

## · 讲座 ·

# 临床实践指南的制定方法和评价原则

袁帅 李亚斐

临床实践指南 (clinical practice guideline, CPG) 能帮助临床医师将最好的证据转化为最好的临床实践<sup>[1]</sup>。目前,许多国家和地区都在积极制定和推广 CPG,以 CPG 为依据规范当前的医疗卫生服务。高质量的 CPG 可降低临床实践的不一致性,减小不同医疗机构和不同临床医师之间医疗水平的差异,提高诊断的准确性,防止采用无效甚至有害的干预措施,使患者得到经济、高效的治疗方案<sup>[2]</sup>。然而,如何制定出科学公正、具有权威指导意义的 CPG 则是要考虑的重要问题。同时,目前国内外发表的 CPG 质量良莠不齐,也需要对这些 CPG 进行科学评价。本文将综述 CPG 的基本概念、制定方法和评价原则,以期帮助国内学者制定出高质量的 CPG。

## 1 CPG 的定义

2011 年美国医学研究所对 CPG 提出的定义为:通过系统综述相关科学证据以及对各种备选干预方式的利弊评价之后提出的最优指导性文件<sup>[3]</sup>。CPG 是公认的声明,是被系统地开发出来的,其内容要经过严格的评价,用于在特定的临床环境中帮助医务工作者对医疗活动进行决策。因此,CPG 具有权威性、宏观性、适用范围广等特点。

可靠的 CPG 必须做到:(1) 基于对现有证据的系统评价;(2) 由来自专业团队、各学科专家和主要相关团体的代表共同制定;(3) 适当的考虑重要患者亚群和患者偏好;(4) 过程透明,使干扰、偏倚和利益冲突最小化;(5) 对各备选干预措施及相应结局之间的关系提供合理的解释,并对证据质量和推荐意见进行分级;(6) 有重要的新证据时,要对原有 CPG 进行合理的重新审议和修订。

## 2 制定 CPG 的流程与方法

现阶段,制定高质量 CPG 是国内外实施正规化医疗、提高医疗质量和降低医疗投入十分有效的方法。CPG 一般分为专家共识指南和循证实践指南两大类<sup>[4]</sup>。

### 2.1 专家共识指南制定法 (consensus guideline development)

专家共识指南制定法分为非正式和正式的专家共识指南制定法。非正式的专家共识指南制定法较为简单,指南制定者组织一组专家就相关临床问题进行一次或多次开会讨论,经过会议讨论后将达成的共识形成推荐意见作为 CPG,由专业学会或政府机构进行发布。这种指南的制定缺乏相关的证据基础,推荐意见易受参会专家各方面因素的影响,专家推荐的干预方式在实际的临床应用中并不能确保真正对患者有利,因此,这种指南的质量一般较差,不具备真正意义的可靠性<sup>[5]</sup>。正式的专家共识指南制定法是针对某一干预方式进行相关研究证据的回顾,并列出可能的适应证,然后提供给制定指南的专家组成员,由专家组成员在第一次开会之前各自对每个适应证进行打分,评价其是否适用,量表为 9 分制,完全适用评 9 分,完全不适用评 1 分,其余情况在 2~8 分之间进行权衡。开会时,各个专家将自己的评分与专家组集体的评分进行比对,找出评分差异的原因,然后以会议上的讨论为基础,对先前自己的评分进行修改。以此方式,专家小组成员得出一致性程度较高的推荐意见<sup>[5]</sup>。正式的专家共识指南制定法虽然回顾了相关的研究证据,但在制定推荐意见时没有评价相关证据的质量,专家的主观意见仍占主导地位。

### 2.2 循证实践指南制定法 (evidence-based guideline development)

近年来,CPG 制定程序不断规范,越来越多的 CPG 都是在系统综述相关证据并对这些证据进行评价之后制定出来的。循证 CPG 现已成为制定

DOI:10.3877/cma.j.issn.1674-0807.2014.04.012

作者单位:400038 重庆,第三军医大学流行病学教研室 循证医学和临床流行病学中心

通信作者:李亚斐,Email:liyafei@tmmu.edu.cn

指南的趋势。循证 CPG 是基于相关证据制定的,并且严格评价了相关证据的质量。与以往的指南不同,循证 CPG 的制定流程包括:组成指南制定专家小组,提出指南拟解决的临床问题,系统全面检索文献,使用科学的方法对证据进行评价,综合考虑证据级别、证据强度和专家组成员的实践经验之后提出推荐意见。此外,为使 CPG 能与时俱进,CPG 推出后还应对其进行系统评估、推广普及和修订更新。下面将重点介绍循证 CPG 的制定方法。

苏格兰校际指南网络(Scottish Intercollegiate Guidelines Network, SIGN; <http://www.sign.ac.uk/>)创始于 1993 年,重点关注慢性非传染性疾病,如癌症、心血管疾病等。该网站下设有临床指南、指南选题、指南制定的方法学等栏目<sup>[6]</sup>。截至 2014 年 7 月 29 日,SIGN 共制定、发布了 140 项循证 CPG。SIGN 推荐的循证指南制定方法更具有代表性,其制定指南的主要步骤如下。

(1) 组建 CPG 制定专家小组 由来自不同地区的多学科专业人员组成,一般 15~20 人。制定指南需要有四项核心技能,即临床专业技能、卫生保健的实践经验、专业知识和严格的评估技能。

(2) 文献检索 CPG 制定小组确定指南拟解决的主要问题后,由专业的文献检索专家进行系统的文献检索,在 Cochrane Library、Medline、Embase、重要的专业学会或协会、指南出版机构的网站以及试验注册资料库和其他相关数据库反复进行检索。首先检索已有的 CPG 及系统评价,其次检索随机对照试验,最后根据提出的问题和获得证据的数量再检索其他类型的临床试验。

(3) 评价证据 CPG 制定小组明确规定文献的纳入标准和排除标准,并严格采用循证医学的评价标准对相关文献进行科学评价。

对证据质量的评价要点包括:①不同的研究提供的证据是否一致,研究对象的基线特征是否一致,研究内容是否一致;②研究结果是否与实际临床运用时的结果一致;③证据是否直接针对 CPG 的目标人群;④研究的样本量大小是否符合统计学要求。

对证据的解释,包括:①权衡利弊,并说明患者的结局指标是否最大限度地改善;②是否与现有的医疗实践有较大的差距;③是否会导致大规模的资源重新分配,卫生系统是否支持改进的

措施。

根据证据的质量对证据进行分级。每一篇文章至少应由两名 CPG 制定小组成员进行评价,如果存在分歧,则应交于第三方进行仲裁。

(4) 制定 CPG 初稿 根据对证据的客观评价结果提出推荐意见,制定出 CPG 初稿。经过严格的证据评价后达成共识,并参照证据水平和推荐意见强度对照表对推荐意见的强度进行标注。有充分证据时,根据证据提出推荐意见;没有证据或证据很弱时,根据讨论达成的共识性意见提出推荐意见。

(5) 咨询和同行评价 召开会议,向 CPG 制定小组提出疑问及对 CPG 初稿做出评价。CPG 制定小组根据建议进一步对 CPG 进行修订。然后送同行专家进行专业性评价。最后,SIGN 编辑组对指南进行审查并作出评价。

(6) 评估 CPG 发布 2 年后再进行评估。根据该领域进展现状决定是否对现有 CPG 进行更新。

(7) 患者参与 CPG 的制定不仅需要临床医师参与,也需要患者参与,应从临床医师、护理人员、患者等多个视角制定 CPG。

(8) 文件存档 在 CPG 制定过程中,需要重点保存以下文件:制定 CPG 的原始提议,制定 CPG 的理由和 CPG 涉及的范围,确定 CPG 的关键问题,检索策略、数据库和文献检索的时间范围,文献评价的纳入和排除标准,对支持建议的文献所用的方法学清单,回答所有关键问题的证据总结表,谨慎判断的表格,列表说明 CPG 制定小组对整体证据的质量和相关推荐分级的结论,总结性大会和同行评议的评论及回复记录。

(9) CPG 的执行 医务人员应根据自己所属的医疗环境以及患者的实际情况执行 CPG。

(10) 资料来源和其他因素 制定 CPG 会耗费大量的金钱和时间,因此,为取得预期效果,项目必须由遵循方法学的专家管理,并在规定的时间内完成。

循证 CPG 建立在科学证据的基础之上,可信度高,同时又标注了推荐意见的强度,便于使用者根据其强度决定是否采用该推荐意见。循证 CPG 一般都标注推荐意见级别和证据等级,但不同国家、不同机构采用的标准也不尽相同。为了建立评价证据和推荐意见分级的国际标准体系,2004

年评估、制定与评价推荐意见的等级 (grading of recommendations assessment, development and evaluation, GRADE) 工作组制定了 GRADE 证据质量和推荐分级<sup>[7]</sup>。GRADE 评级系统突破了单从研究设计角度考虑证据质量的局限性,它依据未来的研究结果是否改变临床医师对目前疗效评价的信心和改变可能性的大小将证据质量分为高、中、低、极低 4 个等级(表 1<sup>[7]</sup>),将推荐意见分为强、弱 2 个级别。当明确证实干预措施利大于弊或者弊大于利时,应评为强推荐;当不确定利弊关系或无论质量高低的证据均显示利弊相当时,则视为弱推荐。基于临床随机对照试验(randomized controlled trial, RCT)的证据在 GRADE 证据质量分级系统中被评为高质量。但是,并非所有 RCT 的质量都一致,下述 5 种情况将降低其证据质量,即研究的局限性、研究结果不一致、间接证据、结果不精确、报告有偏倚。GRADE 涵盖所有医学专业和临床护理领域的各种临床推荐意见,简单明了,易于掌握。Cochrane 协作网、世界卫生组织(WHO)等多个国际组织已经对其提供支持并广泛使用该评级系统。

表 1 GRADE 系统证据质量分级<sup>[7]</sup>

质量等级	定义
高	未来研究不太可能改变现有疗效评价结果的可信度
中	未来研究可能对现有疗效评估有重要影响,可能改变评价结果的可信度
低	未来研究很可能对现有疗效评估有重要影响,很可能改变评估结果的可信度
极低	任何疗效的评估都很不确定

GRADE:评估、制定与评价推荐意见的等级

在 GRADE 系统出现之前,应用较多的是 2001 年 5 月牛津循证医学中心制定的证据水平评价标准。其根据研究设计论证因果关系的强度不同,将证据水平分为 5 级<sup>[8]</sup>。根据证据质量、一致性、临床意义、普遍性等将推荐意见分为 A (优秀)、B (良好)、C (满意)和 D (差) 4 级。其中, A 级推荐意见应来自于 1 级水平的证据,所有研究结论一致,临床意义大,如果证据研究的样本人群与目标人群吻合,该推荐意见可直接应用于临床实践;而 B、C 级推荐意见则在上述各方面存在一定问题,其适用性受到不同限制;D 级推荐意见有待进一步评价,使用时要非常谨慎。表 2<sup>[8]</sup>仅列出治疗性研究证据的质量和推荐级别。

表 2 牛津循证医学中心临床证据水平分级和推荐级别(2001 年)<sup>[8]</sup>

推荐级别	证据水平	证据来源(治疗性研究证据)
A	1a	同质性 RCT 的系统评价
	1b	单个 RCT (置信区间窄)
	1c	“全或无”病案系列(未治疗前所有患者均死亡或部分死亡,治疗后仅部分死亡或全部存活)
B	2a	同质性队列研究的系统评价
	2b	单个队列研究(包括低质量 RCT,如随访率<80%)
	2c	结局性研究(指描述、解释和预测干预措施对临床最终结局的作用和影响的一类研究,最终结局不同于中间指标,主要包括生存、生命质量、疾病负担等)
	3a	同质性病例对照研究的系统评价
	3b	单个病例对照研究
C	4	病例系列研究(包括低质量队列和病例对照研究)
D	5	基于经验,未经严格论证的专家意见

RCT:随机对照研究

3 CPG 的评价

CPG 制定小组在制定 CPG 的过程中需合理控制可能存在的各种偏倚因素,才能制定出高质量的 CPG,确保 CPG 在临床应用中具有可行性。低质量的 CPG 不仅起不到临床指导作用,反而会误导临床医师采取不恰当的医疗手段。因此,对已发表的 CPG,读者应掌握如何评价其质量的高低,以便正确地选择应用。

3.1 评价 CPG 的基本原则

对 CPG 的评价主要分为真实性评价、重要性评价和适用性评价三个方面<sup>[9]</sup>。

首先是真实性评价。高质量的 CPG 应该遵循循证医学的原则和方法。强调 CPG 应以科学证据为基础,并根据证据的可靠程度对提出的建议进行分级。评价真实性的要点包括:(1) CPG 制定者的文献检索策略是否全面、可重复,文献检索是否在过去 1 年内进行;(2) 每项建议是否均标明了相关证据的等级,并准确地提供了原始证据来源。真实性评价主要包括证据的收集、评价和合成,以及如何将推荐意见与相关证据紧密结合等方面。

其次是重要性评价。评价了 CPG 的真实性后,还需要考虑该 CPG 针对的问题是否是临床工作中亟待解决的重要问题。但要注意,临床上经常遇到非常复杂的问题,任何 CPG 都不可能涵盖



所有的临床问题。

最后是适用性评价。适用性评价要点包括:本地区的疾病负担是否很低而无需参考 CPG,对患者治疗的效度评价是否与 CPG 中可比,执行该 CPG 所需的成本是否可接受,将 CPG 应用于患者时是否存在不可克服的困难,等等。

### 3.2 评价 CPG 的常用工具

为了科学客观地评价 CPG,不同的国家和学术团体制定了许多专门的 CPG 评价工具。迄今为止,大约有 20 多种评价工具。目前有 3 种评价工具比较全面、系统,分别是 Cluzeau 量表(37 个条目)、美国指南标准化会议评价标准(18 个条目)和欧洲临床指南研究与评价系统(appraisal of guidelines research and evaluation, AGREE)评价标准(23 个条目)。下面将主要介绍欧洲 AGREE 量表。AGREE 评价工具是欧洲制定的如何撰写和评价 CPG 的“指南”<sup>[10]</sup>,是由 13 个国家的研究者制定的一种 CPG 研究和评价的工具。AGREE 评价工具为 CPG 的质量审查提供了一个框架。AGREE 量表通过 6 个部分(CPG 的范围和目的、参与人员、CPG 开发的严格性、CPG 的清晰性与可读性、CPG 的适用性和 CPG 撰写的独立性)23 个条目对 CPG 进行评分(表 3)<sup>[10]</sup>。

表 3 中每个条目的分数为 1~4 分,完全符合条目要求的为 4 分,完全不符合的为 1 分,介于两者之间的,根据评估员的判断给 2 分或 3 分。由于评估员不可避免地存在着主观性,所以应由 2~4 名参评人员独立评价 CPG。然后,根据公式将所有参评人员的评分进行综合,得到标准化百分比。标准化百分比=(实际总分-最低可能分数)/(最高可能分数-最低可能分数)<sup>[11]</sup>(以表 3 第 1 部分为例计算得分的方法见表 4)。评价原则中 6 个部分的得分是完全独立的,不能被合计为一个总的质量得分。各部分的分数可以用于比较 CPG,以便帮助评估员决定是否使用或者推荐该 CPG。

AGREE 系统最后一部分为全面评估部分,它包含了一系列的选项,即“强烈推荐”、“推荐(补充或改进)”、“不推荐”和“不确定”。这种全面评估需要评估员考虑到每个评估标准,对该 CPG 的质量做出准确判断。

## 4 结语

随着现代医学的发展,尽管 CPG 已经逐渐成

表 3 AGREE 评价工具的 6 个部分 23 项具体条目<sup>[10]</sup>

I 指南的范围和目的(涉及 CPG 的总目的、具体的临床问题和适用的患者)	
(1)	明确说明撰写指南的目的
(2)	明确阐述了 CPG 所涵盖的临床问题
(3)	明确阐述了 CPG 要应用的目标人群,应提及年龄范围、性别、临床类型及共病
II 参与人员	
(4)	CPG 制定小组的成员包括所有相关方面的专家
(5)	CPG 考虑了目标患者的观点和选择
(6)	CPG 的适用者已明确
(7)	CPG 发表前在适用者中已被试行
III CPG 开发的严格性(涉及收集和综合证据的过程、制定和更新推荐建议的方法和步骤)	
(8)	采用严格系统的方法检索证据,应提供检索证据的详细策略,包括使用的检索词、信息来源和文献涵盖的时间
(9)	清楚描述了证据的纳入和排除标准
(10)	详细描述了形成推荐建议的方法
(11)	形成推荐建议时充分考虑了干预措施的获益、风险和和不良反应
(12)	推荐建议与支持证据之间有明确的联系
(13)	CPG 发表前接受了制定小组以外专家的同行评议
(14)	提供了更新 CPG 的程序
IV CPG 的清晰性和可读性	
(15)	推荐建议明确而不含糊
(16)	针对不同临床状况提供了不同的治疗选择
(17)	关键的推荐建议可以很容易找到
(18)	提供了 CPG 方便应用的工具(如概要、教育工具、针对患者的宣传单等)
V CPG 的适用性(CPG 应用时可能涉及的单位、操作和费用问题)	
(19)	讨论了应用该 CPG 的过程中可能遇到的困难
(20)	讨论了应用该 CPG 过程中可能的费用
(21)	提供监督和审查 CPG 执行情况的评估指标
VI CPG 编撰的独立性(推荐建议的独立性和对 CPG 制定小组中各成员利益冲突的说明)	
(22)	CPG 编辑工作独立于赞助单位
(23)	CPG 制定小组成员的利益冲突在 CPG 中有所说明

AGREE:欧洲临床指南研究与评价系统;CPG:临床实践指南

为临床医师科学决策的重要依据,但也应该认识到 CPG 存在一定的局限性。一方面,尽管国内外不断出现新制定的 CPG,但其质量仍然存在良莠不齐的现象,低质量的 CPG 将会误导临床实践;

表 4 AGREE 评价工具第 1 部分得分的计算方法

评估人员	条目 1	条目 2	条目 3	总分
评估员 1	2	3	3	8
评估员 2	3	3	4	10
评估员 3	2	4	3	9
评估员 4	2	3	4	9
总分	9	13	14	36

AGREE:欧洲临床指南研究与评价系统;最高可能分数=4 (完全符合分数)×3(项目数)×4(评估员数)=48;最低可能分数=1 (完全不符合分数)×3(项目数)×4(评估员数)=12;该部分的标准化百分比=(实际总分-最低可能分数)/(最高可能分数-最低可能分数)=(36-12)/(48-12)=0.67×100%=67%

另一方面,实际工作中的临床问题可能会相当复杂,任何 CPG 都不可能涵盖实际工作中的各种问题。国外循证 CPG 的制定方法为国内学者提供了重要的参考依据。国内学者可以以国外的相关循证 CPG 为基础,系统综述来自中国的证据,并对证据进行科学的评价,从而制定出符合自身需要的 CPG,以帮助临床医师和患者针对具体的临床问题做出恰当处理。

【关键词】 循证医学; 指南

【中图法分类号】 R-0 【文献标志码】 A

参 考 文 献

[1] Rosenfeld RM, Shiffman RN, Robertson P, et al. Clinical practice guideline development manual, third edition: a quality-driven approach for translating evidence into action [J]. Otolaryngol Head Neck Surg, 2013, 148(1 Suppl): S1-55.

[2] 柴凡, 姜军. 循证指南和决策支持服务在乳腺癌诊治过程中的作用 [J/CD]. 中华乳腺病杂志;电子版, 2007, 1(6): 248-249.

[3] Institute of Medicine (US) Committee on Standards for Developing Trustworthy Clinical Practice Guidelines, Graham R, Mancher M, et al. Clinical practice guidelines we can trust [M]. Washington, DC: National Academies Press (US), 2011.

[4] 陈薇, 刘建平. 循证临床实践指南的制订和评价 I. 循证临床实践指南编制的方法 [J]. 中华口腔医学杂志, 2013, 48(2): 109-111.

[5] 刘鸣. 临床实践指南意义、建立方法和评价 [J]. 中国卒中杂志, 2006, 1(1): 33-36.

[6] 王波, 詹思延. 国外循证临床实践指南制定的方法与经验 [J]. 中国循证心血管医学杂志, 2013, 5(4): 334-336.

[7] Atkins D, Eccles M, Flottorp S, et al. Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches The GRADE Working Group [J]. BMC Health Serv Res, 2004, 4(1): 38.

[8] 杨学宁, 吴一龙. 临床证据水平分级和推荐级别 [J]. 循证医学, 2003, 3(2): 111-113.

[9] 王家良. 循证医学 [M]. 2 版. 北京: 人民卫生出版社, 2010: 205.

[10] Brouwers MC, Kho ME, Browman GP, et al. AGREE II: advancing guideline development, reporting, and evaluation in health care [J]. Prev Med, 2010, 51(5): 421-424.

[11] 李卫国, 李宇宁, 张秀敏, 等. AGREE II 评价 9 个儿童青少年维生素 D 临床实践指南及推荐意见共识和差异 [J]. 中国循证儿科杂志, 2012, 7(5): 372-379.

(收稿日期:2014-07-31)

(本文编辑:罗承丽)

袁帅, 李亚斐. 临床实践指南的制定方法和评价原则 [J/CD]. 中华乳腺病杂志;电子版, 2014, 8(4): 283-287.