

指南的系统评价：是什么，为什么，怎么做

陈耀龙^{1,2,3}，张静怡⁴，张天嵩⁵，孙凤^{6,7}，倪小佳^{8,9}，杨楠^{1,2,3}，王琪^{10,11}，
王健健⁴，马艳芳^{1,2,3}，商洪才¹²，杨克虎^{1,2,3}

¹兰州大学基础医学院循证医学中心，兰州 730000

²世界卫生组织指南实施与知识转化合作中心，兰州 730000

³GRADE 中国中心，兰州 730000

⁴兰州大学公共卫生学院，兰州 730000

⁵复旦大学附属静安区中心医院中医科，上海 200040

⁶北京大学循证医学中心，北京 100191

⁷北京大学公共卫生学院流行病与卫生统计学系，北京 100191

⁸广东省中医院，广州 510120

⁹广州中医药大学第二临床医学院，广州 510120

¹⁰麦克马斯特大学卫生论坛，加拿大汉密尔顿 L8S 4L8

¹¹麦克马斯特大学健康科学学院卫生研究方法、证据和影响系，加拿大哈密尔顿 L8S 4L8

¹²北京中医药大学东直门医院中医内科学 教育部和北京市重点实验室，北京 100700

通信作者：陈耀龙 电话：0931-8912639，E-mail：chenyaolong@vip.163.com

【摘要】近年来，临床实践指南的数量快速增长，越来越多相同或相似的指南被制订和发表，这些指南在方法学质量和推荐意见等方面存在一定差异和冲突，临床工作人员在使用指南的过程中面临诸多困惑。通过对指南进行系统评价，能够全面、客观地呈现某一领域或具体临床问题所有相关指南的现状和差异，更好地指导临床实践，提升医疗服务质量。本文针对指南系统评价的定义、方法及存在的挑战展开讨论，并就如何开展指南系统评价研究提出思考与建议，供指南的制订者、使用者和研究者参考。

【关键词】临床实践指南；系统评价

【中图分类号】 R331 **【文献标志码】** A **【文章编号】** 1674-9081(2020)03-0320-05

DOI： 10.3969/j.issn.1674-9081.20200126

Systematic Review of Clinical Practice Guidelines: What, Why and How

CHEN Yao-long^{1,2,3}，ZHANG Jing-yi⁴，ZHANG Tian-song⁵，SUN Feng^{6,7}，NI Xiao-jia^{8,9}，YANG Nan^{1,2,3}，
WANG Qi^{10,11}，WANG Jian-jian⁴，MA Yan-fang^{1,2,3}，SHANG Hong-cai¹²，YANG Ke-hu^{1,2,3}

¹Evidence-Based Medicine Center, School of Basic Medical Sciences, Lanzhou University, Lanzhou 730000, China

²WHO Collaborating Center for Guideline Implementation and Knowledge Translation, Lanzhou 730000, China

³Chinese GRADE Center, Lanzhou 730000, China

⁴School of Public Health, Lanzhou University, Lanzhou 730000, China

⁵Department of Traditional Chinese Medicine, Jing'an District Central Hospital, Fudan University, Shanghai 200040, China

⁶Center of Evidence-based Medicine and Clinical Research, Peking University, Beijing 100191, China

基金项目：国家重点研发计划（2018YFC1705500）

利益冲突：无

陈耀龙、张静怡对本文同等贡献

⁷Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing 100191, China

⁸Guangdong Provincial Hospital of Chinese Medicine, Guangzhou 510120, China

⁹The Second Clinical School of Guangzhou University of Chinese Medicine, Guangzhou 510120, China

¹⁰McMaster Health Forum, McMaster University, Hamilton L8S 4L8, Canada

¹¹Department of Health Research Methods, Evidence, and Impact (HEI), Faculty of Health Sciences, McMaster University, Hamilton L8S 4L8, Canada

¹²Key Laboratory of Chinese Internal Medicine of Ministry of Education, Beijing Dongzhimen Hospital, Beijing University of Chinese Medicine, Beijing 100700, China

Corresponding author: CHEN Yao-long Tel: 86-931-8912639, E-mail: chenyaolong@vip.163.com

【Abstract】 In recent years, the number of clinical practice guidelines has been growing rapidly. An increasing number of identical or similar guidelines on the same topic have been developed and published for diseases. There are certain differences or conflicts in methodological quality and contents of recommendation. Clinical staffs face a lot of confusion when using the guidelines. A systematic review of guidelines can present more comprehensively and objectively the current status and differences in guidelines on the same topic, and can better guide clinical practice and improve the quality of healthcare service. Thus, we discussed the definition, methods, and challenges in systematic review of guidelines, and put forward some suggestions on how to develop a systematic review of guidelines, so as to provide references for guideline developers, users, and researchers.

【Key words】 clinical practice guideline; systematic review

Med J PUMCH, 2020, 11(3):320-324

临床实践指南（下文简称“指南”）的数量近30年间快速增长，仅我国1年发表的指南数量即超过100部^[1]，每年全球同行评审期刊发表的英文指南数量超过1000部^[2]。医务人员在使用指南时，可能面临如下问题：（1）针对同一主题疾病，通过哪些平台或途径可以全面获取到不同机构和组织制订的指南？例如，2007至2017年我国围绕糖尿病发布的专家共识和指南达100部^[3]，如何确保内分泌科医师全面、系统地检索到这些指南？（2）对内容相同或相似的指南，如何遴选和评价其质量？例如，2007至2017年全球发布了15部痛风指南，这些指南不同维度的质量得分在23%~82%不等^[4]，风湿免疫科医师如何选择指南用于痛风诊疗的决策？（3）当遴选出的指南，其推荐意见不一致时，应如何处理？例如，有研究分析了全球8部具有代表性的高血压管理指南，27%的推荐意见在推荐方向上一致，但在推荐强度上不一致，41%的推荐意见在推荐强度和方向上均不一致^[5]，针对不一致的推荐意见，心内科医师应如何进行选择和判断？此外，指南制订者在计划启动一部新指南时，应对当前已发表的、相同或相似主题的指南进行全面检索和评估，了解目前的发展现状，更客观地阐述当前新指南制订的必要性，避免重复工作^[6]。

如何对已发表的指南进行全面检索和严格评价，并对推荐意见进行系统比较和分析，已成为指南领域

科研工作者关注的新方向。作为本专栏第6篇文章，本文将从“指南系统评价”的定义、方法及存在的挑战3个方面进行分析、讨论，在此基础上就如何开展指南的系统评价研究提出思考与建议。

1 指南系统评价的定义与类型

指南的系统评价是指“对某一领域或具体临床问题所有相关指南或推荐意见进行系统、全面地检索，对其制订原则和方法进行严格评价，以呈现某领域指南或推荐意见的现状和证据空白”^[7]。与经典干预性系统评价相比，指南的系统评价开始时间相对较晚，其方法仍在进一步完善中，但发展速度很快，涵盖了内外、妇、儿、护理、检验、口腔等多个领域^[8]。

根据研究目的和研究内容，指南的系统评价主要分为以下4种类型（表1）：（1）关注某领域指南方法学质量的系统评价，如使用指南研究与评价工具AGREE II进行质量评价^[9]；（2）关注某领域指南报告质量的系统评价，如使用RIGHT工具进行评价^[10]；（3）关注某领域指南的证据质量及推荐意见方向、强度和内容的系统评价，如使用Alper等学者提出的指南推荐意见分类标准（表2）^[5]；（4）单纯针对指南某部分内容进行的评价，如专家组构成、利益冲突、分级标准等情况，对其进行调

查性分析和研究。

2 指南系统评价的制作步骤

指南系统评价制作的关键步骤与经典干预性系统评价相似，但仍存在一定差异，主要表现在注册、制作团队、结构化问题、对指南的检索与评价、规范化报告和更新等方面（表3）。

3 指南系统评价存在的问题与挑战

系统评价和临床实践指南经过了近30年的发展，各自建立了一套较为完整的方法学体系，具有专门的

组织和机构对其开展持续研究，如Cochrane协作网和国际指南联盟（Guidelines International Network）。但对指南进行系统评价，目前还是一个新的交叉研究方向，其中的很多方面尚缺乏深入细致的研究，国际上也缺乏统一的规范和标准。其中较为突出的问题包括以下6个方面：（1）缺乏规范化制作指南系统评价的标准和流程，包括类似PICO格式化问题的定义、检索策略的构建、数据提取的要求等；（2）缺乏规范化报告指南系统评价的清单；（3）缺乏评价指南系统评价风险偏倚的工具；（4）缺乏对推荐意见进行对比分析的方法；（5）缺乏量化合并数据的方法；（6）缺乏判断指南系统评价是否过期的标准和更新的方法等。

表1 指南系统评价的类型

类型	举例
方法学质量评价	为了解中国指南现状和发展趋势，对已发表的中国指南进行AGREE II评价，共纳入1993至2010年发表在115种医学期刊上的269部指南，结果显示“范围和目的”（19%）及“清晰性”（26%）平均得分较低，而“严谨性”（7%），“参与人员”（8%），“应用性”（6%）及“独立性”（2%）平均得分极低；与其他国家的指南质量相比，中国指南AGREE II评价得分较低，但随着时间的推移，指南质量会有一定提高 ^[11]
报告质量评价	ART的快速发展，使得该领域的指南数量越来越多，但目前ART指南的报告质量未知，采用RIGHT工具的22个条目对15部ART指南进行报告质量评价，结果显示平均每部指南仅报告了1/3的条目，在经费资助方面的报告质量尤其差 ^[12]
证据质量及推荐意见方向、强度和内容的评价	为探究围手术期糖尿病管理指南的质量、一致性和证据支持情况，对14部指南及其342条推荐意见进行分析，发现不同指南对手术当天停止用药情况的推荐意见不一致：两部指南建议，当患者空腹时应停止晨间美格列脲的剂量，而在晨间进行手术时应停止晨间磺脲的剂量，在下午进行手术时应停止一天的磺脲剂量；另一部指南建议，不管手术时间如何，只保留晨间磺脲剂量；多数推荐意见（78.9%）未详细说明证据支持，仅71条（20.8%）推荐意见使用了证据分级标准 ^[13]
针对指南某部分内容的调查性分析	为探究WHO指南DOI和COI报告情况，对176部WHO指南进行调查分析，其中128部（73%）报告了资助者，最常见的是政府资助；156部（89%）报告了外部资助者的DOI，与非经济COI相比，经济COI的报告频率更高，因此，报告内容尚需在资助者及其作用、DOI程序和COI管理方面有所改进 ^[14] ；指南推荐意见表述模糊或格式不统一、不规范，可能会对读者阅读和使用指南造成一定的障碍或误解。以痛风领域指南为例，对15部痛风指南中的推荐意见进行系统分析，结果发现痛风领域指南推荐意见的报告和呈现具有许多共性和差异，如13部指南正文均呈现推荐意见，其中8部指南的推荐意见采用不同于正文的特殊格式，包括加粗、斜体、加框、特殊颜色、缩小字体5种形式 ^[15]

ART：辅助生殖技术；AGREE II：指南研究与评价工具；WHO：世界卫生组织；DOI：利益声明；COI：利益冲突；RIGHT：卫生保健实践指南的报告条目

表2 指南推荐意见分类标准^[5]

分类	内容
推荐意见高度一致	推荐意见均为方向一致的强推荐 (1)不同来源的指南均对临床实践给出了强推荐或对干预措施利弊平衡的结果具有高度确定性 (2)从证据到推荐的过程和解释合理（主要体现在系统评价、一致的多学科专业知识及明确的偏好和价值观方面） (3)对推荐意见说明无争议
推荐意见比较一致	推荐意见方向一致，但强度不一致 (1)不同来源的指南均对临床实践给出了相同方向的推荐 (2)但并非所有推荐意见都是强推荐
推荐意见不一致	符合下列情况之一 (1)不同来源的指南中，≥1条推荐意见支持某种干预措施，同时≥1条推荐意见反对该支持措施 (2)不同来源的指南中，≥1条推荐意见支持某种干预措施，同时推荐意见提出支持或反对使用该干预措施的证据不足

表 3 经典干预性系统评价与指南系统评价制作的差异性

主要步骤	经典干预性系统评价	指南系统评价
注册	PROSPERO 或 Cochrane/Campbell 图书馆	PROSPERO 或 IPGRP
制作团队	多学科团队，主要由医务人员、文献检索专家和统计学家组成	多学科团队，但要求包含指南方法学家，有时也需要指南制订人员的参与
结构化问题	PICO 原则：P (Population) 代表研究对象，I (Intervention) 代表干预措施，C (Comparison) 代表对照措施，O (Outcome) 代表结局	PICAR 原则：P (Population, Clinical Indication and Condition) 代表人群、临床指征及病情，I (Intervention) 代表干预措施，C (Comparator, Comparison, and Content) 代表比较对象，A (Attributes of the CPG) 代表指南的特性，R (Recommendation Characteristics and other Considerations) 代表推荐意见特征及其他方面 ^[7]
检索来源	(1) 经典数据库：英文数据库包括 PubMed、EMBASE、Web of Science 等；中文数据库包括中国知网、中国生物医学文献服务系统及万方数据知识服务平台 (2) WHO 临床试验注册平台、Cochrane 临床对照试验中心注册数据库等 (3) 关键期刊 (4) 追溯纳入研究的参考文献 (5) 灰色文献	除经典数据库外，还需特别检索以下资源： (1) 英国国家健康与临床优化研究所、苏格兰院际指南网络、国际指南协作网、WHO 和其他医学行业学会/协会的网站 (2) 国际实践指南注册平台 (http://www.guidelines-registry.org/) (3) DynaMed、UpToDate 等循证医学数据库
纳入和排除标准	根据 PICO 确定纳入和排除标准	可参考 PICAR，如对指南进行评价，则以指南为单位确定纳入和排除标准；如对某一推荐意见进行对比或分析，则以推荐意见为单位确定纳入和排除标准
方法学质量评价	随机对照试验使用 Cochrane 推荐的 ROB 评估工具，非随机对照试验使用 ROBINS-I 评估工具，队列研究和病例对照研究使用 NOS 量表等	常使用 AGREE II 工具对指南进行评价，此外正在研发的 AGREE-REX 工具可用于评估推荐意见的临床可信度和可实施性 ^[16]
数据提取	对所纳入的原始研究提取预先设定的定性或定量数据	对所纳入的指南或推荐意见提取预先设定的定性或定量数据，定性数据较常见
数据合成	合适的情况下进行 Meta 分析，并进行亚组分析及敏感性分析	通常以描述性分析为主，可尝试使用 Meta 综合或 Meta 流行病学研究 ^[17]
报告规范	PRISMA 声明 ^[18]	尚无专门的报告规范，可参考 ENTREQ ^[19]
更新	出现新证据，能更好地描述亚组特征的指标，更优的治疗方法或采用新的结局指标均可作为更新的原因 ^[20]	尚无具体的更新标准

PROSPERO：国际化前瞻性系统评价注册数据库；IPGRP：国际实践指南注册平台；CPG：临床实践指南；ROB：偏倚风险评价工具；ROBINS-I：非随机干预研究评价工具；PRISMA：系统评价和 Meta 分析优先报告条目；ENTREQ：定性系统评价报告规范

4 对我国学者开展指南系统评价工作的思考与建议

指南系统评价能够全面、客观地呈现某一领域或具体临床问题所有相关指南的现状和差异，对我国指南制订者、临床医务人员和研究人员意义重大，因此，开展指南系统评价相关研究至关重要。以下就如何更好地开展指南系统评价工作，提出 3 个方面的思考和建议。

4.1 指南制订者层面

在开启任何一部新指南制订工作之前，应对当前已有的国内外指南进行系统评价。基于评价结果，讨论是否有必要和有价值继续制订新的指南，或对已有的高质量指南直接使用或进行改编，以避免重复工作。如果决定继续制订新指南，应在撰写

新指南时，将指南的系统评价作为制订新指南的重要构成部分，从而使指南的使用者能够更加全面地了解该领域相关指南的情况。

4.2 医务人员层面

针对本地区或医院的诊疗问题，设立专门的项目，开展指南的系统评价，将结果整合入临床路径、计算机辅助决策系统或电子病历系统。如上述工作无法开展，则检索和评价已发表指南的系统评价，编撰指南工作手册，用于指导临床实践。如以上两方面均不可及，则建议就某个具体的临床问题，对当前指南中列出的推荐意见进行评估和对比，遴选和使用最适的推荐，以避免仅依据某一部或几部指南给出的建议而对临床实践造成误导和偏倚。

4.3 研究人员层面

应积极探索指南系统评价的制订标准，开发更有针对性的指南系统评价质量评价工具。如前所

述, 国内外已有 10 多种针对指南方法学的质量评价工具, 但均非针对指南系统评价而开发。建议参考 Cochrane 偏倚风险评估工具, 开发兼具普适性和针对性的指南偏倚风险评估工具。目前针对系统评价的报告标准 PRISMA 和针对指南的报告标准 RIGHT 及其扩展版超过 10 种, 但尚无针对指南系统评价的报告标准。我国学者可与 PRISMA 或 RIGHT 工作组合作, 讨论联合研发指南系统评价扩展版的可行性。此外, 指南系统评价的更新, 目前尚无相关研究, 可在该空白领域尝试探索。

5 小结

不同指南制订机构在同一领域制订的指南无论在质量上还是推荐意见方面均存在一定共性或差异, 故通过对指南进行系统评价能够更客观地呈现指南的现状和差异。基于流行病学研究的系统评价方法适应指南研究的需求, 但需更有针对性的结构化制作流程及更加清晰透明的报告规范。指南制订者、医务人员及研究人员应充分利用当前已有的指南和推荐意见, 提升医疗服务质量和效率。

参 考 文 献

- [1] 陈耀龙, 罗旭飞, 王吉耀, 等. 如何区分临床实践指南与专家共识 [J]. 协和医学杂志, 2019, 10: 403-408.
- [2] Chen YL, Zhao C, Zhang L, et al. Toward evidence-based Chinese Medicine: status quo, opportunities and challenges [J]. Chin J Integr Med, 2018, 24: 163-170.
- [3] Gao Y, Wang J, Luo X, et al. Quality appraisal of clinical practice guidelines for diabetes mellitus published in China between 2007 and 2017 using the AGREE II instrument [J]. BMJ Open, 2019, 9: e022392.
- [4] Wang D, Yu Y, Chen Y, et al. Assessing the quality of global clinical practice guidelines on gout using AGREE II instrument [J]. J Clin Rheumatol, 2020, 26: 54-59.
- [5] Alper BS, Price A, van Zuuren EJ, et al. Consistency of recommendations for evaluation and management of hypertension [J]. JAMA Netw Open, 2019, 2: e1915975.
- [6] Chen Y, Wang C, Shang H, et al. Clinical practice guidelines in China [J]. BMJ, 2018, 360: j5158.
- [7] Johnston A, Kelly SE, Hsieh SC, et al. Systematic reviews of clinical practice guidelines: a methodological guide [J]. J Clin Epidemiol, 2019, 108: 64-76.
- [8] 王琪. 中国临床实践指南质量评价研究 [D]. 兰州: 兰州大学, 2017.
- [9] Brouwers MC, Kho ME, Browman GP, et al. AGREE II: advancing guideline development, reporting and evaluation in health care [J]. CMAJ, 2010, 182: E839-E842.
- [10] Chen Y, Yang K, Marušić A, et al. A reporting tool for practice guidelines in health care: the RIGHT statement [J]. Ann Intern Med, 2017, 166: 128-132.
- [11] Chen YL, Yao L, Xiao XJ, et al. Quality assessment of clinical guidelines in China: 1993-2010 [J]. Chin Med J (Engl), 2012, 125: 3660-3664.
- [12] Xiao Y, Jiang L, Tong Y, et al. Evaluation of the quality of guidelines for assisted reproductive technology using the RIGHT checklist: a cross-sectional study [J]. Eur J Obstet Gynecol Reprod Biol, 2019, 241: 42-48.
- [13] Song X, Wang J, Gao Y, et al. Critical appraisal and systematic review of guidelines for perioperative diabetes management: 2011-2017 [J]. Endocrine, 2019, 63: 204-212.
- [14] Wang X, Chen Y, Yao L, et al. Reporting of declarations and conflicts of interest in WHO guidelines can be further improved [J]. J Clin Epidemiol, 2018, 98: 1-8.
- [15] Yang N, Yu Y, Zhang A, et al. Reporting, presentation and wording of recommendations in clinical practice guideline for gout: a systematic analysis [J]. BMJ Open, 2019, 9: e024315.
- [16] AGREE-REX Research Team. AGREE-REX: Recommendation Excellence 2019 [EB/OL]. <https://www.agreerest.org/resource-centre/agree-rex-recommendation-excellence/>.
- [17] O'Sullivan JW, Albasri A, Koshiaris C, et al. Diagnostic test guidelines based on high-quality evidence had greater rates of adherence: a meta-epidemiological study [J]. J Clin Epidemiol, 2018, 103: 40-50.
- [18] Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement [J]. BMJ, 2009, 339: b2535.
- [19] Tong A, Flemming K, McInnes E, et al. Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ [J]. BMC Med Res Methodol, 2012, 12: 181.
- [20] Garner P, Hopewell S, Chandler J, et al. When and how to update systematic reviews: consensus and checklist [J]. BMJ, 2016, 354: i3507.

(收稿日期: 2020-05-09)