

Core GRADE 方法学体系: GRADE 证据质量评价与指南制定核心方法^Δ

曾力楠^{1,2,3,4,5*}, 贺彤^{1,2,3,4}, 李海龙^{1,2,3,4}, 邹 锐^{1,2,3,4}, 杨春松^{1,2,3,4}, 刁 莎^{1,2,3,4}, 刘 丹^{1,2,3,4}, 陈 哲^{1,2,3,4}, 倪晓凤^{1,2,3,4}, 程 晓^{1,2,3,4}, 陈婧婧^{1,2,3,4}, 张伶俐^{1,2,3,4,5,6#} (1. 四川大学华西第二医院药学部/循证药学中心, 成都 610041; 2. 儿童药物四川省重点实验室, 成都 610041; 3. 国家药品监督管理局药物制剂体内外相关性技术研究重点实验室, 成都 610041; 4. 出生缺陷与相关妇儿疾病教育部重点实验室, 成都 610041; 5. 四川大学华西生物医学大数据研究院, 成都 610041; 6. 四川大学华西医院中国循证医学中心, 成都 610041)

中图分类号 R-3;R9 文献标志码 A 文章编号 1672-2124(2025)11-1281-06

DOI 10.14009/j.issn.1672-2124.2025.11.001



摘要 为解决推荐分级的评估、制订与评价(GRADE)分级标准在应用过程中出现的操作日趋复杂、方法间不一致的问题,有必要制定更为简洁、高效的证据质量评价体系,Core GRADE 方法学体系应运而生。Core GRADE 方法学体系提供了临床问题构建方法、证据质量评价方法(包括证据质量评价目标判定,证据不精确性、不一致性、偏倚风险、间接性、发表偏倚的评价方法)及从证据到推荐意见形成方法。同时,Core GRADE 方法学体系建议使用结果总结表对证据进行汇总,并介绍了指南制定者与卫生技术评估者将证据转化为推荐意见或决策的核心方法。Core GRADE 方法学体系为系统评价、临床指南及卫生技术评估中的证据质量评价与推荐意见形成提供了更高效、实用的方法。

关键词 GRADE; 系统评价; 临床指南; 卫生技术评估; 证据质量; 推荐意见

Core GRADE Methodological System: Core Approach for Evidence Quality Assessment and Guideline Development in GRADE^Δ

ZENG Linan^{1,2,3,4,5*}, HE Tong^{1,2,3,4}, LI Hailong^{1,2,3,4}, ZOU Kun^{1,2,3,4}, YANG Chunsong^{1,2,3,4}, DIAO Sha^{1,2,3,4}, LIU Dan^{1,2,3,4}, CHEN Zhe^{1,2,3,4}, NI Xiaofeng^{1,2,3,4}, CHENG Xiao^{1,2,3,4}, CHEN Jingjing^{1,2,3,4}, ZHANG Lingli^{1,2,3,4,5,6#} (1. Dept. of Pharmacy/Evidence-Based Pharmacy Center, West China Second University Hospital, Sichuan University, Chengdu 610041, China; 2. Children's Medicine Key Laboratory of Sichuan Province, Chengdu 610041, China; 3. NMPA Key Laboratory for Technical Research on Drug Products in Vitro and in Vivo Correlation, Chengdu 610041, China; 4. Key Laboratory of Birth Defects and Related Diseases of Women and Children, Sichuan University, Ministry of Education, Chengdu 610041, China; 5. West China Biomedical Big Data Center, West China Hospital, Sichuan University, Chengdu 610041, China; 6. Chinese Evidence-based Medicine Center, West China Hospital, Sichuan University, Chengdu 610041, China)

ABSTRACT To address the issues of increasing operational complexity and methodological inconsistencies encountered during the application of the Grading of Recommendations Assessment, Development and Evaluation (GRADE) framework. It is necessary to establish a more concise and efficient evidence quality assessment system, and the Core GRADE methodological system is developed. The Core GRADE methodological system provides methods for constructing clinical questions, evaluating evidence quality (including determining evidence quality objectives, and assessing imprecision, inconsistency, risk of bias, indirectness, and publication bias), and formulating recommendations from evidence. Meanwhile, the Core GRADE methodological system recommends to use summary of

Δ 基金项目:国家自然科学基金面上项目(No. 72474148)

* 主任药师,教授,博士,博士生导师。研究方向:循证药学研究与实践。E-mail:linanzeng@sina.com

通信作者:主任药师,博士,博士生导师。研究方向:循证药物决策与管理、循证临床药学研究与实践。E-mail:zhanglingli@scu.edu.cn

finding tables to summarize evidence and introduces the core methods for guideline developers and health technology assessment professionals in translating evidence into recommendations or decisions. The Core GRADE methodological system provides a more efficient and practical approach for evidence quality assessment and recommendation formulation in systematic reviews, clinical guidelines, and health technology assessments.

KEYWORDS GRADE; Systematic review; Clinical guideline; Health technology assessment; Evidence quality; Recommendation

1 前言

循证医学始终强调将最佳证据、临床经验与患者价值观相结合,从而做出最适宜于患者的医疗决策^[1]。证据质量评价是循证医学的核心内容之一。2000年,包括世界卫生组织(WHO)、Cochrane协作网、加拿大麦克马斯特大学在内的19个组织联合成立了推荐分级的评估、制订与评价(grading of recommendations assessment, development and evaluation, GRADE)工作组,旨在制定出统一、公认的证据质量评价方法,为评价临床研究证据质量以及基于证据制定指南推荐意见提供系统性的方法学体系。

2004年,首个GRADE方法学体系系列介绍在医学权威期刊《British Medical Journal》上发表^[2]。随后,GRADE工作组在《Journal of Clinical Epidemiology》上发表了详尽的方法学体系指南,全面制定了从证据质量评价到推荐意见形成的完整流程。GRADE一经推出,受到了全球的广泛认可,被WHO、Cochrane协作网、国际指南协作网(GIN)等全球120多个组织广泛采纳,成为系统评价、临床指南和卫生技术评估中证据质量评价与推荐意见制定的标准方法^[3]。

然而,随着GRADE方法学的不断发展,GRADE方法也逐渐显现其局限性:一方面,方法体系的持续细化导致证据质量评价的方法日趋复杂,给使用者带来了较高的学习与应用门槛;另一方面,使用者难以及时跟进更新的方法并准确把握证据质量的评估要点^[4]。为解决上述问题,提供更简洁、高效的证据质量评价工具,Core GRADE系列的作者梳理了GRADE方法的核心内容,并更新了部分方法,形成了Core GRADE方法学体系。

2 Core GRADE方法学体系简介

Core GRADE方法学体系针对单一干预措施与单一对照措施比较的“头对头”临床研究和临床决策场景,不仅适用于系统评价的设计、实施与分析,也贯穿于临床实践指南和卫生技术评估从证据到推荐意见的形成过程,见图1。

Core GRADE系列现已发表7篇文章:首篇文章阐述了Core GRADE方法学体系中构建临床问题的方法,证据质量评价及“从证据到推荐意见”的整体框架^[5]。第2篇文章介绍了证据质量评价目标的判定方法及证据不精确性的评价方法^[6]。第3篇文章探讨了证据质量评价中的不一致性评价及发现不一致性时的处理方法^[7]。第4篇文章阐述了原始研究的偏倚风险评估、发表偏倚评估及其对证据质量的影响,同时介绍了针对非随机干预研究(non-randomized studies of interventions, NRSI)在哪些情况下可升高证据质量^[8]。第5篇文章讨论了证据间接性的评估方法及发现间接性时的处理方法^[9]。第6篇文章介绍了结果总结表(summary of findings

tables, SoF表)的构建方法^[10]。第7篇文章则面向指南制定者与卫生技术评估者,介绍了如何基于证据形成推荐意见或卫生决策^[11]。下面将以上述7篇文章为主线,介绍Core GRADE方法学体系的核心内容。

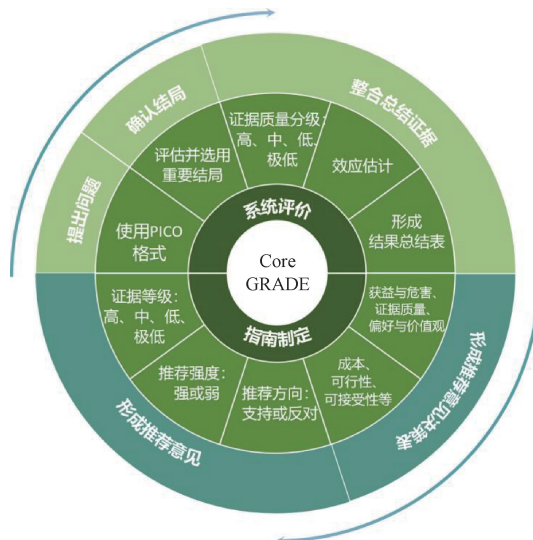


图1 Core GRADE整体框架

2.1 Core GRADE 1:临床问题构建、证据质量评价及“从证据到推荐意见”的方法概述

证据整合的第一步骤是运用PICO原则[人群(patient, P); 干预(intervention, I); 对照(comparison, C); 结局(outcome, O)]明确临床问题。例如,在我国射血分数正常的心力衰竭(heart failure with preserved ejection fraction, HFpEF)患者中,使用达格列净能否降低心血管死亡的风险?该案例中,P为我国HFpEF患者,I为达格列净,C为安慰剂,O为心血管死亡的发生率。

当可能存在亚组效应时(即干预措施的效果在不同患者特征或研究背景下可能存在真实差异),需在问题构建时针对不同亚组人群构建问题^[12]。例如,研究达格列净对降低HFpEF患者心血管死亡风险的效果时,基于病理生理机制[钠-葡萄糖耦联转运体2(SGLT2)抑制剂具有降糖及心肾保护的双重作用]推测其疗效在合并糖尿病的HFpEF患者中可能优于非糖尿病患者。因此,除总体人群问题外,可预先设定针对“合并糖尿病”与“未合并糖尿病”的2个亚组问题。

针对所构建的临床问题,通常检索最为相关的直接证据。当直接证据有限或完全没有时,考虑检索间接证据,可将间接证据用于循证决策。例如,在评估达格列净能否降低HFpEF患者心血管死亡风险时,若缺乏针对HFpEF人群的随机对照试验(RCT)证据,可检索间接证据。如人群间接性证据,基于相似病理生理机制,可考虑该药在射血分数降低的心力衰竭

(heart failure with reduced ejection fraction, HF_rEF) 患者中的 RCT 研究;或结局间接性证据,检索该药在 HF_pEF 患者中针对替代终点(如心力衰竭住院率)的有效性证据。

针对证据质量评价,Core GRADE 体系将“证据质量”定义为“真实效应位于决策阈值某一侧或特定范围内的可信程度”,并将证据质量等级分为高、中、低、极低 4 级。来自 RCT 的证据起始于高质量,而来自 NRSI 的证据,包括队列研究和病例对照研究,则起始于低质量。影响证据质量等级的因素包括证据不精确性、不一致性、偏倚风险、间接性和发表偏倚 5 个降级因素,以及剂量效应、大效应量 2 个用于 NRSI 证据的升级因素。GRADE 方法中的“可能存在的混杂因素”这一升级因素,因其实际评估较困难、适用性低,在 Core GRADE 体系

中被删除。

针对“从证据到推荐意见”的过程,Core GRADE 体系包含了 2 个基本步骤。首先,评估每个重要结局指标的证据质量并形成 SoF 表;然后,综合考虑证据质量、效应量大小、患者价值观等,形成具有明确方向(推荐或不推荐)和强度(强或弱)的推荐意见。对于系统评价团队,使用 SoF 表可清晰呈现每个重要结局的相对效应值、绝对效应值、证据质量等级及其升降级的具体理由。例如,《Percutaneous or surgical revascularization in patients with severe left main coronary artery disease in Latin America: A GRADE clinical practice guideline》中的 SoF 表,包括了结局、相对效应、绝对效应、证据质量及简要总结内容,见表 1^[13]。

表 1 SoF 表举例

结局[受试者数量/例(研究数量、类型)]	相对效应 (95%CI)	预期绝对风险(95%CI)			证据质量	简要总结
		冠状动脉旁路 移植术(CABG)	经皮冠状动脉 介入治疗(PCI)	差异		
30 d 全因死亡率[4 394(4 项 RCT)]	HR=1.09 (0.60~1.97)	1.0%	1.1% (0.6%~2.0%)	增加 0.1%(减少 0.4%至增加 1.0%)	⊕⊕⊕○ 中等	PCI 可能对 30 d 全因死亡率几乎没有影响
5 年心血管死亡率[4 394(4 项 RCT)]	HR=1.07 (0.83~1.37)	5.9%	6.3% (4.9%~8.0%)	增加 0.4%(减少 1.0%至增加 2.1%)	⊕⊕○○ 低	PCI 可能不会影响 5 年心血管死亡率
30 d 卒中发生率[3 957(4 项 RCT)]	RR=0.38 (0.16~0.89)	1.0%	0.4% (0.2%~0.9%)	减少 0.6%(减少 0.9%至减少 0.1%)	⊕⊕⊕⊕ 高	PCI 未增加 30 d 卒中风险(高确定性); PCI 可能略微降低 30 d 卒中风险(低确定性)
30 d 生活质量改善(采用西雅图心绞痛问卷-生活质量量表评分,0~100,分值越高表示生活质量越好)[1 538(1 项 RCT)]	RR=1.07 (1.01~1.14)	71.6%	76.6% (72.3%~81.6%)	增加 5.0%(增加 0.7%至增加 10.0%)	⊕⊕○○ 低	PCI 可能轻微改善 30 d 生活质量
30 d 主要或危及生命的出血事件[3 271(3 项 RCT)]	RR=0.24 (0.18~0.31)	12.0%	2.9% (2.2%~3.7%)	减少 9.1%(减少 9.9%至减少 8.3%)	⊕⊕⊕○ 中等	PCI 可能大幅减少 30 d 主要或危及生命的 出血事件
5 年内再次血运重建(新 PCI 或 CABG [4 037 (4 项 RCT)])	RR=1.73 (1.49~2.02)	11.0%	19.1% (16.5%~22.3%)	增加 8.1%(增加 5.4%至增加 11.3%)	⊕⊕⊕○ 中等	PCI 可能导致 5 年内再次血运重建(PCI 或 CABG)发生中度至大幅增加
5 年内自发性心肌梗死[4 037(4 项 RCT)]	RR=2.02 (1.52~2.70)	3.2%	6.5% (4.9%~8.7%)	增加 3.3%(增加 1.7%至增加 5.5%)	⊕⊕⊕○ 中等	PCI 可能增加 5 年内自发性心肌梗死的风险

注:HR 为风险比;RR 为相对危险度;CI 为置信区间;“⊕”表示对应证据质量评价条目无需降级;“○”表示对应证据质量评价条目需降 1 级。

对于指南制定者或卫生技术评估者,则需基于系统评价结果,起草结构化的推荐意见决策表。起草过程需综合考虑干预的获益与风险、证据质量、患者价值观、成本及可行性等因素。最终,指南制定者或卫生技术评估者综合权衡上述因素,形成包含推荐方向(推荐或不推荐)、强度(强或弱)的推荐意见。

2.2 Core GRADE 2:判定证据质量评价目标及评价证据不精确性的方法

在判定证据质量评价的目标时,可根据所关注的问题选择无效线或最小重要差值(minimal important difference, MID)。当所关注的问题是干预措施与标准治疗措施相比,是否有效时,可选择无效线;当所关注的问题是干预措施与标准治疗措施相比,效果大小是否具有实际意义时,则需选择 MID。例如,在评估 SGLT2 抑制剂降低 HF_pEF 患者心血管死亡风险的治疗效果时,若关注 SGLT2 抑制剂能否降低心血管死亡风险,则选择无效线;若关注 SGLT2 抑制剂降低的心血管死亡风险是否有意义,则选择 MID。

为降低证据质量评价中阈值设定的难度,同时兼顾证据质量评价结果的适用性,Core GRADE 体系建议在证据质量评价中,仅考虑使用无效线或 MID 作为阈值,无需再考虑此前 GRADE 体系中所建议的中等效应量阈值或大效应量阈值。

在证据质量评价时需注意,通常应使用绝对效应值而非相

对效应值。其原因在于,当患者判定是否接受某一干预措施时,该干预措施获益与风险的绝对效应量比相对效应量更有价值。例如,《Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein》针对低密度脂蛋白胆固醇(LDL-C)水平正常但超敏 C 反应蛋白(hs-CRP)升高的人群,探究瑞舒伐他汀能否改善心血管事件的发生风险^[14]。该研究结果显示,对于心血管事件发生率这一结局,瑞舒伐他汀组和安慰剂组的发生率分别为 0.77% 和 1.36% (RR = 0.56, 95%CI = 0.46~0.69, P < 0.000 01);瑞舒伐他汀的治疗使主要心血管事件的相对风险降低 44%,而绝对风险仅降低 0.59%,即每 1 000 例患者中仅有约 6 例患者改善。基于绝对风险结果,患者可能不会接受该干预措施。

当选定阈值后(无效线或 MID),效应量点估计值与阈值的相对位置关系,决定了证据质量评价的目标。例如,选择 MID 为阈值时,效应量的点估计值 ≥ MID,则认为效应达到或超过最小重要差异,干预措施的效果有重要意义(A 点),效应量的点估计值 < MID,则认为干预措施的效果不重要(B 点);选择无效线为阈值时,效应量的点估计值 ≥ 无效线,则认为效应达到或超过无效阈值,干预措施有效(B 点),效应量的点估计值 < 无效线,则认为干预措施无效(C 点),见图 2。

当确定证据质量评价目标后,便可判断证据的不精确性。

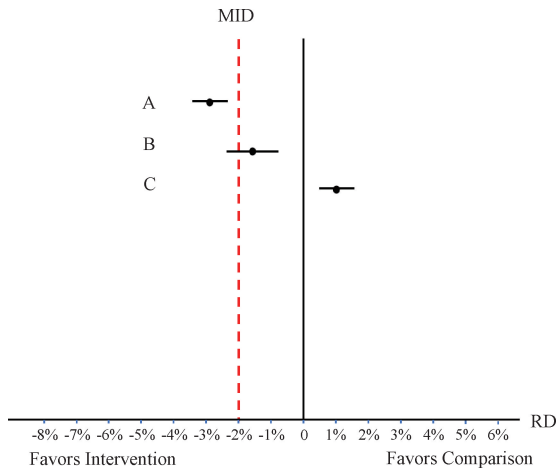


图2 点估计值与阈值示意图

根据 Core GRADE 体系的推荐,证据不精确性的判定方法即根据 *CI* 与阈值的位置关系,决定是否需要因不精确性降级。若 *CI* 未跨越阈值(无效线或 MID),则不需因不精确性降低证据质量等级;若 *CI* 跨越阈值,则需因不精确性降低证据质量等级。例如,图 2 中的 B 点,选择无效线为阈值时,*CI* 未跨越无效线,不需因不精确性而降级;选择 MID 为阈值时,*CI* 跨越了 MID,需因不精确性而降级。值得注意的是,*CI* 跨越多个阈值时(例如,同时包括重要的有益效应和重要的有害效应),可考虑因不精确性降低 2 级。*CI* 未跨越阈值且效应量较大时(例如,相对风险降低>30%甚至 40%),需要检验 Meta 分析的样本量是否大于最优信息样本量(OIS);若 \geq OIS 则不降级,若 $<$ OIS 则需因不精确性而降级。

2.3 Core GRADE 3:评价证据不一致性的方法

评价证据不一致性时,对于二分类结局,Core GRADE 体系建议关注相对效应指标(如 *RR*);对于连续性结局,则关注绝对效应指标[如均数差(*MD*)]。

对于证据不一致性的评价,首要方法是直接观察森林图中各原始研究效应量估计值的相对关系。评估要素包括点估计值的差异程度、*CI* 的重叠情况,以及点估计值与阈值(无效线或 MID)的关系。当不同原始研究间点估计值差异明显、*CI* 重叠小、点估计值分别位于阈值两侧时,应考虑因不一致性降级。第 2 种评估方法是 I^2 统计量, $I^2 < 30\%$ 时,通常提示存在严重不一致的可能性较低; $I^2 \geq 30\%$ 时,可能需降级。例如,1 篇评估肺康复对慢性阻塞性肺疾病患者疗效的 Cochrane 综述评估了肺康复与常规护理相比对“最大运动能力”的影响,在森林图中,各原始研究的点估计值在“-3.50”至“55”均有分布,差异程度较大,点估计值在无效线两侧均有分布,且 *CI* 重叠程度有限,此时需考虑因不一致性而降级证据质量等级;该对比中, $I^2 = 74\%$,明显大于 30%,提示可能存在严重不一致性,与直接观察森林图所得结果一致,见图 3^[15]。但需要注意, I^2 值不能作为评估证据不一致性的唯一方法,否则可能存在误导性。

当发现不一致性时,若能识别异质性来源,则需分开呈现亚组分析结果并独立评估其不一致性。系统评价者应基于前期研究证据(例如,临床试验、Meta 分析或队列研究)及生物学机制,预先提出有限数量且方向明确的亚组分析假设。若研究

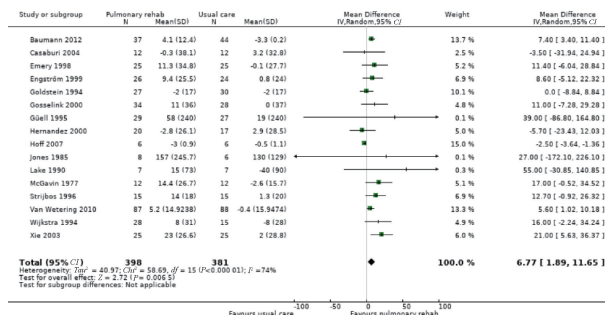


图3 评价证据不一致性森林图

结果间的异质性无法通过亚组分析解决,应考虑因不一致性降级。

Core GRADE 推荐,运用 ICEMAN 工具(instrument to assess the credibility of effect modification analyses)评估亚组分析结果的可靠性^[16]。当亚组分析结果可靠时,Core GRADE 建议分别呈现各亚组的效应量并分别评价其证据质量;当亚组分析结果不可靠时,则呈现全部证据的整合结果,并在必要时因不一致性降低证据质量。关于亚组分析的可靠性评价方法,以及亚组分析可靠性对结果呈现与证据不一致性评级的影响,也是 Core GRADE 方法学体系较 GRADE 体系更为明确和完善之处。

2.4 Core GRADE 4:评估偏倚风险、发表偏倚的方法及针对 NRSI 试验/研究升高证据质量的情况

与 GRADE 体系一致,在 Core GRADE 体系中 RCT 起始于高质量,而 NRSI 起始于低质量。对于 RCT,其证据质量可能因研究设计或执行中的缺陷,导致偏倚风险,从而证据质量降级。Core GRADE 推荐使用 Risk of Bias 2 (RoB 2)^[17] 或 ROBUST-RCT^[18] 等工具评估 RCT 的偏倚风险;对于 NRSI,推荐使用纽卡斯尔-渥太华量表(NOS)^[19] 等工具。

Core GRADE 体系制定了因偏倚风险降低证据质量等级的具体规则。若证据体主要由高偏倚风险的原始研究构成,且偏倚很可能是在 Meta 分析中观察到效应(或未观察到效应)的原因,则降低证据质量。若证据体主要由低偏倚风险的原始研究构成,且其效应估计显著区别于高偏倚风险研究,则在 Meta 分析中仅纳入低偏倚风险的研究(即排除高偏倚风险的研究),证据体不因偏倚风险而降级。若低偏倚风险研究与高偏倚风险研究的结果间无显著差异,则 Meta 分析中可纳入所有研究、合并结果且不因偏倚风险降级。

发表偏倚是指根据研究结果进行选择发表的现象,表现为阳性结果的研究更易被发表,而阴性结果的研究常常未被发表。发表偏倚可通过漏斗图评估,也可通过统计检验(如 Egger 回归)。此外,若研究证据主要来源于小样本且受行业资助的研究,应警惕发表偏倚。必要时,需因发表偏倚风险下调证据质量等级。

在某些情况下,可升高 NRSI 证据质量的等级。当 NRSI 显示较大效应量时,即合并效应量 $RR > 2.0$ 或 < 0.5 时可考虑升高 1 级, $RR > 5.0$ 或 < 0.2 时可考虑升高 2 级。此外,若存在可信的剂量-反应梯度,即干预剂量或强度增加(或减少),效应量发生相应的变化,可增强对其因果关系的信心,从而升高证据质量等级。

2.5 Core GRADE 5:评价证据间接性的方法

当目标临床问题的 PICO 与当前最佳研究证据的 PICO 存在较大差异,且差异可能影响决策时,即产生证据间接性。Core GRADE 建议评估间接性的核心在于判断 PICO 差异的程度,以及差异对证据适用性和效应量估计值可靠性的影响。

PICO 的差异可包括人群特征差异(如年龄、疾病严重程度不同)、干预措施或对照措施差异(如干预强度、持续时间或对照措施不同)、结局指标差异(如随访时长不同)。其中,最值得关注的是使用替代结局指标而非患者认为重要的结局指标。

是否因间接性而降低证据质量等级,关键在于判断 PICO 差异是否可实质性地改变干预措施效应量的大小或方向,从而降低证据的可靠性与适用性。当仅在 PICO 中的一个维度存在差异且差异的影响相对有限时,证据质量降 1 级;若在多个维度存在多个差异或差异影响显著,则可考虑证据质量降 2 级。反之,若存在差异,但几乎不影响效应量的大小或方向,则无需降级。

例如,在探究达格列净对我国 HF_rEF 患者心血管死亡风险的影响时,检索到文献《Dapagliflozin in patients with heart failure and reduced ejection fraction》^[20]。该研究纳入全球 20 个国家的 HF_rEF 患者,虽与目标问题中的我国患者存在地域差异,但此差异几乎不影响效应量方向,从 P(人群)角度无需因间接性降低证据质量等级。该研究主要报告心血管死亡/心力衰竭恶化住院复合终点,复合终点效应可能主要受因心力衰竭恶化而住院影响,与目标问题关注的单纯心血管死亡的效应存在差异,从 O(结局)角度需因间接性降低证据质量 1 级。

值得注意的是,对于直接研究证据,即使在人群、干预措施或次要结局指标方面,研究证据与临床问题一致,但若干依从性过低或研究仅报告替代结局而非患者重要结局,或对照组设置存在明显不合理,可能引入偏倚,此时需考虑因证据间接性降级。例如,假若前述研究仅报告了心力衰竭恶化住院率结局,该结局并非对患者重要的终点指标,需考虑因证据间接性降级。

当直接证据本身质量较低或缺乏直接证据时,可考虑使用间接证据。使用间接证据时,通常因证据的间接性降低证据质量 1 级或 2 级。

2.6 Core GRADE 6:证据呈现的方法——SoF 表

SoF 表以透明、结构化的形式呈现系统评价中的重要研究结果,即对患者重要的结局指标(包括获益与风险相关指标)。SoF 表中呈现的内容包括干预措施对每项重要结局的效应量(包括相对效应量与绝对效应量)、证据质量等级及升级或降级的原因。当亚组分析较为可靠时,SoF 表应针对不同亚组单独呈现结果。

对于二分类结局指标,绝对效应量通常根据相对效应估计值与目标人群的基线风险计算得出。基线风险通常来源于大型观察性研究。若无相应研究,则可使用随机试验中的对照组实际发生率作为基线风险。对于连续性结局指标,若原始研究采用了不同的测量工具(或量表),Core GRADE 建议将原始研究的结果数据转换为单一、首选工具的自然单位后再进行数据合并,合并效应量估计值也以该自然单位呈现。此外,针对结果的通俗语言表述也是 SoF 表中不可或缺的组成部分,便于患者、临床医师及相关人员理解研究结果。

2.7 Core GRADE 7:“从证据到推荐意见”的方法

Core GRADE 方法学体系指出,高或中等质量的证据通常支持强推荐(或强不推荐),而低或极低质量的证据则倾向于产生弱推荐(或弱不推荐)。

Core GRADE 建议,在制定推荐意见时,主要考虑获益、风险、证据质量、患者价值观 4 个主要因素,以及从人群视角出发的成本、可行性和可接受性等次要因素。所有推荐意见均需反映患者关注的事项,并充分考虑其价值观。通过确定最小重要获益,确保推荐意见聚焦于对患者真正重要的临床结局。最后,采用明确的表述、结构化的呈现、优化的展示形式及决策辅助工具,有效促进临床医师与患者间的共同决策。

3 结语

Core GRADE 方法学体系梳理、整合并优化了既往的 GRADE 工具,为证据质量评价与推荐意见形成提供了更高效实用的指导。然而,该体系仍存在局限性,即证据质量评价以及从证据形成推荐意见的过程几乎不可避免地包含主观判断^[4]。而这种主观性是当前多数证据质量评价与临床决策框架的共性。过度追求绝对客观、僵化的规则,反而可能得出偏颇、脱离实际的结论。

Core GRADE 体系以其简洁性和实用性,在推动循证实践规范化和普及化方面有着巨大潜力。随着在系统评价、临床实践指南制定及卫生技术评估等领域的深入应用,该体系将助力使用者更高效地掌握证据质量评价的核心要素,为临床决策和健康政策的制定提供更坚实的循证基础。

参考文献

- [1] STRAUSS S E, GLASZIOU P, RICHARDSON W S, et al. Evidence-based medicine: how to practice and teach EBM[M]. 5th ed. Edinburgh: Elsevier, 2019: 18-21.
- [2] ATKINS D, BEST D, BRISS P A, et al. Grading quality of evidence and strength of recommendations[J]. BMJ, 2004, 328: 1490.
- [3] 叶志康, 翟所迪. 医院药学指南制定中的证据与推荐意见 GRADE 系统评级方法及应用[J]. 中国医院用药评价与分析, 2024, 24(7): 769-773.
- [4] GUYATT G, HULTCRANTZ M, AGORITSAS T, et al. Why core GRADE is needed: introduction to a new series in *the BMJ*[J]. BMJ, 2025, 389: e081902.
- [5] GUYATT G, AGORITSAS T, BRIGNARDELLO-PETERSEN R, et al. Core GRADE 1: overview of the core GRADE approach[J]. BMJ, 2025, 389: e081903.
- [6] GUYATT G, ZENG L N, BRIGNARDELLO-PETERSEN R, et al. Core GRADE 2: choosing the target of certainty rating and assessing imprecision[J]. BMJ, 2025, 389: e081904.
- [7] GUYATT G, SCHANDELMAIER S, BRIGNARDELLO-PETERSEN R, et al. Core GRADE 3: rating certainty of evidence-assessing inconsistency[J]. BMJ, 2025, 389: e081905.
- [8] GUYATT G, WANG Y, EACHEMPATI P, et al. Core GRADE 4: rating certainty of evidence-risk of bias, publication bias, and reasons for rating up certainty[J]. BMJ, 2025, 389: e083864.
- [9] GUYATT G, IORIO A, DE BEER H, et al. Core GRADE 5: rating certainty of evidence-assessing indirectness[J]. BMJ, 2025, 389: e083865.

(下转第 1290 页)