

# 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

叶志康<sup>1\*</sup>, 翟所迪<sup>2,3</sup> (1. 麦克马斯特大学 Michael G. DeGroote 国家疼痛中心, 加拿大 安大略省 汉密尔顿 L8S4L8; 2. 北京大学医学部药物评价中心, 北京 100191; 3. 北京大学第三医院药剂科, 北京 100191)

中图分类号 R9 文献标志码 A 文章编号 1672-2124(2024)07-0769-05

DOI 10.14009/j.issn.1672-2124.2024.07.001



**摘要** 目前,国际上存在多种证据与推荐意见分级系统,对临床医务工作者阅读指南证据和推荐意见造成困扰。证据推荐分级的评估、制定与评价(GRADE)系统目前是全球使用最为广泛的证据与推荐意见分级系统,本文简单地介绍 GRADE 系统证据质量评级及影响推荐意见形成的 4 个主要因素,并结合中国医院药学界发起制定的、在国际上发表的几部药学指南举例阐述,以帮助临床医务工作者了解 GRADE 系统,以批判性思维运用指南证据和推荐意见指导临床实践行为。

**关键词** 临床指南; 证据; 推荐意见; GRADE

## The Application of GRADE System in Clinical Guideline Development Regarding Hospital Pharmacy

YE Zhikang<sup>1</sup>, ZHAI Suodi<sup>2,3</sup> (1. Michael G. DeGroote National Pain Centre, McMaster University, Hamilton, Ontario L8S4L8, Canada; 2. Institute for Drug Evaluation, Peking University Health Science Center, Beijing 100191, China; 3. Dept. of Pharmacy, Peking University Third Hospital, Beijing 100191, China)

**ABSTRACT** There are many systems of rating quality of evidence and grading strength of recommendations, this may cause confusion for clinicians when reading guideline evidence and recommendations. Grading of Recommendations Assessment, Development, and Evaluation (GRADE) system is the most widely used rating system in the world. This article is to briefly introduce the evidence rating and 4 main factors of formulating recommendations in GRADE. We will use several published guidelines led by Chinese hospital pharmacy as examples to illustrate GRADE system, to help clinicians understand GRADE system and apply guideline evidence and recommendations into clinical practice with critical thinking.

**KEYWORDS** Clinical guideline; Evidence; Recommendations; GRADE

### 1 概述

2011 年,美国医学研究所(IOM)将临床实践指南定义更新为包含了一至多条用于完善优化患者监护的推荐意见的表述文本,而推荐意见的形成是基于系统评价全面总结的证据、不同干预措施的利弊平衡以及相应的卫生经济学评估<sup>[1-2]</sup>。基于 IOM 的定义,高质量临床指南应符合以下 6 条要求:(1) 基于总结当下所有证据的系统评价;(2) 由多学科临床专家和患者代表组成的指南专家组制定;(3) 合理考虑患者人群分类和患者意愿价值观;(4) 指南制定是基于清晰、具体且透明的过程,尽可能地减少指南制定过程中的偏倚和利益冲突;(5) 指南提供证据评级和推荐意见强度,且为推荐意见提供清晰且符合逻辑的解释;(6) 新的、重要的证据出现且有可能改变

现有推荐意见时,考虑更新指南。

临床实践指南制定的基础是证据收集合成与评价,全球目前使用最为广泛的证据与推荐意见分级系统是证据推荐分级的评估、制定与评价(grading of recommendations assessment, development, and evaluation, GRADE)系统<sup>[3]</sup>。GRADE 系统是由加拿大麦克马斯特大学发起、领导,联合全球各地的循证医学方法学家共同制定的分级系统。目前在中国已经成立 3 个 GRADE 分中心,越来越多的中国临床指南方法学家协助中国医师、药师和护士发起制定临床指南,并发表在高水平的国际期刊上。

2012 年,中国药理学学会治疗药物监测研究专业委员会发起制定了《万古霉素治疗药物监测指南》,该指南是在国内较早采用 GRADE 系统对证据和推荐意见强度进行分级的临床指南<sup>[4]</sup>。此后,更多的中国医院药学研究者发起制定了关于药

\* 副研究员。研究方向:临床实践指南制定、临床研究方法学。  
E-mail: yez39@mcmaster.ca

物治疗的临床实践指南,其中很多发表在国际期刊上,如2020年发表在英国医学杂志(British Medical Journal, BMJ)的《胃肠道出血预防用药快速临床指南》<sup>[5]</sup>,2020年发表在Clinical Infectious Diseases的《万古霉素治疗药物监测指南更新版》<sup>[6]</sup>。本文将结合这些中国医院药学专家发起的药物治疗指南,对GRADE证据与推荐意见分级系统进行阐述和探讨。

2 证据质量分级

GRADE系统将证据质量分为高、中、低和极低4个等级,随机对照试验(RCT)起始为高质量证据,观察性研究起始为低质量证据<sup>[7]</sup>。证据质量可以因为风险偏倚<sup>[8]</sup>、不一致性<sup>[9]</sup>、不精确性<sup>[10]</sup>、间接性<sup>[11]</sup>和发表偏倚<sup>[12]</sup>而降级。证据质量可以降1级、2级甚至3级,成为极低质量证据。

2.1 GRADE 评价证据整体

如前文所述,临床指南的推荐意见是基于系统评价总结的证据。系统评价可以针对某一临床问题系统地检索纳入所有符合条件的临床研究,全面总结证据。2021年更新发布的

系统评价报告规范(Prisma Checklist 2020)不仅要求评价单个纳入研究的质量和风险偏倚,还要求对整个系统评价的证据整体(body of evidence)进行质量评价<sup>[13]</sup>。

GRADE系统恰恰是对证据整体的质量进行评级,而不是对单个纳入研究的质量进行评价。GRADE系统基于每一个对患者重要的临床结局,把系统评价中关于该结局的所有纳入临床研究当成整体进行证据质量评级。

2.2 GRADE 证据质量概要表

GRADE证据概要表非常精炼地概括了一篇系统评价总结的证据,展示了临床干预措施的效应量大小、利弊和证据可信度,为临床医务工作者节省了时间和精力。GRADE证据概要表针对每一个重要临床结局进行证据汇总和分级,表格纵列依次为临床结局、对应的纳入研究数量和样本量、相对效应量、基线风险和绝对效应量差值、证据降级升级因素、证据等级以及对证据的简要文字描述。大多数指南的GRADE证据概要表标题见图1。

| Absolute effect estimates |                  |                                  |                     |                     | Plain language summary |
|---------------------------|------------------|----------------------------------|---------------------|---------------------|------------------------|
| Outcomes                  | Relative effects | Baseline risk for control group* | Difference (95% CI) | Quality of evidence |                        |

图1 GRADE 证据概要表标题

2.3 绝对效应量与相对效应量

系统评价通过Meta分析合并得出的效应量为相对效应量,如相对危险度(RR)和比值比(OR),但在临床实践指南制定和GRADE系统中,更加依赖于绝对效应量[绝对风险降低(ARR)]的大小。例如,与安慰剂相比,某药物可以显著降低心肌梗死再次发生率,相对效应量RR为0.75,也就是说,该治疗措施能将心肌梗死再次发生率降低25%。粗看起来,该药物对于预防再次心肌梗死的作用非常显著,但这只是相对效应量比较显著,需把相对效应量换算为绝对效应量,换算公式为 $ARR = (RR - 1) \times \text{心肌梗死基线发生率}$ 。心肌梗死基线发生率指安慰剂组或对照组的心肌梗死发生率,见表1。相对效应量RR均为0.75,当安慰剂组的心肌梗死再次发生率基线为1%时,即使能将心肌梗死再次发生率降低25%,也只是降低0.25%的心肌梗死再次发生率,0.25%的降低并没有显著的临床意义。当心肌梗死再次发生率基线为4%时,该药物可以降低1%的心肌梗死再次发生率,1%的降低幅度或许是比较有临床意义的。当心肌梗死再次发生率基线值为10%时,该药物可以降低2.5%的心肌梗死再次发生率,就比较大的临床意义。因此,在临床决策中,应该更注重治疗措施给患者带来的绝对效应量大小。

表1 相对效应量转化为绝对效应量

| 相对效应量 | 心肌梗死再次发生率基线/% | 绝对效应量/% |
|-------|---------------|---------|
| 0.75  | 1             | -0.25   |
| 0.75  | 2             | -0.50   |
| 0.75  | 4             | -1.00   |
| 0.75  | 10            | -2.50   |
| 0.75  | 20            | -5.00   |

2.4 风险偏倚

风险偏倚是GRADE系统评价证据整体质量的第一个维

度。每一篇系统评价都会采用风险偏倚量表评价每一个纳入研究的质量,如采用Cochrane Risk of Bias量表<sup>[14]</sup>或McMaster量表<sup>[15]</sup>评价每一个纳入RCT的质量,评价条目包括随机、分配隐藏、盲法、失访、选择性报告结局或其他风险偏倚等。对于队列研究和病例对照研究,可以采用纽卡斯尔-渥太华量表(Newcastle-Ottawa scale, NOS)<sup>[16]</sup>或McMaster大学量表<sup>[17-18]</sup>评价每一个纳入的队列研究或病例对照研究的风险偏倚质量。

一篇系统评价会纳入多个RCT或观察性研究,有的系统评价甚至可能会纳入超过100个临床研究。实际情况中,不是每一个纳入研究都是高质量研究,此时如果大多数的纳入研究没有严重的风险偏倚,则不会考虑因为风险偏倚而选择降级。但假如多数RCT在盲法这一项存在缺陷,则还是会因为风险偏倚而选择降级。

研究者往往也会对风险偏倚进行亚组分析(低偏倚风险与高偏倚风险),如果低偏倚风险和高偏倚风险亚组分析结果表明两组之间效应量存在明显不一致,则可能会选择低偏倚风险研究合并的结果作为更高质量的证据。例如,胃肠道出血预防用药BMJ快速临床指南及其系统评价<sup>[5,19]</sup>,对于临床显著胃肠道出血这个结局,基于风险偏倚进行亚组分析。质子泵抑制剂对比H<sub>2</sub>受体拮抗剂,低偏倚风险组合并效应量为0.58(0.29~1.17),而高偏倚风险组合并效应量为0.20(0.07~0.54),亚组分析差异P值为0.08,认为有明显的亚组差异,所以最终采用低偏倚风险亚组的效应量。

2.5 不一致性

不一致性考量的是纳入研究之间是否存在统计学和临床意义不一致性的问题。如果纳入的研究之间存在严重不一致性,会降低证据的可信度,此时需要将证据降级。不一致性分

为统计学不一致性和临床意义一致性,统计学异质性除了看  $I^2$  以外,还需要看纳入研究之间的点估计值和置信区间的重合程度。如果纳入研究之间的点估计值和置信区间存在严重不一致,则需要因为不一致性而降级。

图 2 为大血管闭塞性急性缺血性脑卒中患者单独机械取栓对比机械取栓联合阿替普酶的自主行动[改良 Rankin 量表(mRS)评分为 0~2 分]结局的森林图<sup>[20]</sup>。可以看到  $I^2$  为 0%, 6 个纳入 RCT 的置信区间基本相互重合包含。虽然有 2 个 RCT 的点估计值在 1 的右侧,其他 4 个 RCT 的点估计值在 1 的左侧,但总体来说 6 个 RCT 的一致性程度较高,所以没有因为不一致性而降级。

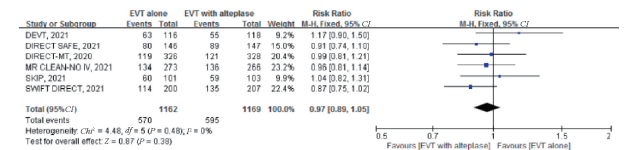


图 2 单独机械取栓对比机械取栓联合阿替普酶的 mRS 评分结局的森林图

除了考虑统计学不一致性外,还需要考虑是否存在临床意义的不一致性,包括人群、研究组和对照组的干预措施以及结局的不一致性。人群的不一致性是指各研究之间的人群年龄、性别、疾病严重程度、合并症等方面是否存在严重不一致;干预措施的不一致性包括剂量、给药途径和频次、手术方式和程度等方面是否存在不一致;结局方面的不一致性需要考虑各研究之间的结局测量方式、结局定义、结局随访时间等是否存在不一致。即使在统计学上没有异质性,但如果存在明显的临床意义不一致性,则建议进行亚组分析,增加结果的可信度。

对于研究之间不一致性的问题,除了对证据整体进行降级以外,建议对结果进行亚组分析并且判断亚组分析结果的可靠性。由加拿大麦克马斯特大学发起制定的效应差异可信度分析量表(ICEMAN)是判断亚组分析结果可靠性的工具<sup>[21]</sup>。

2.6 不精确性

不精确性简单来说主要考虑置信区间和样本量 2 个因素。在临床指南制定中,指南制定专家组会确定重要临床结局的最小临床重要差异(minimum clinically important difference, MCID)。有些治疗措施的疗效具有统计学差异,但可能这种效果太微小以至于对患者没有临床意义,如大多数患者不会认为死亡率绝对值降低 0.1% 有显著的意义。MCID 一般通过问卷形式得出具体数值,即患者个体认为的应该达到的最小临床疗效。例如,大血管闭塞性急性缺血性脑卒中机械取栓与阿替普酶临床指南的专家组成员以患者的角度完成问卷,获得脑卒中后死亡率、自主行动和症状性颅内出血的最小临床意义差值分别为 0.8%、1% 和 1%<sup>[22]</sup>;前蛋白转化酶枯草溶菌素 kexin9 型(PCSK9)抑制剂和依折麦布 BMJ 快速临床指南的专家组成员以患者的角度完成问卷,获得死亡率、非致命性脑卒中和非致命性心肌梗死的最小临床意义差值分别为 0.8%、1.0% 和 1.2%<sup>[23]</sup>。

设定 MCID 后,需考虑该结局绝对效应量的上下区间是否横跨 MCID。以大血管闭塞性急性缺血性脑卒中机械取栓与

阿替普酶临床指南为例,mRS 评分为 0~2 分结局的绝对效应量置信区间为-5.4%~2.5%,而其 MCID 为 1%,其置信区间横跨 MCID。下限为-5.4%,表示可以将 mRS 评分为 0~2 分结局降低 5.4%;上限为 2.5%,表示可以将 mRS 评分为 0~2 分结局增加 2.5%,表明该结局的精确性并不好,所以降 1 级。但考虑到该结局的置信区间较宽,比较不精确,所以将证据质量多降 1 级。因此,因为不精确性将证据质量总共降 2 级。而症状性颅内出血的绝对效应量置信区间为-1.80%~0.27%,其 MCID 为 1%,虽然也横跨 MCID,但是区间明显较窄,因此,因为不精确性将证据质量降 1 级。

2.7 间接性

间接性是指总结的临床证据应用于临床实践时是否存在间接性的问题。间接性的判断可以从人群、研究组和对照组的干预措施以及结局方面去判断。例如,成人的研究结果证据运用于儿童治疗中,就存在着人群的间接性;75 岁以上老年患者补钙治疗的研究证据运用于 50 岁以下成人患者中,也存在着间接性的问题,因为 75 岁以上老年患者补钙治疗的效应量比 50 岁以下成人患者的效应量更大。结局的间接性涉及中间指标这一概念,如血压、血糖和血脂是心血管事件的中间指标,骨密度是骨折发生的中间指标。一个降脂、降糖或降压药物想要取得上市许可,药品生产企业都会选择心血管事件作为 RCT 的主要结局,而不是选择血脂、血压或血糖降低这些中间指标作为 RCT 的主要结局。

2020 年 2 月,新型冠状病毒感染(COVID-19)疫情之初,加拿大麦克马斯特大学临床指南方法学家联合中国和其他一些国家的药学、医学、护理、患者代表、循证医学专家共同发起制定了 COVID-19 药物治疗指南,这是一部快速临床指南,主要针对糖皮质激素、抗病毒药物以及血浆治疗 3 个方面形成快速推荐意见,希望及时为临床医务工作者提供参考<sup>[24]</sup>。当时关于 COVID-19 药物治疗的 RCT 正处于招募受试者的状态,只有几篇样本量很小的观察性研究报告药物治疗效果,而普通大众对于 COVID-19 非常恐慌,盲目抢购服用一些药品,而且临床医师面对这一新型突发传染性疾病,迫切地希望有临床指南和证据指导临床治疗。面对缺乏 COVID-19 药物治疗的直接证据,指南专家组提出总结现有相关或类似疾病的证据作为间接证据来形成推荐意见。以糖皮质激素治疗 COVID-19 的系统评价和 Meta 分析为例<sup>[25]</sup>,因为 COVID-19 危重症患者的症状类似于急性呼吸窘迫综合征(acute respiratory distress syndrome, ARDS)的症状,所以,总结糖皮质激素治疗 ARDS 患者的 RCT 作为 COVID-19 危重症患者糖皮质激素治疗的间接证据。对于 COVID-19 重症患者,收集糖皮质激素治疗严重急性呼吸综合征(severe acute respiratory syndrome, SARS)、中东呼吸综合征(Middle East respiratory syndrome, MERS)、流行性感冒以及社区获得性肺炎的证据作为间接证据来形成 COVID-19 重症患者的推荐意见。对于药物有效性证据,因为 SARS 和 MERS 均为冠状病毒感染,与 COVID-19 更为接近,有效性证据因为间接性只降 1 级。而对于流行性感冒、医院获得性肺炎以及 ARDS,因为间接性程度较高,降 2 级。对于安全性证据,笔者认为安全性证据受间接性影响较少,因此这 5 类



疾病的安全性证据都只降 1 级。

2.8 发表偏倚

除了上述 4 个维度,最后考虑发表偏倚这个维度。期刊和研究者都倾向于发表阳性结果,而阴性结果因为各种原因发表的概率更小,这样就容易存在发表偏倚风险。一般采用漏斗图来检验是否存在发表偏倚。此外,还需要考虑纳入的临床研究是否由药品生产企业资助。最后,如果纳入研究大多数为小样本研究,且由药品生产企业资助,那么发表偏倚的风险也较大。

3 推荐意见强度分级

指南专家组审核确认证据概要表后,开始通过指南专家组全体会议讨论形成推荐意见。GRADE 系统将推荐意见分为强(强推荐和强不推荐)和弱(弱推荐和弱不推荐)两级。在 GRADE 系统中,推荐意见的形成应考虑 4 个主要因素,包括证据质量、利弊平衡、患者价值观和偏好、卫生经济学分析<sup>[2,26]</sup>。

3.1 证据质量

证据质量是确定推荐意见强度的关键因素,证据质量高低代表证据可信度的高低。证据质量越高,可信度越高,越容易形成强推荐或强不推荐<sup>[26]</sup>。但这并不意味着高质量证据一定会形成强推荐,如高质量证据表明某种新上市治疗药物的疗效和安全性显著优于现有治疗药物,但是新药价格很高,超出本国经济及医保承受范围,那么指南专家组可能只是弱推荐使用该新药。同样,低质量的证据一般不建议形成强推荐,但在某些特殊情况也有可以形成强推荐。

3.2 利弊平衡

任何治疗或诊断措施对患者都有收益和弊端,当该治疗或诊断措施给患者带来的收益明显多于弊端时,推荐意见强度更有可能成为强推荐;当该治疗或诊断措施给患者带来的收益与弊端相当时,推荐意见强度更有可能成为弱推荐<sup>[2]</sup>。例如,阿司匹林可以显著降低既往发生心血管事件患者的死亡率和再发心肌梗死,且阿司匹林引起的不良反应比较少,药品价格也较低,这种情况下,指南专家组往往会形成推荐。但很多时候收益和弊端并不容易权衡,以大血管闭塞性急性缺血性脑卒中患者机械取栓与阿替普酶快速临床指南为例,与机械取栓联合静脉用阿替普酶组相比,机械取栓单独治疗可以降低 1.0% 的症状性颅内出血(中等质量证据),但可能会增加 1.2% 的死亡率和降低 1.5% 的自主行动比例(低质量证据)<sup>[22]</sup>。指南专家组更信任中等质量的证据表明机械取栓单独治疗会降低症状性颅内出血,虽然机械取栓单独治疗可能会增加死亡率和降低自主行动比例,但因为是低质量证据,可信度较低,相对没有那么信任。权衡利弊,以及结合卫生经济学分析,指南专家组最后决定弱推荐机械取栓单独治疗。

3.3 患者价值观和偏好

患者价值观和偏好是推荐意见形成的另一个主要因素。患者价值观和偏好定义比较广泛,包括患者的角度、治疗理念、治疗预期和生命健康目标,也包括患者权衡不同治疗或诊断措施所带来的收益、弊端、花费负担的过程<sup>[27]</sup>。

指南专家组获得患者价值观和偏好可以依次考虑以下 4 种方式<sup>[2]</sup>。(1)检索已发表的关于该指南治疗领域患者价值观和偏好的系统评价。(2)如果没有已发表的系统评价,指南

专家组可以开展制定关于该指南治疗领域患者价值观和偏好的系统评价,前提是有已发表的患者价值观和偏好的临床研究,钠-葡萄糖协同转运蛋白 2(SGLT-2)抑制剂或胰高血糖素样肽-1(GLP-1)受体激动剂临床指南就是采用该方法<sup>[28-29]</sup>。(3)如果没有系统评价和临床研究,指南专家组可以设计问卷调查该领域患者价值观和偏好,但这种方法较费时。(4)因此,指南专家组可以采用第 4 种方法,即每名指南专家组成员基于自己的医疗经验代替患者,以患者的角度填写一份问卷,统计分析大概 20 份问卷结果,进而获得患者价值观和偏好。例如,胃肠道出血预防用药指南采用该方法获得的最小临床有效阈值为 1%<sup>[5]</sup>。

患者意愿和价值观可能会出现分歧较大的情况,即使形成了推荐意见,也难以将这条推荐意见适用于绝大部分患者。因此,在患者意愿和价值观不一致的情况下,更容易形成弱推荐或不形成推荐意见。

3.4 卫生经济学分析

指南推荐意见的形成是否考虑医疗卫生资源和开展卫生经济学分析,由指南制定出发的角度所决定。如果是从患者角度制定全球通用的临床实践指南,则不考虑开展卫生经济学分析,只考虑干预措施对患者的收益和弊端。而且,各个国家经济情况不一样,无法对每个国家开展卫生经济学分析。

对于发展中国家,由于医疗资源整体有限,更有必要邀请卫生经济学家加入指南制定,开展卫生经济学分析,为推荐意见形成提供更多参考。以大血管闭塞性急性缺血性脑卒中机械取栓与阿替普酶治疗的快速临床指南为例,与机械取栓联合静脉用阿替普酶相比,机械取栓单独治疗可以降低症状性颅内出血(中等质量证据),但可能会增加死亡率和降低自主行动比例(低质量证据)<sup>[22]</sup>。该指南同时在加拿大和中国开展经济学模型分析比较机械取栓单独治疗组与机械取栓联合阿替普酶组,模型结果分析不认为机械取栓联合静脉用阿替普酶组更具有经济学效益<sup>[30]</sup>。

综上所述,GRADE 系统因为其科学性、透明性、严谨性和全面性成为全球最广为接受的证据与推荐意见分级系统,从风险偏倚、不一致性、不精确性、间接性和发表偏倚 5 个维度全面考虑证据整体可能存在的问题和维度,批判性地评价解读证据。GRADE 系统也全面考虑到形成推荐意见的 4 个主要因素,即证据质量、利弊平衡、患者价值观和偏好以及卫生经济学分析。

本文简单地概述了 GRADE 证据与推荐意见分级系统,并结合中国医院药学界发起制定的药学指南举例阐述说明,希望临床医务工作者能对 GRADE 系统有一个基本全面的了解,提升临床医务工作者解读证据和指南推荐意见的能力,有助于临床医务工作者以批判性的思维运用证据和推荐意见指导临床实践行为。

参考文献

[1] Institute of Medicine, Board on Health Care Services, Committee on Standards for Developing Trustworthy Clinical Practice Guidelines. Clinical practice guidelines we can trust[EB/OL]. [2024-04-20]. <https://nap. nationalacademies. org/initiative/committee-on-standards-for-developing-trustworthy-clinical-practice-guidelines>.

- [2] 叶志康, 杨辉. 临床实践指南的制定、存在问题与运用: 由表及里, 临床医生如何读懂临床指南的内在[J]. 中国全科医学, 2023, 26(13): 1559-1567.
- [3] 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.
- [4] YE Z K, CHEN Y L, CHEN K, et al. Therapeutic drug monitoring of vancomycin: a guideline of the division of therapeutic drug monitoring, Chinese Pharmacological Society [J]. J Antimicrob Chemother, 2016, 71(11): 3020-3025.
- [5] YE Z K, REINTAM BLASER A, LYTVYN L, et al. Gastrointestinal bleeding prophylaxis for critically ill patients: a clinical practice guideline[J]. BMJ, 2020, 368: l6722.
- [6] HE N, SU S, YE Z K, et al. Evidence-based guideline for therapeutic drug monitoring of vancomycin: 2020 update by the division of therapeutic drug monitoring, Chinese Pharmacological Society[J]. Clin Infect Dis, 2020, 71(Suppl 4): S363-S371.
- [7] BALSHEM H, HELFAND M, SCHÜNEMANN H J, et al. GRADE guidelines: 3. Rating the quality of evidence [J]. J Clin Epidemiol, 2011, 64(4): 401-406.
- [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] GUYATT G H, OXMAN A D, KUNZ R, et al. GRADE guidelines: 7. Rating the quality of evidence—inconsistency[J]. J Clin Epidemiol, 2011, 64(12): 1294-1302.
- [10] GUYATT G H, OXMAN A D, KUNZ R, et al. GRADE guidelines 6. Rating the quality of evidence—imprecision [J]. J Clin Epidemiol, 2011, 64(12): 1283-1293.
- [11] GUYATT G H, OXMAN A D, KUNZ R, et al. GRADE guidelines: 8. Rating the quality of evidence—indirectness[J]. J Clin Epidemiol, 2011, 64(12): 1303-1310.
- [12] GUYATT G H, OXMAN A D, MONTORI V, et al. GRADE guidelines: 5. Rating the quality of evidence—publication bias[J]. J Clin Epidemiol, 2011, 64(12): 1277-1282.
- [13] PAGE M J, MCKENZIE J E, BOSSUYT P M, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews[J]. BMJ, 2021, 372: n71.
- [14] STERNE J A C, SAVOVI J, PAGE M J, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials[J]. BMJ, 2019, 366: l4898.
- [15] McMaster University. Tool to assess risk of bias in randomized controlled trials [EB/OL]. [2024-04-24]. <https://www.distillersr.com/resources/methodological-resources/tool-to-assess-risk-of-bias-in-randomized-controlled-trials-distillersr>.
- [16] WELLS G A, SHEA B, O'Connell D, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses[EB/OL]. [2024-04-24]. [https://www.ohri.ca/programs/clinical\\_epidemiology/oxford.asp](https://www.ohri.ca/programs/clinical_epidemiology/oxford.asp).
- [17] McMaster University. Tool to assess risk of bias in cohort studies [EB/OL]. [2024-04-25]. <https://www.distillersr.com/resources/methodological-resources/tool-to-assess-risk-of-bias-in-cohort-studies-distillersr>.
- [18] McMaster University. Tool to assess risk of bias in case control studies [EB/OL]. [2024-04-26]. <https://www.distillersr.com/resources/methodological-resources/tool-to-assess-risk-of-bias-in-case-control-studies-distillersr>.
- [19] WANG Y, YE Z K, GE L, et al. Efficacy and safety of gastrointestinal bleeding prophylaxis in critically ill patients: systematic review and network meta-analysis[J]. BMJ, 2020, 368: l6744.
- [20] WANG X, YE Z K, BUSSE J W, et al. Endovascular thrombectomy with or without intravenous alteplase for acute ischemic stroke due to large vessel occlusion: a systematic review and meta-analysis of randomized trials [J]. Stroke Vasc Neurol, 2022, 7(6): 510-517.
- [21] SCHANDELMAIER S, BRIEL M, VARADHAN R, et al. Development of the instrument to assess the credibility of effect modification analyses (ICEMAN) in randomized controlled trials and meta-analyses[J]. CMAJ, 2020, 192(32): E901-E906.
- [22] YE Z K, BUSSE J W, HILL M D, et al. Endovascular thrombectomy and intravenous alteplase in patients with acute ischemic stroke due to large vessel occlusion: a clinical practice guideline[J]. J Evid Based Med, 2022, 15(3): 263-271.
- [23] HAO Q K, AERTGEERTS B, GUYATT G, et al. PCSK9 inhibitors and ezetimibe for the reduction of cardiovascular events: a clinical practice guideline with risk-stratified recommendations[J]. BMJ, 2022, 377: e069066.
- [24] YE Z K, ROCHWERG B, WANG Y, et al. Treatment of patients with nonsevere and severe coronavirus disease 2019: an evidence-based guideline[J]. CMAJ, 2020, 192(20): E536-E545.
- [25] YE Z K, WANG Y, COLUNGA-LOZANO L E, et al. Efficacy and safety of corticosteroids in COVID-19 based on evidence for COVID-19, other coronavirus infections, influenza, community-acquired pneumonia and acute respiratory distress syndrome: a systematic review and meta-analysis[J]. CMAJ, 2020, 192(27): E756-E767.
- [26] 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.
- [27] MONTORI V, DEVEREAUX P, STRAUS S, et al. Decision making and the patient [M]//GUYATT G, RENNIE D, MEADE M, et al. Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice. New York: McGraw-Hill Education, 2008:1-20.
- [28] LI S Y, VANDVIK P O, LYTVYN L, et al. SGLT-2 inhibitors or GLP-1 receptor agonists for adults with type 2 diabetes: a clinical practice guideline[J]. BMJ, 2021, 373: n1091.
- [29] GONZÁLEZ-GONZÁLEZ J G, DÍAZ GONZÁLEZ-COLMENERO A, MILLÁN-ALANÍS J M, et al. Values, preferences and burden of treatment for the initiation of GLP-1 receptor agonists and SGLT-2 inhibitors in adult patients with type 2 diabetes: a systematic review [J]. BMJ Open, 2021, 11(7): e049130.
- [30] YE Z K, ZHOU T, ZHANG M M, et al. Cost-effectiveness of endovascular thrombectomy with alteplase versus endovascular thrombectomy alone for acute ischemic stroke secondary to large vessel occlusion[J]. CMAJ Open, 2023, 11(3): E443-E450.

(收稿日期:2024-05-15 修回日期:2024-06-05)