

公共卫生决策证据质量的 Meta 流行病学研究

黄家艺^{1,2,3*} 邓欣欣^{1,2,3} 王汉彬^{1,2,3} 胡晓晔^{1,2,3} 梁 翠^{1,2,3} 崔璐^{1,2,3} 杨克虎^{1,2,3,4} 李秀霞^{1,3,4}

1. 兰州大学卫生技术评估中心/兰州大学循证社会科学研究中心 兰州大学公共卫生学院 甘肃兰州 730000
2. 兰州大学循证医学中心 兰州大学基础医学院 甘肃兰州 730000
3. 甘肃省循证医学重点实验室 甘肃兰州 730000
4. 兰州大学医疗器械监管研究中心 甘肃兰州 730000

【摘要】目的:比较公共卫生决策证据质量评价体系(PHE-Grading)与推荐分级的评价、制定与评估体系(GRADE)在评价公共卫生决策证据质量方面的差异性。方法:计算机检索 Cochrane Library 数据库,获取建库起至2024年2月27日发表的有关“Public health”主题的系统评价。采用 EndNote 20 软件进行文献筛选,Excel 2021 和 SPSS 22.0 软件进行数据整理和分析,使用 RevMan 5.4.1 软件绘制森林图。结果:最终纳入 61 篇系统评价进行证据质量评价。GRADE 和 PHE-Grading 证据分级结果的森林图显示,证据等级为高[OR:2.39,95%CI(1.21,4.75)]、中[OR:0.40,95%CI(0.31,0.52)]、低[OR:0.37,95%CI(0.29,0.46)]和极低[OR:85.11,95%CI(34.80,208.11)],证据质量分级结果的差异有统计学意义。结论:与 GRADE 相比,PHE-Grading 对公共卫生决策证据确信度的把握可能更为精准,目前公共卫生决策证据质量仍集中在中低水平,仍需加强高质量研究以支持科学决策。

【关键词】公共卫生;决策;证据质量;GRADE 方法;PHE-Grading
中图分类号:R197 文献标识码:A doi:10.3969/j.issn.1674-2982.2024.10.011

The evidence quality of public health decision-making: A meta-epidemiological study

HUANG Jia-yi^{1,2,3}, DENG Xin-xin^{1,2,3}, WANG Han-bin^{1,2,3}, HU Xiao-ye^{1,2,3}, LIANG Cui^{1,2,3}, CUI Lu^{1,2,3}, YANG Ke-hu^{1,2,3,4}, LI Xiu-xia^{1,3,4}

1. Health Technology Assessment Center/Evidence Based Social Science Research Center, School of Public Health, Lanzhou University, Lanzhou Gansu 730000, China
2. Evidence Based Medicine Center, School of Basic Medical Sciences, Lanzhou University, Lanzhou Gansu 730000, China
3. Key Laboratory of Evidence Based Medicine of Gansu Province, Lanzhou Gansu 730000, China
4. Research Center for Medical Device Regulatory Science, Lanzhou University, Lanzhou Gansu 730000, China

【Abstract】 Objective: To compare the difference between the Evidence Quality Grading System for Public Health Decision-making (PHE-Grading) and the Grading of Recommendations Assessment, Development and Evaluation (GRADE) System in evaluating the quality of evidence for public health decision-making. Methods: Systematic reviews about topic “Public health” were electronically searched in the Cochrane Library database from inception to February 27, 2024. EndNote 20 software was used for literature screening, Excel 2021 and SPSS 22.0 software were used for data collation and analysis, and the forest plot was drawn by RevMan 5.4.1 software. Results: A total of 61 systematic reviews were finally included for evidence quality evaluation. The forest plot of GRADE and PHE-Grading evidence grading results showed that high grade [OR:2.39, 95%CI (1.21 to 4.75)], moderate grade [OR:0.40, 95%CI (0.31 to 0.52)], low grade [OR:0.37, 95%CI (0.29 to 0.46)], and extremely low grade [OR:85.11, 95%CI (34.80 to 208.11)], and the differences in evidence quality grading results between the two systems were statistically significant. Conclusions: Compared with GRADE, PHE-Grading may be more accurate in

* 基金项目:国家自然科学基金项目(72074103)
作者简介:黄家艺(2001 年—),女,硕士研究生,主要研究方向为循证卫生决策与管理。E-mail:2623460875@qq.com
通讯作者:李秀霞。E-mail:lixiaxia@lzu.edu.cn

grasping the certainty of public health decision-making evidence. Currently, the quality of public health decision-making evidence is still concentrated in low and middle level, and high-quality research still needs to be strengthened to support scientific decision-making.

[Key words] Public health; Decision making; Evidence quality; GRADE approach; PHE-Grading

1 背景

公共卫生决策旨在促进和保护群体、社区和人口的健康和福祉^[1],其决策的科学性非常关键。随着国内外医学环境的不断发展与进步以及医疗资源紧缺压力的不断增加,世界各国的卫生决策越来越多地依赖于科学证据和实证研究,同时也对公共卫生决策的科学性提出了更高层次的要求。^[2-3]循证公共卫生决策是采用循证医学的思想和理念处理和解决人群的公共卫生问题,证据质量分级是循证决策的精髓。^[4-5]近年来,卫生政策研究者和制定者对于研究质量的关注日益增加,但研究证据的评估部分尚需进一步细化和完善。^[6]2023 年,商雪等构建了公共卫生决策证据质量评价体系(Evidence Quality Grading System for Public Health Decision-making, PHE-Grading),证据质量评价采用评分制而非原有的升降级制,弥补了证据质量和推荐意见评级系统(Grades of Recommendations Assessment, Development and Evaluation, GRADE)缺乏量化考虑的不足。^[7]然而,新体系的建立需要经过严格的检验和验证,以确保其在实践中的有效性和可靠性,但此前尚未有研究系统的检验 PHE-Grading 在公共卫生决策领域内对证据确信度把握的精确性。Meta 流行病学是将流行病学和 Meta 分析结合起来的一门交叉学科,可用于研究方法学质量评价或探究各种研究特征的报告情况。^[8]

实现健康中国建设 2030 年的宏伟目标以及强化和完善公共卫生体系,亟需大量高质量的证据支撑。^[9]通过研究循证公共卫生决策,能够不断丰富和完善公共政策制定的理论与方法体系。^[2] Cochrane Library 数据库的系统评价因其方法严谨和连续更新的特点,被认为是世界上最高级别的证据之一,已成为许多发达国家公共卫生决策的参考依据。^[10-11]因此,本研究聚焦证据质量板块,以 Cochrane Library 数据库为例,运用 Meta 流行病学的方法,比较并分析 PHE-Grading 与 GRADE 在评价公共卫生决策证据质量方面的差异性和精准性,通过不断完善和优化证据质量评价体系,从而推动公共卫生政策创新,以适应公共卫生领域的新需求和新挑战。

2 资料与方法

2.1 资料来源与数据提取

计算机检索 Cochrane Library 数据库,检索时间为建库起至 2024 年 2 月 27 日,不限制发表语言,以检索词“Public health”为主题进行文献检索,选取“Cochrane Reviews”部分根据文献纳入、排除标准筛选相关文献。

本研究仅纳入关注公共卫生研究问题并且已对其结局指标进行了 GRADE 评级的系统评价,其中公共卫生具体包括对重大疾病,尤其是传染病(如结核、艾滋病、SARS、新冠肺炎等)的预防、监控和治疗,对食品、药品、公共环境卫生的监督管理,以及相关的卫生宣传、健康教育、免疫接种等方面。^[12]同时排除计划书、已撤稿的文献、系统评价再评价、GRADE 评级信息不全以及方法学研究的文献。

检索到的 117 篇文献随后被导入 EndNote 20 软件进行自动去重处理,之后由研究者根据纳入、排除标准初筛题目和摘要,排除明显不符合的文献后复筛全文,以确定最终纳入 61 篇文献用于证据质量评价。数据提取部分,研究利用 Excel 2021 软件预设表格,提取纳入文献的结局指标、GRADE 评级结果及证据升降级原因。文献筛选和数据提取全过程均由两名研究者独立进行,交叉核对,分歧由第三人裁定,以保证研究质量。文献筛选流程见图 1。

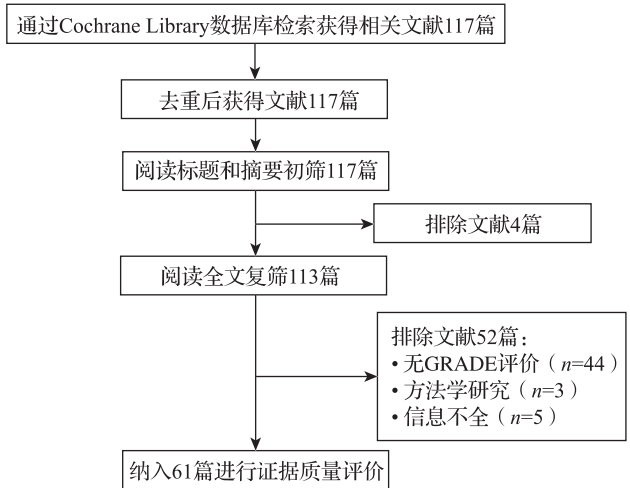


图 1 文献筛选流程图

2.2 证据质量评价

PHE-Grading 是对公共卫生决策进行证据质量分级的特色理论体系,由 15 个一级条目和 55 个二级条目组成,采用评分制,15 个一级条目满分共 15 分,根据得分最终转化为 4 个级别的证据强度,各等级得分区间为:高(>11 分)、中(8~11 分)、低(4~8 分)和极低(≤4 分)。[7]本研究按照公共卫生决策证据质量评价体系(PHE-Grading)的评价标准,对纳入文献,针对其结局指标进行证据质量评价,将评价结果记录到预设的 Excel 2021 表格里。证据质量评价的全过程由两名研究者独立进行,交叉核对,分歧由第三人裁定,以保证研究质量。

2.3 统计分析

采用 Excel 2021 软件整理数据,使用 SPSS 22.0 软件进行 Shapiro-Wilk 检验以探究 PHE-Grading 证据质量评分的正态性, $P<0.05$ 为偏态分布。采用 RevMan 5.4.1 软件绘制森林图比较 GRADE 和 PHE-Grading 系统证据质量评价结果的差异性,优势比(Odds Ratio, OR)的 95% 置信区间(Confidence Interval, CI)与无效线相交为差异无统计学意义,反之则为有。结果以文字、百分数结合图表直观展示。

3 结果

3.1 纳入研究基本特征

最终纳入 61 篇文献用于证据质量评价,其发表年份在 2011—2024 年之间,其中有 46 篇(75.41%)系统评价进行了 Meta 分析,15 篇(24.59%)系统评价采用叙事综合的方法进行证据合成,系统评价的第一作者来自英国(14 篇,22.95%)、澳大利亚(9 篇,14.75%)、加拿大(5 篇,8.20%)等 18 个国家,包括营养与食品健康(19 篇,31.15%)、职业病与职业健康(7 篇,11.48%)、预防伤害(6 篇,9.84%)等 19 个主题。

3.2 GRADE 证据质量评价结果

本研究共有 635 个结局指标,其中 480 个(75.59%)评价起始于高级,155 个(24.41%)评价起始于低级,456 个(71.81%)因偏倚风险而降级,144 个(22.68%)因不一致性而降级,303 个(47.72%)因不精确性而降级,105 个(16.54%)因间接性而降级,14 个(2.20%)因发表偏倚而降级,13 个(2.05%)因大效应量而升级,2 个(0.31%)因剂量—反应关系而升级,没有结局指标因负偏倚而

升级。

最终,256 个(40.31%)结局指标被评为极低,234 个(36.85%)结局指标被评为低级,117 个(18.43%)结局指标被评为中级,28 个(4.41%)结局指标被评为高级(表 1)。

表 1 证据质量评价结果

GRADE 评价结果	PHE-Grading 评价结果				合计
	极低	低	中	高	
极低	5	209	42	0	256
低	0	138	96	0	234
中	0	42	74	1	117
高	0	0	17	11	28
合计	5	389	229	12	635

3.3 PHE-Grading 证据质量评价结果

应用 PHE-Grading 对 635 个结局指标进行证据质量评价,Shapiro-Wilk 检验结果显示,证据质量评分为偏态分布($P<0.05$),中位数为 8.00 分,四分位数间距为 1.50 分,本次评价中 15 个一级条目总得分范围为 4~12。结局指标在研究设计、直接性、稳健性、因果推断重要性、资助偏倚和可推广性 6 个一级条目上被评为“1 分”的百分比高于 50%,在研究执行质量、精确性、大效应量和负偏倚 4 个一级条目上被评为“0 分”的百分比高于 50%,值得注意的是,没有结局指标存在负偏倚。此外结局指标在发表偏倚、剂量—反应关系以及证据的阻碍和促进因素 3 个一级条目上的得分也较低,它们被评为“0.5 分”的百分比分别为 95.91%、98.90% 和 70.55%(图 2)。

最终,389 个(61.26%)结局指标被评为低级,229 个(36.06%)结局指标被评为中级,12 个(1.89%)结局指标被评为高级,5 个(0.79%)结局指标被评为极低(表 1)。

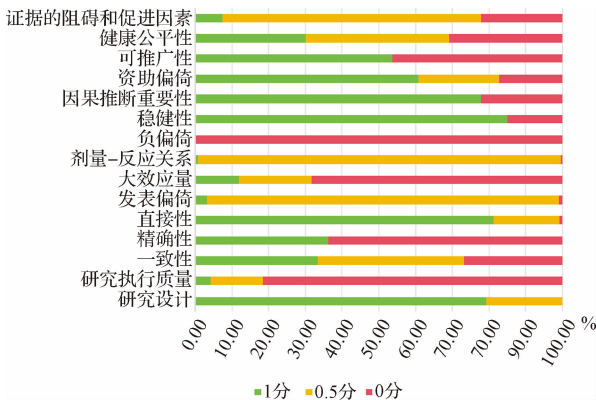


图 2 PHE-Grading 证据质量评价结果

3.4 证据质量评级的 Meta 比较

GRADE 系统与 PHE-Grading 两种证据质量评价系统均得出公共卫生决策系统评价主要在研究设计上容易升级/得分,在研究执行质量,即偏倚风险和精确性上容易降级/失分。

Meta 比较森林图显示,各证据级别 OR 值的 95% CI 横线与无效线不相交,说明应用两种系统进行证据质量分级结果的差异有统计学意义。GRADE 系统分级结果在高[OR:2.39,95% CI(1.21,4.75)]和极低[OR:85.11,95% CI(34.80,208.11)]两个证据级别上的数量多于 PHE-Grading 分级结果,在中[OR:0.40,95% CI(0.31,0.52)]和低[OR:0.37,95% CI(0.29,0.46)]两个证据级别上的数量少于 PHE-Grading 分级结果。应用 GRADE 系统进行证据

质量评价的结果多集中在极低级和低级,共 490 个(77.17%),应用 PHE-Grading 进行证据质量评价的结果多集中在低级和中级,共 618 个(97.32%),两种系统进行证据质量评价得出的高质量证据占比均较少(图 3)。

与应用 GRADE 系统进行证据质量评价相比,应用 PHE-Grading 评价后有 348 个(54.80%)结局指标的证据等级得到了升高,其中 209 个由极低级升高到了低级,42 个由极低级升高到了中级,96 个由低级升高到了中级,1 个由中级升高到了高级。59 个(9.29%)结局指标的证据等级降低了,其中有 42 个由中级降到了低级,17 个由高级降到了中级。此外,还有 228 个(35.91%)结局指标的证据级别没有发生变化(表 1)。

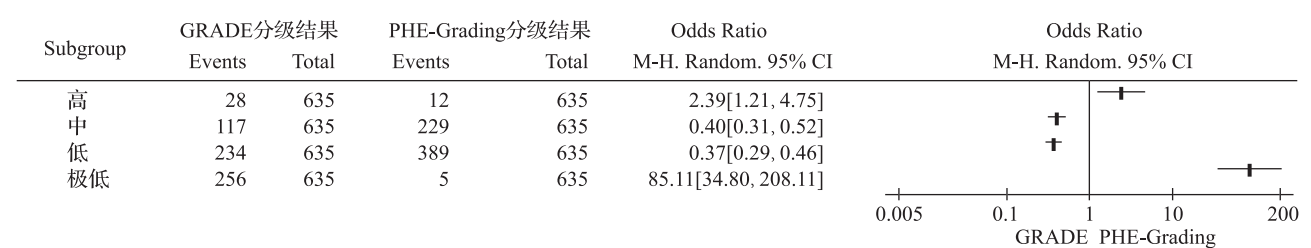


图 3 证据质量评级的 Meta 比较森林图

4 讨论

4.1 GRADE 与 PHE-Grading 证据质量评级的差异性分析

GRADE 系统对证据质量的评估框架严谨,它对证据质量的判断始于研究设计,如随机对照试验通常获得最高初始评级,观察性研究则始于低级别评级。^[13]随后通过细致考量偏倚风险、不一致性、不精确性、间接性和发表偏倚五个因素而降级,同时也将大效应量、存在剂量—反应关系和负偏倚三个因素作为升级依据。^[14]Meta 比较的结果反映出应用两种系统进行证据质量分级结果的差异有统计学意义,GRADE 系统在高和极低证据级别上的分级结果多于 PHE-Grading,而在中和低证据级别上则少于后者。从证据质量评价结果可以看出,应用 GRADE 评价多起始于高级,多因研究偏倚风险、不精确性、不一致性和间接性而降级,升级情况相对罕见,根据研究设计和证据质量评价的“五降三升”因素对结局指标进行评级,证据级别多集中在低和极低水平,中高水平的证据相对较少,尤其缺乏高级别证据。

而 PHE-Grading 系统在设计时充分考虑了公共卫生决策的特定需求,其评价在一级条目“稳健性、因果推断重要性、资助偏倚、可推广性”上得分均较多,在“健康公平性、证据的阻碍和促进因素”两个一级条目上得分适中,凸显出这些维度对于公共卫生决策尤为重要。通过这一多维度的评估框架,PHE-Grading 能够更全面地捕捉影响证据在公共卫生实践中应用效果的关键因素。除此之外,经 PHE-Grading 证据质量评级,过半的证据在评级后得到了升级,且证据级别多集中在中低水平,反映出 PHE-Grading 系统在实际应用中显示出对证据级别的提升作用更为显著。这一现象表明,PHE-Grading 可能更适用于在复杂多变的公共卫生环境中,识别并赋予研究证据以更精准的确信度,从而有助于决策者做出更为精准和有效的决策。

尽管两个系统均致力于提升公共卫生决策的科学性,但 PHE-Grading 通过其独特的评估框架和更加贴近公共卫生实践需求的考量条目,可能在把握证据确信度方面展现出更高的精准性。同时高质量证据占比均较少,提示未来仍需加强高质量研究以支持科学决策。

4.2 PHE-Grading 证据质量评级的精准性分析

基于上述分析,PHE-Grading 在评估公共卫生决策证据的确信度方面,相较于普遍适用的 GRADE 系统,可能展现出更为精准的特点。这主要得益于 PHE-Grading 在设计时对评价形式的思维转变和对公共卫生领域特定需求的充分考虑。一方面,PHE-Grading 中有 9 个一级条目与 GRADE 系统相同,但因其整体采用量化评分制而非升降级制且具体评价标准有所不同^[7],在前 9 个一级条目上的评价也有所不相同。如使用 GRADE 系统对来自观察性研究的公共卫生证据进行质量评级有困难,难以区分所有类型的非随机研究,这可能会导致在向政策制定者传达信息时存在对证据的误解^[15],而 PHE-Grading 在“研究设计”这一条目则考虑了多种研究设计,对应不同的分值,使得评价过程更加清晰明了,此外,其在“一致性”条目中也充分考虑到了单个研究和定性研究等。另一方面,PHE-Grading 新增健康公平性、资助偏倚、可推广性、稳健性、因果推断重要性以及证据的阻碍和促进因素 6 个一级条目,充分考虑了公共卫生领域的独特性和特定需求,证据质量评级更具针对性。例如,在面对复杂的公共卫生问题时,决策者通常将公平性作为关键的实施考虑因素之一来帮助做出决策^[16],而 PHE-Grading 也将健康公平性作为评估条目,旨在确保所有公众成员都能公平、无障碍地享有健康服务与资源。^[7]

证据质量评价形式从“升降级制”优化为“量化评分制”,评价领域从“大众领域”细化到“特定领域”,反映出了公共卫生决策研究证据质量的分级系统在不断优化和向更深层次发展。专家学者积极探索和研究公共卫生决策证据质量分级方法有利于推动循证公共卫生决策的发展。^[17]

4.3 公共卫生决策研究证据质量分析

在进行公共卫生决策时通常需要高质量的研究证据,通过对 GRADE 与 PHE-Grading 证据质量评级的差异性分析发现,现存的公共卫生决策研究证据质量多集中在中低甚至更低水平。与本研究发现一致的是,2018 年 Jenny Moberg 发表的一篇文章中也谈到卫生系统和公共卫生干预措施的证据质量通常处于较低或极低水平^[16],这说明当前公共卫生决策研究高质量证据较为缺乏,未来仍需加强高质量研究以支持科学决策。从证据质量评价的角度,即 PHE-Grading 证据质量评价的结果来看,纳入研究的

结局指标在研究执行质量、精确性、大效应量和负偏倚 4 个一级条目上被评为“0 分”的比例高于 50%,因此,当前公共卫生决策缺乏高质量研究证据的主要原因可能体现在以下四个方面:一是在研究执行质量方面,公共卫生决策研究的偏倚风险普遍偏高,研究过程中可能存在偏差或操作不规范,影响数据的准确性和可靠性;二是在精确性方面,公共卫生决策研究可能对其结果的测量和效应量估计值的把握存在误差,导致证据的精确性不足;三是在大效应量方面,公共卫生决策研究结局指标的合并效应量普遍为无效应,即相应检验无统计学意义;四是在研究的负偏倚方面,当前公共卫生决策研究普遍不存在负偏倚,可能存在高估其真实效果的情况。除此之外,研究的发表偏倚、剂量—反应关系以及证据的阻碍和促进因素等其他 11 个方面也对公共卫生决策研究证据的质量存在一定影响,这些同样值得今后的研究者、专家或学者去关注与重视。

5 建议

一是建议今后对公共卫生决策研究证据质量的评价优先考虑采用 PHE-Grading 系统。该系统在设计时充分考虑了公共卫生决策的特定需求,与 GRADE 相比,PHE-Grading 对公共卫生决策证据确信度的把握可能更为精准。这不仅有助于提升决策的科学性和准确性,还能更好地权衡公共卫生行动的利与弊,以促进政策目标的实现,最终达到提升社会整体福祉的效果。

二是建议公共卫生决策研究领域的专家或学者继续加强高质量研究以支持科学决策。尽管 PHE-Grading 系统为公共卫生决策提供了有力支持,但当前公共卫生决策证据的整体质量仍集中在中低水平。系统评价作为高质量证据的来源,能为卫生服务决策者提供宝贵的信息支持,然而低质量的系统评价则可能误导决策^[18],因此,加强高质量公共卫生研究,提升证据的生产能力和水平,对于科学决策、优化公共卫生政策尤为重要。公共卫生决策相关研究者未来可以从研究执行质量、精确性、大效应量和负偏倚四个主要方面进行突破,结合 PHE-Grading 系统的其他 11 个评价条目,加强对这些方面的研究和评估,同时加大对公共卫生研究的投入,鼓励跨学科合作,推动创新方法和技术在公共卫生研究中的应用,以产生更全面、可靠和高质量的证据,为公共卫生政策的制定提供有力支撑。

此外,研究初次应用 PHE-Grading 进行证据质量评价,为完善和进一步探索当代公共卫生领域证据质量评价体系提供研究参考,进而促进公共卫生领域的科学决策,具有一定创新性。然而,新生的证据质量分级系统 PHE-Grading 还未得到推广使用,因此用户对该系统的使用情况尚未展开调查,在未来的研究中将作进一步探索。

作者声明本文无实际或潜在的利益冲突。

参 考 文 献

[1] Kneale D, Rojas-García A, Raine R, et al. The Use of Evidence in English Local Public Health Decision-Making: A Systematic Scoping Review[J]. Implementation Science, 2017, 12(1):1-12.

[2] 陈静静, 潘琳敏, 周波. 循证公共卫生决策的发展与应用[J]. 智慧健康, 2020, 6(8): 45-48.

[3] 李幼平, 杨晓妍, 陈耀龙, 等. 我国公共卫生领域的循证决策与管理:挑战与探索[J]. 中国循证医学杂志, 2008, 8(11): 945-950

[4] 童峰, 林移刚, 张冲. 循证决策:一种忠于证据的公共卫生决策模式[J]. 医学与哲学(B), 2015, 36(5): 4-7.

[5] MuirGray, 唐金陵. 循证医学:循证医疗卫生决策[M]. 北京大学医学出版社, 2004.

[6] 石振宇, 加焱冰, 徐进, 等. 我国卫生政策与体系研究的政策转化分析[J]. 中国卫生政策研究, 2020, 13(10): 41-47.

[7] 商雪, 邓欣欣, 郭康乐, 等. 运用德尔非法构建公共卫生决策证据质量分级系统[J]. 中国卫生政策研究, 2023, 16(10): 71-78.

[8] 龙囿霖, 林春滢, 郭嘉琪, 等. Meta 流行病学的产生与发展[J]. 中国循证医学杂志, 2022, 22(6): 716-721.

[9] 孟庆跃. 卫生政策与体系研究能力述评[J]. 中国卫生

政策研究, 2020, 13(10): 14-19.

[10] 贾文琴, 杨克虎, 田金徽, 等. Cochrane 系统评价发表状况调查[J]. 中国循证医学杂志, 2009, 9(6): 635-639.

[11] 张鸣明, 李幼平, 帅晓. Cochrane 系统评价精萃循证医学实践高质量的实时证据[J]. 中国循证医学, 2002, 2(1): 63-64.

[12] 赵冰. “非典”疫情之后公共卫生问题之我见[J]. 上海预防医学杂志, 2003, 15(10): 485-486.

[13] 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]. The British Medical Journal, 2008, 336(7650): 924-926.

[14] Guyatt G, Oxman A D, Akl E A, et al. GRADE Guidelines: Introduction-GRADE Evidence Profiles and Summary of Findings Tables [J]. Journal of Clinical Epidemiology, 2011, 64(4): 383-394.

[15] Rehfuess E A, Akl E A. Current experience with applying the GRADE approach to public health interventions: an empirical study[J]. Bmc Public Health, 2013, 13(1): 9-22.

[16] Jenny M, Oxman A D, Sarah R, et al. The GRADE Evidence to Decision (EtD) framework for health system and public health decisions[J]. Health Research Policy & Systems, 2018, 16(1): 45-60.

[17] 杨超群, 刘文迪, 王子怡, 等. 公共卫生决策证据质量分级方法问题的定性系统评价[J]. 中国循证医学杂志, 2023, 23(12): 1413-1418

[18] 王靖, 刘琴, 翁淳光, 等. 国内公共卫生研究领域系统评价/Meta 分析的质量评价[J]. 中国循证医学杂志, 2010, 10(12): 1367-1374.

[收稿日期:2024-07-16 修回日期:2024-09-24]

(编辑 赵晓娟)