

STAR 指南评级提高检验医学指南与共识的科学性、透明性和适用性

STAR 检验医学专科委员会

通信作者: 门剑龙, Email: menjianlong@tmu.edu.cn; 彭明婷, Email: mtpeng@nccl.org.cn

【摘要】 在医疗技术迭代发展以及健康理念快速更新的背景下, 临床实践指南和专家共识对推动临床诊疗规范化、提高临床干预安全性和有效性具有重要意义。然而, 专家团队对指南制订的方法学和相关规范认识尚不全面, 导致部分指南质量存在不足。为进一步提高中国指南和共识的整体水平, 指南科学性、透明性和适用性的评级工作组(简称 STAR 工作组)在 2022 年制定了综合性指南评级系统, 为国内指南评价体系提供了有力的评价工具。该文结合 STAR 评级工具的具体要求, 对指南和共识制订过程中的 11 个领域(包括注册、计划书、资助、工作组、利益冲突、临床问题、证据、共识方法、推荐意见、可及性和其他)进行分析和说明, 为指南制订者提供参考。

【关键词】 检验医学; 指南; 共识; 评级工具

STAR guideline rankings improve the scientificity, transparency and applicability of laboratory medicine guidelines and consensus

STAR Specialized Committee on Laboratory Medicine

Corresponding author: Men Jianlong, Email: menjianlong@tmu.edu.cn; Peng Mingting, Email: mtpeng@nccl.org.cn

【Abstract】 In the background of the iterative development of medical technology and the rapid update of health concepts, clinical practice guidelines and expert consensus play a central role on promoting the standardization of clinical diagnosis and treatment and improving the safety and effectiveness of clinical interventions. However, insufficient understanding of the methodology and relevant norms for guideline the development of clinical practice guidelines and expert consensus by the expert team might lead to deficiencies in the quality of some guidelines and expert consensus. In order to further improve the overall level of Chinese guidelines and consensus, the Working Group on the Scientific, Transparent and Applicable Rankings (STAR) for guidelines had developed a comprehensive guideline evaluation system in 2022, and provided an evaluation tool for the domestic guideline evaluation system. Based on the detailed specific requirements of the STAR ranking tool, present manuscript analyzed and explained the 11 fields related to the guideline and consensus development processes (including registration, plans, funding, working groups, conflicts of interest, issues of clinical, evidence, consensus methods, recommendations, accessibility and others) to provide reference for guideline makers.

【Key words】 Laboratory medicine; Guideline; Consensus; GRADE approach

临床实践指南(简称指南)是针对临床问题, 基于系统评价的证据, 在通过权衡不同干预措施利弊的基础上, 为医疗服务提供的最佳推荐意见^[1], 旨

在指导医务人员、政策决策者及患者作出合理决策。随着诊疗技术及疾病谱的不断变化, 临床实践日益复杂, 因此指南在构建规范化诊疗路径中具有

DOI: 10.3760/cma.j.cn114452-20240729-00408

收稿日期 2024-07-29 本文编辑 干岭

引用本文: STAR 检验医学专科委员会. STAR 指南评级提高检验医学指南与共识的科学性、透明性和适用性[J]. 中华检验医学杂志, 2025, 48(1): 20-27. DOI: 10.3760/cma.j.cn114452-20240729-00408.



重要作用,以确保临床救治的安全性和有效性。与原始研究、系统评价及荟萃分析等研究模式不同,指南的特点在于综合分析最新临床研究证据、针对具体临床问题提出推荐和建议。高质量的指南可规范医疗行为、提升临床决策水平和服务质量,同时也是医疗资源合理使用及医疗政策制定的可靠依据^[2]。

随着检验医学技术的不断发展,各亚专业围绕着临床应用、质量控制、技术规范以及实验室建设等热点问题制订了大量指南及专家共识(简称共识),推动了实验室流程管理及临床实践能力的提升。然而,由于编制团队对指南方法学和相关规范尚缺乏全面认识,导致部分指南和共识在科学性、透明性和适用性方面存在争议与不足,影响了其有效性和安全性,也造成了资源浪费^[3-4]。因此,本文结合指南编制的关键环节和评级要求,对指南和共识的制订过程进行解析,旨在为检验医学及相关领域的指南制订者提供参考,提升指南和共识的质量与实用性。

一、指南制订的指导文件

近十年,用于指导指南制订的文件主要包括 AGREE 清单^[5]、卫生保健实践指南的报告规范清单^[6]、中国临床实践指南评价体系^[7]。2022 年,《中国制订/修订临床诊疗指南的指导原则(2022 版)》^[1]进一步整合了国内外指南方法学的研究进展,为指南制订提供了最新参考。该指导原则主要从目的、总则、程序与方法、传播、更新与修订、实施与评价 6 个部分进行阐述。其中,程序与方法涉及指南编写过程的具体要求,包括启动与规划、确定指南类型、注册、撰写计划书、成立工作组、管理利益冲突、调研临床问题、检索和评价证据、形成推荐意见、撰写与发表等。

二、STAR 工作组与指南评级

STAR 工作组是一个由研究者主导的全国性学术组织。该组织由世界卫生组织指南实施与知识转化合作中心、中华医学会、中国医学科学院联合发起,于 2021 年在兰州成立。STAR 工作组的主要职能包括指南相关的评级、培训、认证、研究和实施。2022 年,STAR 工作组发表了《针对临床实践指南科学性、透明性和适用性的评级工具研发》^[8],为指南评价体系提供了有效工具,推动了高质量指南和专家共识的制订与应用。STAR 评级体系涵盖了 11 个领域(包括注册、计划书、资助、工作组、利益冲突、临床问题、证据、共识方法、推荐意见、可及

性和其他),共 39 项条目,每项条目根据权重赋值(0.9~6.4 分),总计 100 分^[8],以全面评估指南质量(表 1)。

三、STAR 评级体系的 11 个领域

1. 注册:注册是确保指南制订过程透明性的重要环节。在指南制订启动前,应通过国际指南注册与透明化平台登记指南标题、目的、参与人员、制订方法及利益冲突等重要信息。这不仅保证了制订过程的透明性,还可避免指南的重复制订,同时通过公开注册信息,提高了知晓率,可进一步促进指南的实施与传播。注册核心内容包括:制订团队基本信息,指南制订背景,文献证据的检索策略、证据合成与分级方法、资助来源与利益冲突管理等。在指南投稿或发布时应提交注册平台与注册号。

STAR 评级关于“注册”的评价条目包括:(1)是否进行了注册;(2)是否在指南中提供了注册平台和注册号信息。根据 STAR 工作组对 2022 年医学期刊发表的国内指南与共识的评级数据分析,纳入的 22 篇在国内杂志上发表的检验学科指南与共识(包括 2 篇指南及 20 篇共识)中,仅 2 篇共识完成了注册,仅 1 篇共识提供了注册号。这表明,目前国内的专家团队在指南与共识注册方面的意识普遍不足,导致多数共识和指南在制订之初就缺乏规范性。

2. 计划书:计划书是一份规划指南核心内容设计与制订流程的计划性文件,主要包括:(1)指南的基本信息(包括标题、副标题、术语和缩略语、通信作者等);(2)背景(包括临床必要性的简要描述、总目标和具体目的、目标人群、指南使用者和应用环境、指南制订小组等);(3)制订方法;(4)证据(包括证据的检索与评价等);(5)推荐意见(包括形成临床问题、推荐意见的原理和解释说明、从证据到推荐的决策过程和方法等);(6)评审和质量保证(包括外部评审、质量保证等);(7)资助与利益冲突管理;(8)工作组的构成;(9)指南可及性等^[1]。指南计划书可明确提供工作架构、形成路径图及时间表,进一步推动指南制订的高效率、高质量完成。计划书应由指南的核心专家组拟定,并由所有参编专家共同审核定稿。为提升指南制订过程的透明性和科学性,计划书应可从公开平台获取(如期刊发表、注册平台或专业网站等)。在指南发表时,应明确说明指南计划书的先前发布信息、获取渠道、内容调整说明以及计划书是否按预定要求完成。

STAR 评级关于“计划书”的评价条目包括:

(1)撰写了计划书(包含完整的制订方法和时间节点);(2)计划书能够在公开平台获取。然而,2022年的STAR评级结果显示,22篇检验相关指南和共识中,仅3篇能提供计划书,且均无法从公共平台获取。上述数据表明,检验领域多数指南和共识在启动之初,尽管成立了专家组、分配了编写任务,并初步确定了成稿及定稿会时间等,但通常未制定完整的计划书,且缺乏获取计划书的公共渠道。

3.资助:对资助的正确管理是指南制订规范化的重要保障,在此过程中应注意资助使用的透明性和合理性,避免资助对指南推荐意见的科学性产生影响。STAR评级关于“资助”的评价条目包括:(1)指南中说明资助来源(如说明资助具体来源、提及某制订步骤的资助来源);(2)说明资助在指南制订中的作用(如资金在指南中的具体作用、提及某制订步骤的资助来源);(3)说明指南推荐意见未受资助影响^[8]。长期以来,资金管理及利益冲突处理一直是影响指南“编撰独立性”的重要因素,大多数WHO指南会在文中报告资金来源,充分公开利益关系和外部捐助者的利益冲突。2022年的STAR评级结果显示,22篇检验相关指南和共识中,有7篇说明了资助来源,但均未说明资助在指南制订中的作用;其中仅有2篇说明了其资金来源于政府基金、公益基金或其他非商业性资金,但未明确说明指南制订或推荐意见形成未受资助影响。《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》^[4]指出,国内334篇指南和1143篇共识中,“说明资助在指南制订中的作用”条目的达标率偏低,表明目前国内指南和共识在该方面的不足,降低了指南和共识的透明度,也难以证明其制订过程和推荐意见可规避潜在的利益干扰和偏倚风险。

4.工作组:工作组负责指南的制订或修订,由具有专业知识和经验的专家组成,通过共同合作以确保指南的准确性、实用性和权威性。不同专业领域指南,应根据指南制订的目的、具体内容及涉及领域对工作组类型、规模、成员组成及职责分工进行调整。检验专业指南主要聚焦于检验技术,内容涵盖方法学、质量控制、流程管理和临床实践等方面,工作组的人员构成以检验专家为主,同时应纳入循证医学专家(遴选临床问题、检索和评价研究证据)和卫生经济学专家(减少临床专家在经济和专业方面潜在的利益冲突),如果指南涉及临床应

用的推荐意见,应纳入临床专家;如果涉及临床用药和护理的推荐意见,应纳入临床药师与护师,如果涉及患者意愿或患者阅读版的指南,还应纳入患者代表^[1-2]。构建多学科指南制订团队有助于获得客观全面的推荐意见,这也是评价指南方法学质量的重要标准。指南工作组通常应包括指导委员会(设立首席专家和首席方法学家)、秘书组、证据评价组、推荐意见共识组和外审组等。

STAR评级关于“工作组”的评价条目包括:(1)列出参与人员的机构(描述参与指南制订的所有个人的专业职称和/或职务、工作单位、所在部门等信息);(2)说明参与人员的分组情况;(3)说明参与人员的职责;(4)明确提出纳入除本专业以外的其他2个及2个以上专业的人员;(5)明确提出工作组包含方法学或循证医学专家^[8]。2022年STAR评级结果显示,在22篇检验相关的指南和共识中,全部文章列出了参与人员的所属机构(说明了专业领域),但均未提供完整的分组信息(包含证据组和共识组);只有3篇在方法部分说明了各组人员(或参与人员)的分工与职责;有12篇明确提出纳入了至少2个本专业以外人员;仅1篇明确说明指南制订组包括指南方法学家或循证医学专家。上述数据表明,尽管部分检验专业指南制订团队已认识到分组合作、职责分工以及建立多学科团队的重要性,但在分组分工规范性方面仍有不足。由于大多数指南和共识仅列出了参与者,因此对指南和共识的科学性和适用性具有潜在影响。需强调的是,流行病学专家不应被认定为指南方法学家,应通过专家机构(如循证医学中心等)对方法学或循证医学专家进行专业认定。

5.利益冲突:在制订指南过程中,当主要利益(公众利益)受次要利益(个人利益)过度影响时,会产生利益冲突,干扰指南制订的独立性和可靠性^[9],引起有益效果被高估而危害性被低估,进而可能导致指南制订过程及推荐意见发生严重偏倚^[1-2]。利益冲突通常分为经济与非经济利益冲突,或直接与间接利益冲突。存在利益冲突的专家在收集和遴选临床问题、制作和使用系统评价、讨论和形成推荐意见等环节中,更易作出倾向性的判断和决定。因此,在指南制订前,应建立独立的利益冲突管理小组(或委员会)并制订相应的管理办法。所有参与指南制订的成员均需填写《利益冲突声明表》,以此评估其利益冲突程度,并采取相应措施(如存在严重利益冲突的成员不应参与指南制订

工作),且在指南发表时应公布《利益冲突声明表》。值得注意的是,尽管国际上已存在关于利益冲突的管理方案,但不应简单套用,需充分结合国内本专业领域的实际情况进行优化完善,以增强管理的可行性及有效性。

STAR 评级关于“利益冲突”的评价条目包括:(1)说明有无利益冲突(提供《利益冲突声明表》);(2)提供详细的利益冲突管理办法^[8]。2022 年的 STAR 评级结果显示,在 22 篇检验相关的指南和共识中,2 篇未做任何利益冲突说明,16 篇只报告了无利益冲突,4 篇能够提供《利益冲突声明表》并报告无任何利益冲突;仅 1 篇完整说明了利益关系的评估方法和利益冲突的处理方式。上述数据表明,国内多数指南制订团队虽已认识到利益冲突可对指南的客观性及公正性产生影响,但因缺乏实际操作经验,且部分指南、共识形成是基于对国外指南共识的借鉴或翻译,导致对利益冲突处理既不充分也不规范,进而影响了读者及使用者对指南和共识的信任度。

6. 临床问题:临床问题涉及临床诊疗、技术规范、患者管理、卫生经济学等方面,是在临床实践中亟待解决的关键问题。构建合理且被临床广泛认可的临床问题,是形成高质量推荐意见的重要基础。临床问题与临床实践的相关性、表达形式及其能否满足指南使用者的需求,直接影响指南与共识的普及程度及应用效果。临床问题的形成过程包括 2 个关键环节,一是收集和遴选临床问题,二是解构临床问题。

临床问题的来源主要通过 3 个途径:文献检索(如相关指南和共识、系统评价及原始研究)、用户调查及专家咨询。指南制订团队需对收集的原始问题进行整合、精练、去重及合并,并根据其临床问题发生顺序、逻辑关系以及临床价值等形成合理的问题表述。

经遴选后的临床问题应以 PICO^[1] [即人群(population, P)、干预(intervention, I)、对照(comparison, C)和结局(outcome, O)^[1-2]] 形式进行解构。在检验医学专业中,“人群”是描述检验技术或检验指标适用的群体特征(如性别、年龄、疾病类型及诊疗目的);“干预”则涵盖了检验技术或指标的应用场景、目的、规则、流程、检验前评估、临界值、联合应用等;“对照”是比较检验技术或指标在不同患者群体、健康人、病情阶段、治疗方案间的应用价值及效果,或比较不同检验技术或指标的差

异;“结局”是指检验技术或指标在诊断和鉴别诊断、风险分层、预后判断、患者风险与获益评估、卫生经济学等方面的临床应用价值。

STAR 评级关于“临床问题”的评价条目包括:(1)明确提出指南拟解决的临床问题;(2)说明通过文献调研(指南、系统评价及原始研究)、用户调查或专家咨询收集临床问题;(3)说明临床问题遴选的方法;(4)临床问题以 PICO 形式进行解构^[8]。2022 年的 STAR 评级结果显示,22 篇检验相关的指南和共识中,18 篇在文中明确列出拟解决或关注的临床问题;12 篇说明了通过文献调研、用户调查或专家咨询收集临床问题;3 篇阐述了临床问题遴选方法;3 篇采用 PICO 等形式列出了临床问题,另有 3 篇虽提出按 PICO 或其他框架进行解构,但未明确列出具体临床问题。上述数据表明,尽管多数专家团队在制订指南和共识时,对拟解决的临床问题有较为明确的规划,但在临床问题的遴选过程及方法上存在缺失与不足,且未能广泛采用 PICO 的形式解构临床问题,这可能影响临床问题的合理性及全面性。此外,由于缺乏基于 PICO 的问题分析,也容易导致临床问题,例如新指标或新技术,在临床适用性方面可能存在缺陷。

7. 证据:指南制订团队应围绕临床问题尽可能收集所有相关的研究证据,并进行筛选和评价。首先,依据不同数据库特性,设计相应的检索策略,针对临床问题从 PICO 4 个方面设置检索词(包括主题词和自由词)。其次,制定纳入、排除标准,以确保文章筛选效率和证据质量,同时也可节约时间及成本。系统评价(包括定量的 Meta 分析和定性的系统评价)是支撑指南、共识推荐意见的最高质量证据^[2, 10],因此应首先检索与临床问题相关的系统评价,并采用系统综述质量评价工具^[11]或系统综述偏倚风险评估工具^[12]对检索出的系统评价进行质量分析和适用性评估。鉴于时间和资源的有限性,亦可直接采用近 2 年内发表的高质量系统评价。然而,若出现系统评价缺失、适用性不足、时效性较差或质量较低等情况,应检索原始研究证据,并自行开展系统评价。需指出的是,无论是已发表还是自行开展的系统评价,均需进行证据质量评价和质量分级(如 GRADE 分级)^[13]。

STAR 评级关于“证据”的评价条目包括:(1)主要推荐意见有明确的参考文献;(2)说明系统检索证据(检索数据库和检索词);(3)说明证据纳入排除标准;(4)评价证据的偏倚风险或方法学质量;

(5)对证据结果进行汇总分析;(6)说明证据质量分级标准;(7)提供证据总结表或分级依据;(8)可追溯到系统评价全文;(9)列出缺乏证据的临床问题,提供未来研究方向^[8]。2022 年的 STAR 评级结果显示,22 篇检验相关的指南和共识中,18 篇的推荐意见有明确的参考文献,仅 1 篇完整说明了系统检索证据(检索数据库和检索词),6 篇仅说明了检索数据库或检索词,仅有 5 篇说明了纳入标准,全部指南、共识均未说明排除标准;在对研究证据的质量评价中,仅 1 篇对研究的偏倚风险进行评价且提供了所有纳入研究的偏倚风险结果,1 篇说明了汇总研究证据的具体分析结果,2 篇仅说明了汇总研究证据的结论,4 篇提供了证据质量分级标准;此外,只有 1 篇提供了每条推荐意见关联证据的具体分级依据,1 篇的推荐依据部分所引用的系统评价或补充材料可追溯到系统评价全文;有 9 篇说明了目前的研究空白或未来的研究方向。上述数据表明,由于缺乏规范的检索流程和排除标准,导致多数指南和共识纳入的证据质量无法评估,甚至质量较低;同时由于缺失系统评价过程,未能对证据质量进行评价和分级,使推荐意见的合理性和安全性缺乏可信度。

8. 共识方法:共识形成方法包括非正式和正式两种方法。非正式共识形成方法指在无规范化流程情况下,通过自由讨论对某一问题达成共识。而正式共识形成方法指在预先制定原则和程序下,经专家团队认同,通过讨论等方式获得共识或推荐意见。基于循证的指南和共识需采用正式共识形成方法,如德尔菲法、名义群体法、共识会议法及 GRADE 网格法^[2]。

德尔菲法是指通过反复的结构化的方式搜集参与者意见,其主要流程包括专家遴选、制作专家调查问卷、发送调查问卷^[14-15]。德尔菲法通常需实施 2~4 轮次,具体轮次需根据专家意见的协调情况而定(以专家意见趋于一致为问卷工作结束标准)。

名义群体法是指在决策过程中对群体成员的讨论或人际沟通加以限制,群体成员独立思考、平等参与(避免在讨论甚至产生意见冲突的状态下提交自己的想法和建议)^[16]。然后再通过群体讨论,分析、评价和排序每个想法和建议,最终选择综合排序最高的想法,该过程可进行多次直到达成一致。

共识会议法是指组织一组人员参与会议,根据现有证据就某一问题进行讨论并达成共识^[2, 17]。

共识会议的主要环节包括遴选会议成员(应尽可能选择不同观点的专家,给出较客观且专业化的意见)、组织会议(包括公开讨论会和委员会)、形成共识声明或推荐意见(通过投票、排序、公开讨论等非结构化的互动方式提供证据,整合信息并达成共识或推荐意见)。

GRADE 网格法是指在达成共识存在困难时,当推荐或反对某项建议的比例 $>50\%$ 时,视为达成共识;若该比例 $\geq 70\%$ 则可视作强推荐意见;如未达成共识,需进行下一轮投票以重新审议^[18]。德尔菲法、名义群体法及共识会议法在达成共识意见时均可采用 GRADE 网格法作为辅助工具。

STAR 评级关于“共识方法”的评价条目包括:(1)说明推荐意见的共识方法;(2)说明如何基于证据质量以外的其他因素进行共识;(3)提供完整的共识过程记录^[8]。2022 年的 STAR 评级结果显示,在 22 篇检验相关的指南和共识中,5 篇提供了形成推荐意见的共识方法;3 篇明确提出了除基于证据质量以外的其他共识形成因素(如经济学、患者偏好和价值观、利弊权衡、可及性、公平性及可接受性等);3 篇提供了完整的共识过程记录,包括共识形成轮次、每轮达成的推荐意见数量、各推荐意见共识度及相应标准,同时收集、反馈专家意见。上述数据表明,多数指南、共识在形成方法设计上存在不足,使得形成过程缺乏透明性、公平性,进而影响推荐意见的科学性和适用性,导致指南发表后,推荐意见的代表性不足,引起相关专业领域的争议。

9. 推荐意见:推荐意见的表达应明确、清晰且具有可操作性,同时需明确给出推荐强度及证据等级,尽可能按照 PICO 的构架阐释推荐意见,并结合相关证据对推荐意见进行解释或说明^[1-2]。STAR 评级关于“推荐意见”的评价条目包括:(1)明确列出推荐意见,如采用图表、放大或加粗字体、下划线等方式表达;(2)说明每条推荐意见的推荐强度;(3)提供每条推荐意见的解释说明;(4)说明推荐意见实施过程中的注意事项^[8]。2022 年的 STAR 评级结果显示,在 22 篇检验相关的指南和共识中,15 篇有明确可识别的推荐意见,如以图表,放大或加粗字体,下划线等方式呈现,5 篇仅提及了“推荐”“建议”,2 篇未提及任何推荐或建议;4 篇推荐意见明确提供了推荐强度;5 篇对各条推荐意见做了解释说明(如证据描述、经济学考量、患者偏好和价值观、利弊分析、可及性、公平性、可接受性等),6 篇提供了部分说明;10 篇阐述了推荐意见在实施过

程中的注意事项。上述数据表明,多数指南、共识因前期证据评价不足,未能提供推荐强度,多数指南、共识对推荐意见的解释和注意事项说明不充分,从而影响了推荐意见在临床实践中的参考价值。

10. 可及性:可及性是指指南能否被广泛知晓、普及应用并充分发挥积极效果的关键因素。STAR 评级关于可及性的评价条目包括:(1)通过指南文库、会议、网络等多平台发布指南;(2)提供不同用户版本的指南;(3)以图片、视频等其他形式发布指南或推荐意见;(4)指南可被免费获取^[8]。2022 年的 STAR 评级结果显示,在 22 篇检验相关的指南和共识中,仅 4 篇通过多种平台(如微信、会议照片、指南文库、新闻媒体等)发布;1 篇提供了针对不同用户群体的版本;2 篇以图片、视频等可视化形式展示推荐意见;仅 10 篇可免费获取(如在发表指南的杂志官网或微信免费获取全文)。上述数据表明,现阶段指南和共识的可及性不足,成为阻碍其在临床实践中广泛应用主要瓶颈,可及性的缺失不仅限制了指南的临床价值实现,也是对专家智慧与努力的一种浪费。

11. 其他:根据 STAR 评级要求,指南应制订推荐意见路径图^[8]。2022 年的 STAR 评级结果显示,在 22 篇检验相关的指南和共识中,仅 4 篇提供了包含部分或全部推荐意见信息的路径图,如诊疗流程图、临床路径图等。值得注意的是,鉴于检验医学领域的特殊性,许多指南共识涉及实验方法学、技术规范及临床应用,其推荐意见之间可能不存在逻辑顺序或诊断关联性,因此检验领域的指南共识是否需制作推荐意见路径图,各制订团队应结合自身学科特点进行选择 and 评估。

四、2022 年检验医学领域指南和共识 STAR 评级结果

STAR 工作组和 STAR 检验医学专科委员会对 2022 年检验医学领域指南及共识进行了 STAR 评级,共纳入 2 篇指南和 20 篇专家共识。

统计 11 个领域中每个条目的得分率(%),计算公式为:条目的得分率(%)=条目实际分数/条目理论总分数×100%。每个领域的得分率(%),计算公式为:领域下所有条目实际分数总和/所有条目理论总分数×100%。

结果显示,得分率最高的 3 个领域分别为临床问题(47.2%)、推荐意见(41.5%)和工作组(33.1%)。得分率最低的 3 个领域分别为注册

(4.5%)、计划书(5.1%)和资助(11.6%)(表 1)。

各领域具体条目得分情况:4 个条目得分率>70%,其中条目 8(说明了参与人员的机构)为 100%、条目 19(主要推荐意见有明确的参考文献)和条目 15(明确提出指南拟解决的临床问题)均为 81.8%、条目 31(明确列出了推荐意见,例如以图表、字体放大或加粗、下划线等方式呈现)为 79.5%。

13 个条目得分率<10%,其中条目 4(计划书能够在公开平台获取)和条目 6(说明了资助在指南制订中的作用)均为 0,条目 1(进行了注册)、条目 2(提供注册的平台和注册号信息)、条目 7(说明指南推荐意见未受资助影响)、条目 12(明确提出工作组包含方法学家或循证医学专家)、条目 22(评价证据的偏倚风险或方法学质量)、条目 26(可追溯到系统评价全文)和条目 36(提供不同用户版本的指南)均为 4.5%,条目 14(提供详细的利益冲突管理办法)和条目 25(提供了证据总结表或分级依据)均为 6.8%,条目 23(对证据结果进行汇总分析)和条目 37(以图片、视频等其他形式发布指南或推荐意见)均为 9.1%(表 1)。

本次 STAR 评级得分为 25.5(19.7, 31.8)分,较高的分别为 69.1 分^[19]、67.3 分^[20]和 46.5 分^[21],最低分 9.62 分。

五、总结

随着医疗技术的迅猛发展以及先进诊疗理念的不断更新,基于循证医学的指南与专家共识对推动临床诊疗规范化、保障干预措施安全性及有效性方面具有重要的意义。近年来,检验专业领域发表了许多共识与指南,为临床诊疗提供了高水平推荐意见,但许多指南制订团队对方法学和规范流程缺乏了解,导致在注册、计划书、资助管理和说明、分组与职责、利益冲突管理等环节存在不足,影响了指南的透明性、客观性及避免偏倚的能力。

此外,多数指南与共识在临床问题收集、遴选和 PICO 解构方面也表现不充分,对纳入的研究证据未设计排除标准、未评价研究证据的偏倚风险或方法学质量等,这均限制了指南推荐意见的指导作用。

此外,2022 年检验领域专家共识与指南的 STAR 评级分数普遍较低,少数指南与共识以综述或操作规范的形式呈现,而非基于循证医学证据构建。以上问题,均会对检验领域指南与共识的普及与应用产生影响,降低读者和使用者的认可度。因

表 1 纳入 2022 年 STAR 评级的 22 部检验医学领域指南、共识各条目得分率

领域	条目	赋值 (分)	得分率(%)	
			领域	条目
注册			4.5	
	1. 进行了注册	1.458		4.5
计划书	2. 提供注册的平台和注册号信息	3.517		4.5
			5.1	
资助	3. 撰写了计划书	1.878		13.6
	4. 计划书能够在公开平台获取	3.098		0
工作组			11.6	
	5. 说明了资助来源	0.958		31.8
利益冲突	6. 说明了资助在指南制订中的作用	0.905		0
	7. 说明指南推荐意见未受资助影响	1.273		4.5
临床问题			33.1	
	8. 说明了参与人员的机构	0.935		100.0
证据	9. 说明了参与人员的分组情况	1.004		45.5
	10. 说明了参与人员的职责	1.276		13.6
共识方法	11. 明确提出纳入除本专业以外的其他 2 个及 2 个以上专业的人员	1.330		54.5
	12. 明确提出工作组包含方法学家或循证医学专家	2.764		4.5
推荐意见			29.5	
	13. 说明有无利益冲突	4.368		54.5
可及性	14. 提供详细的利益冲突管理办法	4.838		6.8
			47.2	
其他	15. 明确提出指南拟解决的临床问题	6.421		81.8
	16. 说明了通过文献调研(指南、系统评价及原始研究)、用户调查或专家咨询收集临床问题	2.483		54.5
注册	17. 说明了临床问题遴选的方法	3.350		13.6
	18. 临床问题以 PICO 形式解构	4.781		20.5
计划书			19.8	
	19. 主要推荐意见有明确的参考文献	1.662		81.8
资助	20. 说明了系统检索证据	2.224		18.2
	21. 说明了证据纳入排除标准	1.526		11.4
工作组	22. 评价证据的偏倚风险或方法学质量	1.925		4.5
	23. 对证据结果进行汇总分析	2.128		9.1
利益冲突	24. 说明了证据质量分级标准	2.247		18.2
	25. 提供了证据总结表或分级依据	2.368		6.8
临床问题	26. 可追溯到系统评价全文	1.726		4.5
	27. 列出了缺乏证据的临床问题,提供未来研究方向	1.228		40.9
证据			18.4	
	28. 说明了推荐意见的共识方法(德尔非法、名义群体法、共识会议、GRADE 网格法等)	5.130		22.7
共识方法	29. 说明了如何基于证据质量以外的其他因素(经济学、患者偏好和价值观、利弊权衡、可及性、公平性、可接受性等)进行共识	3.811		13.6
	30. 提供了完整的共识过程记录	1.797		15.9
推荐意见			41.5	
	31. 明确列出了推荐意见,譬如以图表、放大或加粗字体、下划线等方式呈现	4.088		79.5
可及性	32. 说明了每条推荐意见的推荐强度	6.252		18.2
	33. 提供了每条推荐意见的解释说明	3.943		36.4
其他	34. 说明了推荐意见实施过程中的注意事项	2.753		45.5
			22.8	
注册	35. 通过指南文库、会议、网络等多平台发布指南	2.547		18.2
	36. 提供不同用户版本的指南	1.361		4.5
计划书	37. 以图片、视频等其他形式发布指南或推荐意见	1.108		9.1
	38. 指南可被免费获取	2.293		45.5
资助			18.2	
	39. 提供指南的推荐意见路径图	1.246		18.2

此规范指南制订流程,提高指南与共识的科学性、透明性和适用性,对提升检验医学指南与共识质量,推动其普及应用及临床实践具有重要意义。

执笔人:张珠博(天津医科大学总医院),马睿(天津医科大学总医院),任静(天津医科大学总医院)

STAR 检验医学专科委员会主任委员:门剑龙(天津医科大学总医院)

副主任委员:彭明婷(国家卫生健康委临床检验中心),干岭(中华检验医学杂志编辑部),任静(天津医科大学总医院)

委员:陈凯(天津市北辰医院),丁伟峰(南通大学附属医院),王国庆(天津市口腔医院),公志华(山西白求恩医院),马睿(天津医科大学总医院),朱国庆(中国医学科学院血液病医院),乔蕊(北京大学第三医院),李宏峰(天津市中医药研究院),张珠博(天津医科大学总医院)

STAR 工作组执行委员会

代表:陈耀龙(兰州大学基础医学院 兰州大学健康数据科学研究院)

STAR 工作组秘书处

代表:王旭(重庆医科大学附属儿童医院),杨楠(兰州大学基础医学院),刘辉(兰州大学基础医学院)

利益冲突 所有作者声明无利益冲突

志谢 兰州大学基础医学院、兰州大学健康数据科学研究院陈耀龙教授指导

参 考 文 献

[1] 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. DOI: 10.3760/cma.j.cn112137-20211228-02911.

[2] 王小钦, 王吉耀. 循证临床实践指南的制订与实施[M]. 北京:人民卫生出版社, 2016.

[3] 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, 世界卫生组织指南实施与知识转化合作中心. 2021 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, 2022, 102(30): 2319-2328. DOI: 10.3760/cma.j.cn112137-20220602-01232.

[4] 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. DOI: 10.3760/cma.j.cn112137-20230724-00076.

[5] Brouwers MC, Kerkvliet K, Spithoff K, et al. The AGREE reporting checklist: a tool to improve reporting of clinical practice guidelines[J]. BMJ, 2016, 352: i1152. DOI: 10.1136/bmj.i1152.

[6] 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(2):128-132. DOI: 10.7326/M16-1565.

[7] 王吉耀, 王强, 王小钦, 等. 中国临床实践指南评价体系的制定与初步验证 [J]. 中华医学杂志, 2018, 98(20): 1544-1548. DOI: 10.3760/cma. j. issn. 0376-2491.2018. 20.004.

[8] 杨楠, 赵巍, 潘旻, 等. 针对临床实践指南科学性、透明性和适用性的评级工具研发 [J]. 中华医学杂志, 2022, 102(30): 2329-2337. DOI: 10.3760/cma. j. cn112137-20220219-00340.

[9] Chen Y, Yang K, Norris SL. Managing conflicts of interest in practice guidelines panels[J]. JAMA, 2017, 318(9): 866-867. DOI: 10.1001/jama.2017.9672.

[10] 田金徽, 陈杰峰. 诊断试验系统评级: Meta 分析指导手册 [M]. 北京:中国医药科技出版社, 2015.

[11] Shea BJ, Bouter LM, Peterson J, et al. External validation of a measurement tool to assess systematic reviews (AMSTAR) [J]. PLoS One, 2007, 2(12): e1350. DOI: 10.1371/journal.pone.0001350.

[12] 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. DOI: 10.1016/j.jclinepi.2015.06.005.

[13] Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations[J]. BMJ, 2008, 336(7650): 924-926. DOI: 10.1136/bmj.39489.470347.AD.

[14] Shang Z. Use of Delphi in health sciences research: a narrative review[J]. Medicine (Baltimore), 2023, 102(7): e32829. DOI: 10.1097/MD.00000000000032829.

[15] Savic LC, Smith AF. How to conduct a Delphi consensus process[J]. Anaesthesia, 2023, 78(2): 247-250. DOI: 10.1111/anae.15808.

[16] McMillan SS, King M, Tully MP. How to use the nominal group and Delphi techniques[J]. Int J Clin Pharm, 2016, 38(3):655-662. DOI: 10.1007/s11096-016-0257-x.

[17] Ponce OJ, Alvarez-Villalobos N, Shah R, et al. What does expert opinion in guidelines mean? a meta-epidemiological study[J]. Evid Based Med, 2017, 22(5):164-169. DOI: 10.1136/ebmed-2017-110798.

[18] 田金徽, 李伦. 网状 Meta 分析方法与实践[M]. 北京:中国医药科技出版社, 2017.

[19] 中国中西医结合学会检验医学专业委员会. 流式细胞术在嵌合抗原受体-T 细胞免疫治疗相关检验中的应用专家共识 [J]. 中华检验医学杂志, 2022, 45(8): 790-801. DOI: 10.3760/cma.j.cn114452-20220218-00081.

[20] 中国医师协会检验医师分会心血管专家委员会. B 型利钠肽及 N 末端 B 型利钠肽前体实验室检测与临床应用中国专家共识[J]. 中华医学杂志, 2022, 102(35): 2738-2754. DOI: 10.3760/cma.j.cn112137-20220714-01553.

[21] 中国中西医结合学会检验医学专业委员会. 临床实验室精液常规检验中国专家共识 [J]. 中华检验医学杂志, 2022, 45(8): 802-812. DOI: 10.3760/cma.j.cn114452-20220112-00016.