopy (ESGE) guideline development policy [J]. Endoscopy, 2012, 44 (6): 626-629.
- [6] 孙鑫, 杨克虎. 循证医学 [M]. 第 2 版. 北京: 人民卫生出版社, 2021.
- Sun X, Yang K H. Evidence-based medicine [M]. 2nd ed. Beijing: People's Medical Publishing House, 2021.
- [7] Li H, Chen Y L, Wei D, et al. Development and revision methods of clinical practice guideline (CGP) on traditional Chinese medicine (integrated medicine): Clinical issues construction [J]. China J Tradit Chin Med Pharm, 2016, 31 (6): 2202-2205.
- [8] World Health Organization. WHO handbook for guideline development [M]. 2nd ed. Geneva: World Health Organization, 2014.
- [9] Ito Y, Tawada M, Yuasa H, et al. New Japanese society of dialysis therapy guidelines for peritoneal dialysis [J]. Contrib Nephrol, 2019, 198: 52-61.
- [10] Diekemper R L, Patel S, Mette S A, et al. Making the GRADE: CHEST updates its methodology [J]. Chest, 2018, 153 (3): 756-759.
- [11] National Institute for Health and Care Excellence. Developing NICE guidelines: the manual [M/OL]. (2014-10-31) [2025-02-19]. <https://www.nice.org.uk/process/pmg20/resources/developing-nice-guidelines-the-manual-pdf-72286708700869>.
- [12] Higgins J P T, Green S. Cochrane handbook for systematic reviews of interventions version 5. 1. 0 [EB/OL]. [2025-02-19]. <http://handbook-5-1.cochrane.org/>.
- [13] Schünemann H, Brożek J, Guyatt G, et al. GRADE handbook [EB/OL]. [2023-06-15]. <https://gdt.gradepro.org/app/handbook/handbook.html>.
- [14] Guyatt G H, Oxman A D, Vist G E, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations [J]. BMJ, 2008, 336 (7650): 924-926.
- [15] Centre for Evidence-Based Medicine. Levels of evidence: an introduction [EB/OL]. [2025-02-19]. <https://www.cebm.ox.ac.uk/resources/levels-of-evidence/levels-of-evidence-introductory-document>.
- [16] Jacobs A K, Kushner F G, Ettinger S M, et al. ACCF/AHA clinical practice guideline methodology summit report: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines [J]. Circulation, 2013, 127(2): 268-310.
- [17] Center for Evidence Based Medicine. OCEBM levels of evidence [EB/OL]. [2025-02-19]. <https://www.cebm.ox.ac.uk/resources/levels-of-evidence/ocebml-levels-of-evidence>.
- [18] Guyatt G, Oxman A D, Akl E, et al. GRADE 指南: I. 导论: GRADE 证据概要表和结果总结表 [J]. 中国循证医学杂志, 2011, 11 (4): 437-445. Guyatt G, Oxman A D, Akl E, et al. GRADE guidelines: 1. Introduction to GRADE

- evidence profiles and summary of findings tables [J]. Chin J Evid Based Med, 2011, 11 (4): 437-445.
- [19] Davino-Ramaya C, Krause L K, Robbins C W, et al. Transparency matters: Kaiser Permanente's National Guideline Program methodological processes [J]. Perm J, 2012, 16 (1): 55-62.
- [20] Estonian Health Insurance Fund, University of Tartu, Republic of Estonia Ministry of Social Affairs. Estonian handbook for guidelines development 2020 [EB/OL]. [2025-02-19]. https://ravijuhend.ee/uploads/userfiles/Estonian_Handbook_for_Guidelines_Development_2020_copy.pdf.
- [21] 李柄辉, 訾豪, 李路遥, 等. 医学领域一次研究和二次研究的方法学质量(偏倚风险)评价工具 [J]. 医学新知, 2021, 31 (1): 51-58.
- Li B H, Zi H, Li L Y, et al. Methodological quality (risk of bias) assessment tools for primary and secondary medical studies: what are they and which is better? [J]. New Med, 2021, 31 (1): 51-58.
- [22] 陈耀龙. GRADE 在系统评价和实践指南中的应用 [M]. 第2版. 北京: 中国协和医科大学出版社, 2021.
- Chen Y L. Application of GRADE in system evaluation and practice guidelines [M]. 2nd ed. Beijing: Peking Union Medical College Press, 2021.
- [23] Iolascon A, Bianchi P, Andolfo I, et al. Recommendations for diagnosis and treatment of methemoglobinemia [J]. Am J Hematol, 2021, 96 (12): 1666-1678.
- [24] Shukla A, Rockey D C, Kamath P S, et al. Non-cirrhotic portal fibrosis/idiopathic portal hypertension: APASL recommendations for diagnosis and management [J]. Hepatol Int, 2024, 18 (6): 1684-1711.
- [25] Patton C, Bradfield L. ASTRO clinical practice guideline methodology guide [EB/OL]. [2025-02-19]. https://www.astro.org/ASTRO/media/ASTRO/Patient%20Care%20and%20Research/PDFs/ASTRO_GuidelineMethodology.pdf.
- [26] Dewidar O, Lotfi T, Langendam M W, et al. Good or best practice statements: proposal for the operationalisation and implementation of GRADE guidance [J]. BMJ Evid Based Med, 2023, 28 (3): 189-196.
- [27] Guyatt G H, Schünemann H J, Djulbegovic B, et al. Guideline panels should not GRADE good practice statements [J]. J Clin Epidemiol, 2015, 68 (5): 597-600.
- [28] Guyatt G H, Alonso-Coello P, Schünemann H J, et al. Guideline panels should seldom make good practice statements: guidance from the GRADE Working Group [J]. J Clin Epidemiol, 2016, 80: 3-7.
- [29] Andrews J, Guyatt G, Oxman A D, et al. GRADE guidelines: 14. Going from evidence to recommendations: the significance and presentation of recommendations [J]. J Clin Epidemiol, 2013, 66 (7): 719-725.
- [30] Alonso-Coello P, Schünemann H J, Moberg J, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction [J]. BMJ, 2016, 353: i2016.
- [31] 令娟, 姚亮, 王明霞, 等. 临床实践指南制订中患者价值观与偏好问卷设计考虑因素的系统评价 [J]. 中国循证医学杂志, 2018, 18 (12): 1359-1367.
- Ling J, Yao L, Wang M X, et al. Consideration of factors for questionnaires on patient values and preferences in clinical practice guidelines: a systematic review [J]. Chin J Evid Based Med, 2018, 18 (12): 1359-1367.
- [32] Scottish Intercollegiate Guidelines Network. A handbook for patient and carer representatives [M/OL]. [2025-02-19]. <https://www.sign.ac.uk/media/1052/sign100.pdf>.
- [33] 周奇, 王琪, 俞阳, 等. 临床实践指南制定中的共识方法 [J]. 药品评价, 2016, 13 (16): 13-17.
- Zhou Q, Wang Q, Yu Y, et al. Consensus methods in the development of clinical practice guidelines [J]. Drug Eval, 2016, 13 (16): 13-17.
- [34] 朱迪, 罗旭飞, 王子君, 等. 专家证据、专家意见及专家共识在指南制订中的区别与联系 [J]. 协和医学杂志, 2024, 15 (4): 936-941.
- Zhu D, Luo X F, Wang Z J, et al. The differences and connections among expert evidence, expert opinion and expert consensus in the development of guidelines [J]. Med J PUMCH, 2024, 15 (4): 936-941.

(收稿: 2025-02-28 录用: 2025-03-31 在线: 2025-04-02)

(本文编辑: 李玉乐)