



协和医学杂志

Medical Journal of Peking Union Medical College Hospital

ISSN 1674-9081, CN 11-5882/R



《协和医学杂志》网络首发论文

题目： 正确理解和应用低质量证据形成指南推荐意见
作者： 史乾灵，刘辉，王子君，罗旭飞，王冰怡，杨楠，孟文勃，陈耀龙
收稿日期： 2024-04-29
网络首发日期： 2024-05-29
引用格式： 史乾灵，刘辉，王子君，罗旭飞，王冰怡，杨楠，孟文勃，陈耀龙. 正确理解和应用低质量证据形成指南推荐意见[J/OL]. 协和医学杂志.
<https://link.cnki.net/urlid/11.5882.R.20240528.1706.010>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

• 临床实践指南 •

正确理解和应用低质量证据形成指南推荐意见

史乾灵^{1,2,3,4}, 刘辉^{2,3,4,5}, 王子君^{2,3,4,5}, 罗旭飞^{2,3,4,5}, 王冰怡^{2,3,4,5},

杨楠^{2,3,4,5}, 孟文勃⁶, 陈耀龙^{2,3,4,5,7,8}

兰州大学¹第一临床医学院, ⁴健康数据科学研究院, ⁵基础医学院循证医学中心,
兰州 730000

²中国医学科学院循证评价与指南研究创新单元 兰州大学基础医学院, 兰州 730000

³甘肃省循证医学重点实验室, 兰州 730000

⁶兰州大学第一医院普外科, 兰州 730000

⁷中华医学会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: 陈耀龙 chevidence@lzu.edu.cn

孟文勃 mengwb@lzu.edu.cn

【摘要】 推荐意见是临床实践指南的核心。基于低质量证据作出强推荐意见是当前国内外指南普遍存在的现象。医务人员认为缺乏高质量的证据就无法制订高质量的指南或仅能够制订专家共识是典型的误解。本文以 GRADE 分级系统为切入点, 阐述了低质量证据的基本内涵和临床作用, 介绍了基于低质量证据形成指南推荐意见的方法, 以期为我国指南制订者和使用者提供参考。

【关键词】 临床实践指南; 推荐意见; 证据质量; GRADE

【中图分类号】 R181.2 **【文献标志码】** A

【文章编号】 1674-9081 (2024) 03-000-00

DOI: 10.12290/xhyxzz.2024-0278

How to Correctly Understand and Use the Low-quality Evidence to Formulate Recommendations in Guidelines

SHI Qianling^{1,2,3,4}, LIU Hui^{2,3,4,5}, WANG Zijun^{2,3,4,5}, LUO Xufei^{2,3,4,5},

WANG Bingyi^{2,3,4,5}, YANG Nan^{2,3,4,5}, MENG Wenbo⁶, CHEN Yaolong^{2,3,4,5,7,8}

¹The First School of Clinical Medicine, ⁴Institute of Health Data Science, ⁵Evidence-based Medicine Center, School of Basic Medical Sciences, Lanzhou University, Lanzhou 730000, China

²Research Unit of Evidence-Based Evaluation and Guidelines, Chinese Academy of Medical Sciences, School of Basic Medical Sciences, Lanzhou University, Lanzhou 730000, China

³Key Laboratory of Evidence-based Medicine of Gansu Province, Lanzhou 730000, China

⁶Department of General Surgery, The First Hospital of Lanzhou University, Lanzhou 730000, China

史乾灵、刘辉对本文同等贡献

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)

引用本文: 史乾灵, 刘辉, 王子君, 等. 正确理解和应用低质量证据形成指南推荐意见[J]. 协和医学杂志, 2024, 15(3). doi: 10.12290/xhyxzz.2024-0278.

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: CHEN Yaolong, chevidence@lzu.edu.cn

MENG Wenbo, mengwb@lzu.edu.cn

【Abstract】 The essence of clinical practice guidelines lies in their recommendations. It is very common to find strong recommendations supported by low-quality evidence in current published guidelines. There is also a typical misunderstanding among medical professionals that without high-quality evidence, there is no way to develop high-quality guidelines or only expert consensus can be developed. Based on the GRADE approach, this paper explains the concept and clinical significance of low-quality evidence, and introduce the methods for formulating recommendations based on low-quality evidence in guidelines, aiming to provide reference for guideline developers and users in China.

【Key words】 clinical practice guidelines; recommendations; quality of evidence; GRADE

Funding: CAMS Innovation Fund for Medical Sciences-Research Unit of Evidence-based Evaluation and Guidelines (2021RU017)

虽然临床实践指南（下文简称“指南”）的质量在过去三十年间稳步提升^[1-3]，但制订者对其原理和方法的理解仍存在误区。其中之一是关于低质量证据、强推荐级别以及高质量指南之间的关系。本文以推荐分级的评估、制订与评价（grading of recommendations assessment, development, and evaluation, GRADE）分级系统为切入点^[4]，阐述低质量证据的基本内涵和临床作用，介绍基于低质量证据形成指南推荐意见的方法，以期为我国指南制订者和使用者提供参考。

1 关于低质量证据的正确理解

1.1 低质量证据是指证据体，其中也可能包含若干个高质量的原始研究

指南的新定义要求推荐意见必须基于系统评价的证据^[1]。该定义表明，对于某一具体的临床问题，指南应尽可能全面检索和收集符合纳入标准的所有原始研究（primary study），并进行严格评价、综合和分析后基于证据综合（evidence synthesize）得出结论。因为不同研究者针对同一临床问题所开展的研究，可能因设计和实施存在差异，而产生不同，甚至矛盾的结果。若仅基于单个研究进行临床实践，可能作出有偏倚甚至弊大于利的临床决策^[5]。因此，本文所说的证据，特指系统检索后能够回答某一临床问题的所有符合标准的研究，它们被视为一个整体，称为“证据体或证据群（body of evidence）”，通常采用 GRADE 方法评估其质量。

在 GRADE 分级系统中，证据质量（quality of evidence）被定义为“对观察值的真实性有多大把握”^[4]。对于某一具体的临床问题，当研究被系统检索和汇总后，可基于 5 个降级因素和 3 个升级因素将证据体质量分为高、中、低、极低 4 个等级（表 1）^[6-7]。为便于理解和应用，本文将 GRADE 分级系统中的 C 级（低）和 D 级（极低）证据，均归类为“低质量证据”。在证据分级的过程中，指南制订者易将证据质量与原始研究本身的质量混淆，但后者只是影响证据质量的因素之一，GRADE 手册将其称为研究的局限性或偏倚风险（图 1）^[8]。在进行证据分级时还需考虑研究结果的不精确性、不一致性和间接性，以及发表偏倚等^[7-8]。以“瑞德西韦能否有效治疗新型冠状病毒感染”为例，一篇发表在《柳叶刀呼吸病学》的系统评价最终纳入 9 篇随机对照试验研究进行分析^[9]，其中部分单个研究的质量很高或偏倚风险很低^[10-11]。但汇总分析结果显示，因研究总体估计效应的精确度显著不足，

对于接受通气治疗的住院患者，瑞德西韦能否降低病死率仍存在较大的不确定性（对应 GRADE 分级系统中的 C 级证据）。

表 1 GRADE 分级系统中证据质量与推荐强度的含义^[4,6]

Tab.1 Quality of evidence and strength of recommendations in GRADE system

证据质量	具体描述	表达符号
高	非常有把握观察值接近真实值	A
中	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	B
低	对观察值的把握有限：观察值可能与真实值有很大差别	C
极低	对观察值几乎没有把握：观察值与真实值极可能有极大差别	D

推荐强度	具体描述	表达符号
强*	明确显示实施某项推荐意见利大于弊或弊大于利	1
弱**	实施某项推荐意见的利弊不确定，或无论质量高低的证据均显示利弊相当	2

注：*包括强推荐和强不推荐；**包括弱推荐和弱不推荐；GRADE（Grading of Recommendations Assessment, Development, and Evaluation）：推荐分级的评估、制订与评价

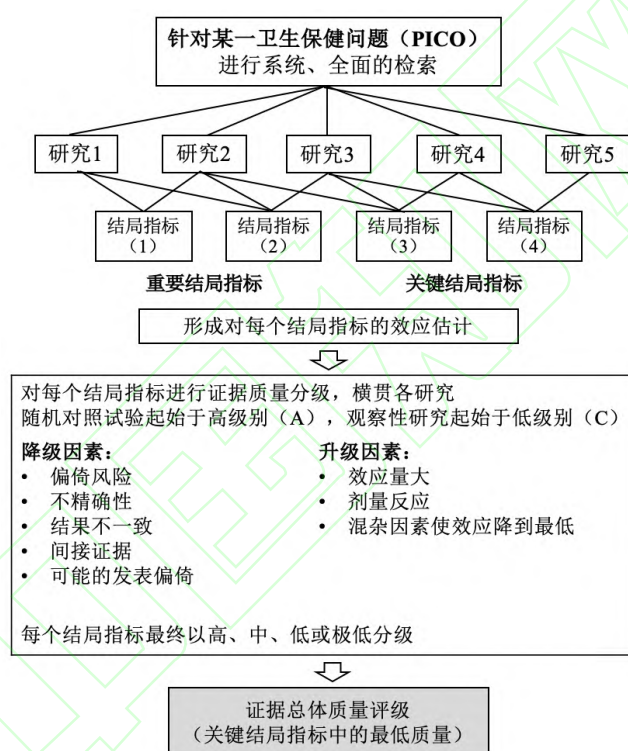


图 1 应用 GRADE 分级系统评估某一临床问题中不同结局指标的证据体质量^[7]

Fig. 1 Evaluating the quality of evidence for different outcomes in PICO question using GRADE

注：PICO（Participants/Patients, Intervention, Control/Comparison, Outcome）：研究对象，干预措施，对照或比较措施，结局指标；GRADE：同表 1

1.2 低质量证据也是当前可得的最佳证据

仍以上文“瑞德西韦能否有效治疗新型冠状病毒感染”为例，尽管瑞德西韦在降低需接受通气治疗的新型冠状病毒感染患者病死率方面的证据质量为“低”，但相对于单个研究或专家经验而言，其仍属于当前可得的“最佳”证据，可为决策者提供重要的判断依据。类似情况还存在于某些经常被临床研究排除的特殊群体（如儿童、孕产妇或老年患者）、罕见病患者，以及其他无法开展随机对照试验的情况^[12-13]。针对此类人群开展的临床研究，进行系统评价后所得到的结果，尽管使用 GRADE 方法有可能判断为低或极低质量，依旧是当前可获得的最佳证据，能够为指南和临床实践提供重要信息^[14]。另一种情况可见于突

发公共卫生事件，如新冠疫情初期，专家需在短时间内作出决策。彼时直接证据很少甚至没有，但仍可借鉴间接证据开展决策。譬如《国际儿童新型冠状病毒疾病管理快速建议指南》^[15]，结合严重急性呼吸综合征（severe acute respiratory syndrome coronavirus, SARS）、中东呼吸综合征（Middle East respiratory syndrome, MERS）和流感的间接证据，制作了快速系统评价，就儿童新冠感染患者的临床表现、筛查、治疗、母乳喂养、健康教育等方面给出了10条具体推荐意见，其中8条基于低或极低质量证据。同样，2018年 *Lancet* 发表的《埃博拉病毒病患者支持治疗的循证指南》^[16]，在系统检索文献的基础上，结合与埃博拉病毒病相似的其他疾病证据（如休克、霍乱、败血症和其他严重腹泻疾病），指南制订小组共形成了8条推荐意见，其中3条基于低质量证据，4条基于中等质量证据，仅1条基于高质量证据。

1.3 低质量证据在证据谱中占主要比例

相关研究基于 GRADE 分级系统分析了已发表指南和系统评价中的证据质量，笔者检索了 MEDLINE 数据库，对其进行初步汇总。检索策略为("systematic review*" [Title] OR "meta-analys*" [Title] OR "meta analys*" [Title] OR "guideline*" [Title] OR "recommendation*" [Title]) AND ("certainty" [Title] OR "confidence" [Title] OR ("evidence" [Title] AND ("quality" [Title] OR "level" [Title] OR "levels*" [Title] OR "underlying" [Title] OR "strength" [Title] OR "behind" [Title]))). 检索时间为建库至 2024 年 4 月 30 日，共获取 22 篇相关研究^[17-38]。结果显示，低（C 级）和极低（D 级）质量证据在证据谱中占主要比例（图 2）。表 2 呈现了各领域临床实践指南和系统评价中不同证据质量的详细情况。

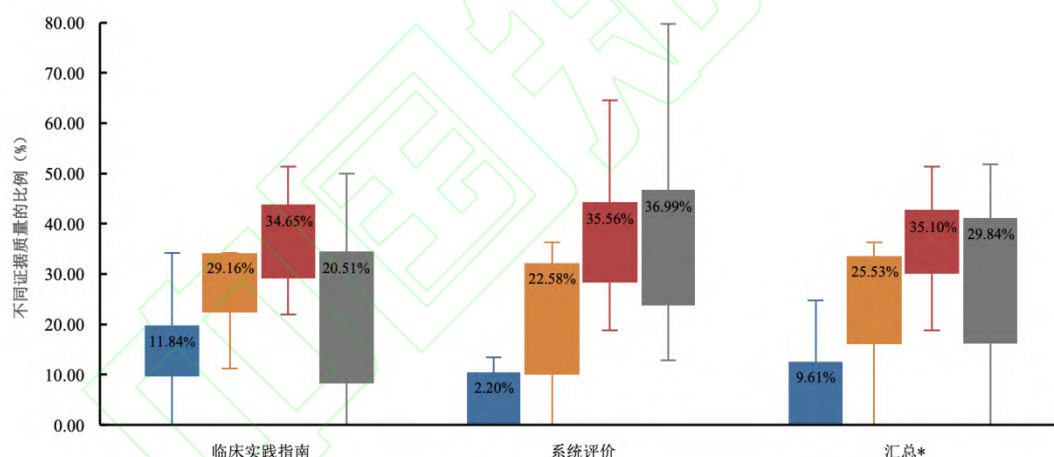


图 2 临床实践指南和系统评价中不同证据质量的比例

Fig. 2 Quality of evidence in clinical practice guidelines and systematic reviews

注：图中标记的具体数据表示中位数；蓝色代表高质量证据（A 级），橙色代表中等质量证据（B 级），红色代表低质量证据（C 级），灰色代表极低质量证据（D 级）；*包括临床实践指南和系统评价

表 2 各领域临床实践指南和系统评价中不同证据质量的比例^[17-38]

Tab. 2 Quality of evidence in clinical practice guidelines and systematic reviews across different areas

领域	高质量证据 (%)	中等质量证据 (%)	低质量证据 (%)	极低质量证据 (%)
临床实践指南 (n=9)				
幽门螺旋杆菌感染	34.26	26.73	21.98	17.03
胃肠病	24.70	34.26	32.67	8.37
结直肠癌筛查	14.47	34.21	51.32	0.00
综合 (NCEC 指南)	13.81	23.20	25.97	37.02
综合 (WHO 指南)	11.84	21.93	34.65	31.58
重症医学	10.91	29.16	34.37	25.55
康复	10.26	30.77	38.46	20.51
传染病	9.31	33.69	48.56	8.45

银屑病性关节炎	0.00	11.25	38.75	50.00
系统评价 (n=13)				
干预类 (Cochrane) ^a	13.49	30.76	31.74	24.01
干预类 (Cochrane) ^b	12.06	30.12	34.01	23.80
麻醉、重症监护和急诊医学 (Cochrane)	10.68	33.98	40.78	14.56
公共卫生	9.91	24.32	25.23	40.54
干预类 ^c	4.14	11.62	32.40	51.83
雷公藤多苷治疗糖尿病肾病	4.11	17.81	41.10	36.99
口腔健康	2.20	21.98	47.25	28.57
丹参酮 II A 磺酸钠注射液治疗冠心病	0.00	36.36	21.21	42.42
中药治疗银屑病	0.00	33.33	35.56	31.11
中医护理	0.00	22.58	64.52	12.90
急性肾损伤	0.00	8.64	40.74	50.62
中药治疗偏头痛	0.00	1.45	18.84	79.71
针灸治疗乳腺癌相关淋巴水肿	0.00	0.00	61.54	38.46

注: NCEC (National Clinical Effectiveness Committee): 爱尔兰国家临床有效性委员会; WHO (World Health Organization): 世界卫生组织; ^a纳入 Cochrane 协作网发布的 1394 篇与医疗或健康干预有关的系统评价; ^b纳入 Cochrane 协作网发布的 2428 篇干预类系统评价; ^c纳入 76 篇发表于 *Annals of Internal Medicine*、*BMJ*、*JAMA*、*Pediatrics* 等杂志的干预类系统评价

1.4 低质量证据与高质量指南之间并无直接关系

在指南制订的过程中, 医务人员普遍存在一个误区, 即如果大部分推荐意见的支持证据都是低质量的, 是否会对指南质量产生影响。事实上, 指南质量的高低与其纳入证据的质量并无直接关系^[14]。目前国际上公认的指南方法学质量评价工具 AGREE II (appraisal of guidelines research and evaluation II) 和指南报告质量评价工具 RIGHT (reporting items for practice guidelines in healthcare)^[39-40], 以及 2022 年推出的指南科学性、透明性和适用性评级工具 STAR (scientific, transparent and applicable rankings tools for clinical practice guidelines)^[41], 均没有任何一条评价标准要求指南必须纳入高质量的研究证据, 也就是说, 指南纳入证据质量的高低与指南本身的质量不相关。恰恰在指南中详细说明通过何种方法检索和评估证据质量, 并给出具体的分级结果, 才是重要的评分项。然而这一点却常被指南制订者忽略, 应用 GRADE 分级系统评估证据质量的指南更是少之又少^[42-43]。

2 基于低质量证据形成推荐意见的方法

国内外不同指南制订机构就如何基于证据质量 (包括低质量证据) 进行推荐给出了相应的方法和建议 (表 3)^[1, 44-48]。其中, 基于 GRADE 研发的 EtD (evidence to decision) 框架, 因其科学的设计、透明的流程和清晰的表述, 成为包括诸多国际组织在内的指南制订者形成推荐意见时参考的主要依据^[49]。

表 3 主要指南制订手册关于从证据至推荐的一般要求和对低质量证据的考虑

Tab. 3 Considerations for evidence to recommendation and low-quality evidence in main guideline development manuals

组织	从证据至推荐的一般要求	关于低质量证据的描述
WHO ^[44]	每个关键问题均需进行系统评价, 使用 GRADE 方法评估每个重要结局指标的证据质量, 并确定证据的总体质量。在形成推荐意见时, 证据质量、价值观和偏好、利弊平衡和资源影响是决定其方向和强度的主要因素, 其他因素还包括可接受性、可实施性、公平性等。	强推荐在指南中并不常见; 当利弊不确定时, 可考虑形成有条件的推荐或弱推荐; 特殊情况下可能因证据不足而不作出推荐。若无可用的证据, 应明确说明证据缺乏, 并将病例报告和国家或个人经验汇编作为形成推荐意见的基础, 并明确记录过程及其局限性

NICE ^[45]	识别和选择相关证据时需进行系统检索，使用 GRADE 或 GRADE-CERQual 方法评估各结局指标的证据质量。若使用其他方法，需相关人员讨论和同意，并说明原因。形成推荐意见时需同时考虑证据的总体质量、利弊平衡、成本效益、可实施性、公平性等各种因素。	证据不充分时（如无证据、证据质量太低，或无法得出确切结论），委员会可考虑如下选择：（1）基于共识形成推荐；（2）形成对未来研究的建议，包括在指南制订中识别的证据不确定性或研究空白；（3）不形成推荐，该方式慎用
GIN ^[46]	通过系统评价总结每个重要结局指标的最佳证据。也可不制作系统评价，但需提供解释。使用分级工具（如 GRADE、美国预防服务工作组分级标准等）评估每个重要结局指标的证据质量和证据的总体质量。对影响推荐意见形成的其他因素也要总结，如价值观和偏好、必要性、公平性、可行性及资源利用等。	当证据不充分或质量极低时，在形成推荐意见时应做好准备，如弱推荐的判断依据应透明。若指南制订者认为作出错误决策的风险很大，可不形成推荐意见；在新的研究出现前，应指出目前的推荐是基于当前研究的最佳方案
IOM ^[1]	无论证据是否充足，指南工作组均应遵循严格透明的步骤，汇总当前可得的所有研究证据，对其数量、质量和一致性进行描述，提供证据质量分级结果，明确证据空白，并同时考虑价值观和偏好、利弊平衡、成本等其他因素。	当指南小组确信实施某项措施利大于弊或弊大于利时，基于低或极低质量证据有可能形成强推荐，但这种情况并不常见。利弊平衡通常十分复杂或很难确定，面对低或极低质量，甚至高质量证据，都可能形成弱推荐
SIGN ^[47]	对于每个需要解决的关键问题，均应对证据进行系统检索，使用 GRADE 方法评估每个关键问题的证据体质量。使用 EtD 框架形成推荐意见，并考虑价值观和偏好、资源可及性、公平性、成本等其他有可能影响推荐意见的因素。	以下四种情况应形成弱推荐：（1）证据基础薄弱；（2）对观察值的把握有限；（3）需要平衡利弊；（4）患者接受度不同。当缺乏证据时可考虑形成 GPP。若工作组认为有必要形成推荐，可基于极低质量证据形成弱推荐，但必须有来自工作组之外的证据支持。若不存在此类证据，应通过正式的方法形成基于共识的建议（consensus-based recommendation）
中华医学会 ^[48]	推荐意见的形成需要基于系统评价的证据，可在评估质量和时效性后使用已发表的系统评价，也可重新制作。推荐使用 GRADE 方法对证据体的质量和推荐意见强度进行分级。使用 EtD 框架形成推荐意见，既要基于当前可得的最佳研究证据，也要综合考虑资源利用、价值观与偏好、公平性等多方面因素。	当证据质量低或缺乏直接证据时，需考虑专家意见和专家证据在其中发挥的作用

NICE（National Institute for Health and Care Excellence）：英国国家卫生与临床优化研究所；GIN（Guidelines International Network）：国际指南联盟；IOM（Institute of Medicine）：美国医学科学院；SIGN（Scottish Intercollegiate Guidelines Network）：苏格兰学会联合指南网络；EtD（Evidence to Decision）：证据到推荐；GPP（Good Practice Points）：良好实践声明；WHO：同表 2；GRADE：同表 1

2.1 正确判断证据质量

尽管低质量证据十分常见，使用 GRADE 方法判断证据质量仍需谨慎，特别注意以下两点。（1）GRADE 的证据质量分级不是对单个临床研究或系统评价的分级，而是针对报告了某个结局指标的证据体的质量分级（图 1）^[4, 6]。即使是同一临床问题的不同结局指标，证据质量也可能存在很大差别。（2）GRADE 将结局指标分为三类：关键（critical）、重要而非关键（important but not critical）、重要性有限（limited importance）^[50]。当一项措施可同时影响多个结局指标时，该措施总体证据质量的判断取决于关键结局指标的最低证据

质量^[50-51]。例如，胰腺癌或壶腹周围癌患者应使用幽门保留术，还是标准的 Whipple 胰十二指肠切除术？针对此问题开展的系统评价最终纳入 6 项随机对照试验研究进行分析^[52-53]。结局指标包括围术期病死率（低质量证据，基于 6 项研究共 490 例患者）、5 年病死率（中等质量证据，基于 3 项研究共 229 例患者）、胆汁渗漏（低质量证据，基于 3 项研究共 268 例患者）、胃排空延迟（极低质量证据，基于 5 项研究共 442 例患者）等。若指南制订小组认为胃排空延迟是关键指标，则该问题的证据总体质量为“极低”；若胃排空是重要而非关键指标，围术期病死率是关键指标，则该问题的证据总体质量为“低”。各升降级因素的详细解释和进一步指导可参考 GRADE 工作组发表的系列论文及相关专著^[54-55]。

2.2 准确把握指南中强、弱推荐的含义

推荐强度是指遵守推荐意见对目标人群产生的利弊程度有多大把握。GRADE 将其分为“强（strong）”“弱（weak）”两类^[5, 7]，后者也可表示为“conditional”“discretionary”或“qualified”（表 1）。对于不同的指南使用者，强、弱推荐的具体内涵略有区别。以临床医生为例，强推荐表示应对几乎所有患者推荐该方案，弱推荐表示不同患者有各自适合的选择，应帮助每位患者作出体现其价值观与偏好的决定。因此，弱推荐并非指推荐意见质量差或不应采用，而是应更加谨慎或结合患者的具体情况进行决策。

2.3 GRADE 方法中基于低或极低质量证据形成强推荐的情况

关键结局指标的证据质量越高，越有可能形成强推荐。但 GRADE 工作组提出，在 5 种特殊情况下也存在基于低（C 级）或极低（D 级）质量证据作出强推荐的可能性（表 4）^[20, 56]。研究显示，基于低质量证据的强推荐在指南中并不少见，但仅少数符合 GRADE 工作组的要求。大部分强推荐意见不合理，未提供相关解释^[57-60]，也忽略了患者个体需求的复杂性^[61]。

表 4 GRADE 分级系统中基于低或极低质量证据形成强推荐的 5 种情况^[20, 56]

Tab.4 Paradigmatic situations in which a strong recommendation may be made based on low or very low certainty evidence

序号	具体情况	示例
1	危及生命（或灾难性）的情况下，低质量证据表明某种措施可能降低死亡率（有关危害的证据可能很低或很高）	两性霉素与伊曲康唑用于治疗威胁生命的播散性芽生菌病患者。高质量证据表明两性霉素产生的毒性比伊曲康唑更大，但低质量证据表示两性霉素在此情形下可降低病死率。考虑到危及生命的情况，指南制订小组强推荐使用两性霉素
2	低或极低质量证据表明某种措施不确定是否获益，但中等或高质量证据表明该措施有害或成本很高	行手术治疗的白内障患者可接受球后麻醉或局部麻醉。低质量证据显示，球后麻醉较局部麻醉具有更多益处，如降低术中小韧带撕裂、虹膜脱垂的风险和手术疼痛等。但高质量证据表明，球后麻醉会带来严重危害，包括球结膜水肿、眶周血肿，甚至产生危及生命的并发症。考虑到这一情况，指南制订小组强推荐白内障手术患者采取局麻替代球后麻醉
3	低质量证据表明备选方案的获益大小相似，但高质量证据表明其中一种危害更小或成本更低	黏膜相关淋巴组织淋巴瘤患者存在幽门螺杆菌感染，通常行放射或胃切除术治疗。低质量证据表明，使用抗生素治疗幽门螺杆菌与放射或胃切除术具有相似的完全缓解率。同时高质量证据表明，抗生素可明显减少危害、降低发病率和成本。故指南强推荐此类患者使用抗生素根除幽门螺杆菌，而不是进行放疗或胃切除

4	高质量证据表明备选方案的获益大小相似，但低质量证据表明其中一种危害更大或成本更高	亚临床甲状腺功能减退症患者能否从甲状腺激素治疗中获益尚不确定。中等至高质量证据表明，无论是否服用激素，患者在甲状腺相关症状和生活质量方面的改善是相似的，但低质量证据表明，服用激素并非无害。故指南强推荐不为此类成人患者提供甲状腺激素治疗
5	某项措施不确定是否获益，或高质量证据表明中等获益，但低或极低质量证据表明可能会发生灾难性的危害	对于前列腺癌患者或有前列腺癌风险的男性，高质量证据表明睾酮对改善患者的骨矿物质密度和肌肉力量有中等效益；但低质量证据表明，睾酮同时会造成伤害（如增加癌细胞转移风险）。故指南强推荐前列腺癌患者不应服用睾酮治疗

*GRADE: 同表 1

2.4 证据质量是影响推荐意见的重要因素，但不是唯一因素

除证据质量外，成本费用、价值观和偏好、可及性、公平性等其他因素也可对推荐意见产生影响，指南制订者需进行综合考虑^[52, 56]。以价值观和偏好为例，其通常是指患者或公众对健康的看法、认知、期望和目标。针对同一问题，当多个有效方案并存时，不同方案的风险如何，是否更加耗时、成本更高、投入精力更多，均可对最终决策产生影响，尤其是证据质量较低时。EtD 方法能够系统总结不同因素对推荐意见的影响，促进整个过程更加科学透明。对于指南使用者，EtD 框架有助于其理解推荐意见的产生过程，促进推荐意见的实施^[49, 62]。鉴于篇幅，本文将不对其具体内容进行阐释，感兴趣的读者可参考相关研究论文。

2.5 低质量证据或弱推荐对未来临床研究开展具有重要启示

在开展新的原始研究之前，理论上必须先作针对该问题的系统评价^[63]。对于指南的使用者，推荐意见不仅是帮助其开展循证决策、规范诊疗行为的重要参考，更是发现科学问题的突破口，尤其是指南中基于低质量证据形成的弱推荐。对于指南的制订者，在形成推荐意见后，可基于对已有证据的掌握，提出研究空白或未来研究的建议，并在指南全文中汇总呈现^[40, 64]，促进未来高质量证据的生产。以 2023 年世界卫生组织发布的《6~23 月龄婴幼儿辅食喂养指南》为例^[65]，该指南就“母乳喂养、乳制品选择、辅食的引入及选择、营养补充品、强化食品”等多个方面给出了 15 条推荐意见，其中 3 条基于极低质量证据，9 条基于低质量证据，并提出了 16 条具体的临床研究选题。

3 小结

低质量证据在临床实践中普遍存在，但其并非无用的证据，也不决定指南的质量，甚至可以支持高级别推荐。基于低质量证据形成的推荐意见，仍然可以为临床实践提供重要指导，同时也可临床研究的开展提供优先和重要选题。正确理解和应用低质量证据，对于指南制订者、使用者和临床研究的开展者均具有重要意义和价值。

作者贡献：史乾灵、陈耀龙负责论文构思；史乾灵、刘辉负责论文初稿撰写；王子君、罗旭飞、王冰怡、杨楠负责论文修订；陈耀龙和孟文勃负责研究指导和论文审校。

利益冲突：所有作者均声明不存在利益冲突

参考文献

- [1] Institute of Medicine. Clinical practice guidelines we can trust[M]. Washington, D.C.: The National Academies Press, 2011.

- [2] Mc Allister M, Florez I D, Stoker S, et al. Advancing guideline quality through country-wide and regional quality assessment of CPGs using AGREE: a scoping review[J]. *BMC Med Res Methodol*, 2023, 23(1): 283.
- [3] Zhou Q, Wang Z J, Shi Q L, et al. Clinical epidemiology in China series. Paper 4: the reporting and methodological quality of Chinese clinical practice guidelines published between 2014 and 2018: a systematic review[J]. *J Clin Epidemiol*, 2021, 140: 189-199.
- [4] 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.
- [5] 周奇, 董冲亚, 王业明, 等. 循证医学理念在指导临床研究与实践中的作用: 基于对抗病毒药物治疗新型冠状病毒肺炎的思考[J]. *中国循证医学杂志*, 2022, 22(4): 373-379.
Zhou Q, Dong C Y, Wang Y M, et al. The role of evidence-based medicine in guiding clinical research and practice: expert opinion based on studies on antiviral drugs for the treatment of COVID-19[J]. *Chin J Evid-Based Med*, 2022, 22(4): 373-379.
- [6] Guyatt G, Oxman A D, Akl E A, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables[J]. *J Clin Epidemiol*, 2011, 64(4): 383-394.
- [7] 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.
- [8] Guyatt G H, Oxman A D, Vist G, et al. GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias)[J]. *J Clin Epidemiol*, 2011, 64(4): 407-415.
- [9] Amstutz A, Speich B, Mentré F, et al. Effects of remdesivir in patients hospitalised with COVID-19: a systematic review and individual patient data meta-analysis of randomised controlled trials[J]. *Lancet Respir Med*, 2023, 11(5): 453-464.
- [10] Beigel J H, Tomashek K M, Dodd L E, et al. Remdesivir for the treatment of covid-19 - final report[J]. *N Engl J Med*, 2020, 383(19): 1813-1826.
- [11] Wang Y M, Zhang D Y, Du G H, et al. Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial[J]. *Lancet*, 2020, 395(10236): 1569-1578.
- [12] Jacobson R M, Pignolo R J, Lazaridis K N. Clinical trials for special populations: children, older adults, and rare diseases[J]. *Mayo Clin Proc*, 2024, 99(2): 318-335.
- [13] Hanlon P, Hannigan L, Rodriguez-Perez J, et al. Representation of people with comorbidity and multimorbidity in clinical trials of novel drug therapies: an individual-level participant data analysis[J]. *BMC Med*, 2019, 17(1): 201.
- [14] 陈耀龙, 杨克虎. 正确理解、制订和使用临床实践指南[J]. *协和医学杂志*, 2018, 9(4): 367-373.
Chen Y L, Yang K H. How to correctly understand, develop, and apply clinical practice guidelines[J]. *Med J PUMCH*, 2018, 9(4): 367-373.
- [15] Liu E M, Smyth R L, Luo Z X, et al. Rapid advice guidelines for management of children with COVID-19[J]. *Ann Transl Med*, 2020, 8(10): 617.
- [16] Lamontagne F, Fowler R A, Adhikari N K, et al. Evidence-based guidelines for supportive care of patients with Ebola virus disease[J]. *Lancet*, 2018, 391(10121): 700-708.
- [17] Ou J Y, Li J Y, Liu Y, et al. Quality appraisal of clinical guidelines for *Helicobacter pylori* infection and systematic analysis of the level of evidence for recommendations[J]. *PLoS One*, 2024, 19(4): e0301006.
- [18] Murayama A, Kamamoto S, Murata N, et al. Evaluation of financial conflicts of interest and quality of evidence in Japanese gastroenterology clinical practice guidelines[J]. *J Gastroenterol Hepatol*, 2023, 38(4): 565-573.
- [19] Weissman S, Goldowsky A, Aziz M, et al. Colorectal cancer screening guidelines are primarily based on low-moderate-quality evidence[J]. *Dig Dis Sci*, 2021, 66(12): 4208-4219.
- [20] Chong M C, Sharp M K, Smith S M, et al. Strong recommendations from low certainty evidence: a cross-sectional analysis of a suite of national guidelines[J]. *BMC Med Res Methodol*, 2023, 23(1): 68.
- [21] Alexander P E, Bero L, Montori V M, et al. World Health Organization recommendations are often strong based on low confidence in effect estimates[J]. *J Clin Epidemiol*, 2014, 67(6): 629-634.
- [22] Zhang Z H, Hong Y C, Liu N. Scientific evidence underlying the recommendations of critical care clinical practice guidelines: a lack of high level evidence[J]. *Intensive Care Med*, 2018, 44(7): 1189-1191.
- [23] 周奇, 王玲, 杨楠, 等. 基于 GRADE 康复临床实践指南证据质量与推荐强度研究[J]. *中国康复理论与实践*, 2020, 26(2): 156-160.
Zhou Q, Wang L, Yang N, et al. Quality of evidence and strength of recommendations of clinical practice guidelines of rehabilitation based on GRADE[J]. *Chin J Rehabil Theory Pract*, 2020, 26(2): 156-160.
- [24] Miles K E, Rodriguez R, Gross A E, et al. Strength of recommendation and quality of evidence for recommendations in current Infectious Diseases Society of America guidelines[J]. *Open Forum Infect Dis*, 2021, 8(2): ofab033.
- [25] Mamada H, Murayama A, Kamamoto S, et al. Evaluation of financial and nonfinancial conflicts of interest and quality of evidence underlying psoriatic arthritis clinical practice guidelines: analysis of personal payments from pharmaceutical companies and authors' self-citation rate in Japan and the United States[J]. *Arthritis Care Res (Hoboken)*, 2023, 75(6): 1278-1286.
- [26] Fleming P S, Koletsis D, Ioannidis J P A, et al. High quality of the evidence for medical and other health-related interventions was uncommon in Cochrane systematic reviews[J]. *J Clin Epidemiol*, 2016, 78: 34-42.
- [27] Howick J, Koletsis D, Ioannidis J P A, et al. Most healthcare interventions tested in Cochrane reviews are not effective according to high quality evidence: a systematic review and meta-analysis[J]. *J Clin Epidemiol*, 2022, 148: 160-169.

- [28] Conway A, Conway Z, Soalheira K, et al. High quality of evidence is uncommon in Cochrane systematic reviews in anaesthesia, critical care and emergency medicine[J]. *Eur J Anaesthesiol*, 2017, 34(12): 808-813.
- [29] Xun Y Q, Guo Q Q, Ren M J, et al. Characteristics of the sources, evaluation, and grading of the certainty of evidence in systematic reviews in public health: a methodological study[J]. *Front Public Health*, 2023, 11: 998588.
- [30] Kane R L, Butler M, Ng W. Examining the quality of evidence to support the effectiveness of interventions: an analysis of systematic reviews[J]. *BMJ Open*, 2016, 6(5): e011051.
- [31] Shi H S, Deng P, Dong C D, et al. Quality of evidence supporting the role of tripterygium glycosides for the treatment of diabetic kidney disease: an overview of systematic reviews and meta-analyses[J]. *Drug Des Devel Ther*, 2022, 16: 1647-1665.
- [32] Pandis N, Fleming P S, Worthington H, et al. The quality of the evidence according to GRADE is predominantly low or very low in oral health systematic reviews[J]. *PLoS One*, 2015, 10(7): e0131644.
- [33] Peng L F, Fan M X, Li J H, et al. Evidence quality assessment of sodium tanshinone IIA sulfonate injection intervention coronary heart disease angina pectoris: an overview of systematic reviews and meta-analyses[J]. *Medicine (Baltimore)*, 2023, 102(44): e35509.
- [34] Zhang J, Yu Q Y, Peng L, et al. Benefits and safety of Chinese herbal medicine in treating psoriasis: an overview of systematic reviews[J]. *Front Pharmacol*, 2021, 12: 680172.
- [35] Jin Y H, Wang G H, Sun Y R, et al. A critical appraisal of the methodology and quality of evidence of systematic reviews and meta-analyses of traditional Chinese medical nursing interventions: a systematic review of reviews[J]. *BMJ Open*, 2016, 6(11): e011514.
- [36] Kong Y K, Wei X Q, Duan L, et al. Rating the quality of evidence: the GRADE system in systematic reviews/meta-analyses of AKI[J]. *Ren Fail*, 2015, 37(7): 1089-1093.
- [37] Fu G J, Fan X M, Liang X, et al. An overview of systematic reviews of Chinese herbal medicine in the treatment of migraines[J]. *Front Pharmacol*, 2022, 13: 924994.
- [38] Wang L, Du X Y, Hu P, et al. Quality of evidence supporting the role of acupuncture for breast cancer-related lymphoedema: an overview of systematic reviews and meta-analyses[J]. *J Cancer Res Clin Oncol*, 2023, 149(18): 16669-16678.
- [39] Brouwers M C, Kho M E, Browman G P, et al. AGREE II: advancing guideline development, reporting and evaluation in health care[J]. *CMAJ*, 2010, 182(18): E839-E842.
- [40] Chen Y L, Yang K H, Marušić A, et al. A reporting tool for practice guidelines in health care: the RIGHT statement[J]. *Ann Intern Med*, 2017, 166(2):128-132.
- [41] 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.
- [42] 中华医学会杂志社指南与标准研究中心, 中国医学科学院循证评价与指南研究创新单元(2021RU017), 世界卫生组织指南实施与知识转化合作中心, 等. 2022年医学期刊发表中国指南和共识的科学性、透明性和适用性的评级[J]. *中华医学杂志*, 2023, 103(37): 2912-2920.
Guidelines and Standards Research Center Chinese Medical Association Publishing House, Research Unit of Evidence-Based Evaluation and Guidelines (2021RU017), Chinese Academy of Medical Sciences, WHO Collaborating Centre for Guideline Implementation and Knowledge Translation, et al. Evaluation and ranking for scientific, transparent and applicable of Chinese guidelines and consensus published in the medical journals in 2022[J]. *Natl Med J China*, 2023, 103(37): 2912-2920.
- [43] Barker T H, Dias M, Stern C, et al. Guidelines rarely used GRADE and applied methods inconsistently: a methodological study of Australian guidelines[J]. *J Clin Epidemiol*, 2021, 130: 125-134.
- [44] World Health Organization. WHO handbook for guideline development[M]. 2nd ed. Geneva: WHO Press, 2014.
- [45] National Institute for Health and Care Excellence. Developing NICE guidelines: the manual[M/OL]. London: National Institute for Health and Care Excellence, 2024(2024-01-17)[2024-04-02].
<https://www.nice.org.uk/process/pmg20/resources/developing-nice-guidelines-the-manual-pdf-72286708700869>.
- [46] Guidelines International Network, McMaster University. GIN-McMaster guideline development checklist[EB/OL]. [2024-04-02]. <https://macgrade.mcmaster.ca/resources/gin-mcmaster-guideline-development-checklist/>.
- [47] Scottish Intercollegiate Guidelines Network (SIGN). A guideline developer's handbook[M/OL]. [2024-04-02]. <https://www.sign.ac.uk>.
- [48] 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 developing and updating Chinese diagnosis and treatment guidelines(version 2022)[J]. *Natl Med J China*, 2022, 102(10): 697-703.
- [49] 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.
- [50] Guyatt G H, Oxman A D, Kunz R, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes[J]. *J Clin Epidemiol*, 2011, 64(4): 395-400.
- [51] Guyatt G, Oxman A D, Sultan S, et al. GRADE guidelines: 11. Making an overall rating of confidence in effect estimates for a single outcome and for all outcomes[J]. *J Clin Epidemiol*, 2013, 66(2): 151-157.

- [52] Guyatt G H, Oxman A D, Kunz R, et al. What is “quality of evidence” and why is it important to clinicians?[J]. *BMJ*, 2008, 336(7651): 995-998.
- [53] Karanicolas P J, Davies E, Kunz R, et al. The pylorus: take it or leave it? Systematic review and meta-analysis of pylorus-preserving versus standard whipple pancreaticoduodenectomy for pancreatic or periampullary cancer[J]. *Ann Surg Oncol*, 2007, 14(6): 1825-1834.
- [54] Guyatt G H, Oxman A D, Schünemann H J, et al. GRADE guidelines: a new series of articles in the *Journal of Clinical Epidemiology*[J]. *J Clin Epidemiol*, 2011, 64(4): 380-382.
- [55] 杨克虎, 陈耀龙, 孙凤, 等. 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: 3.
Yang K H, Chen Y L, Sun F, et al. The application of GRADE in systematic reviews and practice guidelines[M]. 2nd ed. Beijing: Peking Union Medical College Press, 2021: 3.
- [56] Andrews J C, Schünemann H J, Oxman A D, et al. GRADE guidelines: 15. Going from evidence to recommendation-determinants of a recommendation's direction and strength[J]. *J Clin Epidemiol*, 2013, 66(7): 726-735.
- [57] Bautista-Orduno K G, Dorsey-Trevino E G, Gonzalez-Gonzalez J G, et al. American Thyroid Association guidelines are inconsistent with grading of recommendations assessment, development, and evaluations-a meta-epidemiologic study[J]. *J Clin Epidemiol*, 2020, 123: 180-188.e2.
- [58] 谢方瑜, 孙秀杰, 芦秀燕, 等. 低质量证据强推荐意见在中国护理实践指南中的现状分析[J]. *护理学报*, 2021, 28(20): 39-43.
Xie F Y, Sun X J, Lu X Y, et al. Current status of recommendation of low-quality evidence in Chinese nursing practice guidelines[J]. *J Nurs*, 2021, 28(20): 39-43.
- [59] Alexander P E, Brito J P, Neumann I, et al. World Health Organization strong recommendations based on low-quality evidence (study quality) are frequent and often inconsistent with GRADE guidance[J]. *J Clin Epidemiol*, 2016, 72: 98-106.
- [60] Yao L, Ahmed M M, Guyatt G H, et al. Discordant and inappropriate discordant recommendations in consensus and evidence based guidelines: empirical analysis[J]. *BMJ*, 2021, 375: e066045.
- [61] Yao L, Guyatt G H, Djulbegovic B. Can we trust strong recommendations based on low quality evidence?[J]. *BMJ*, 2021, 375: n2833.
- [62] Meneses-Echavez J F, Bidonde J, Montesinos-Guevara C, et al. Using evidence to decision frameworks led to guidelines of better quality and more credible and transparent recommendations[J]. *J Clin Epidemiol*, 2023, 162: 38-46.
- [63] Lund H, Brunnhuber K, Juhl C, et al. Towards evidence based research[J]. *BMJ*, 2016, 355: i5440.
- [64] 刘辉, 兰慧, 赵思雅, 等. 2019 年期刊公开发表的中国临床实践指南文献调查与评价: 研究空白[J]. *协和医学杂志*, 2022, 13(3): 498-505.
Liu H, Lan H, Zhao S Y, et al. Investigation and evaluation of Chinese clinical practice guidelines published in medical journals in 2019: research gaps[J]. *Med J PUMCH*, 2022, 13(3): 498-505.
- [65] WHO. WHO guideline for complementary feeding of infants and young children 6-23 months of age[M]. Geneva: World Health Organization, 2023.

(收稿: 2024-04-29 录用: 2024-05-21)

(本文编辑: 李玉乐)