

· 论 坛 ·

国内循证医学证据分级体系的引入及其在中医药领域面对的挑战

陈 薇¹ 方赛男¹ 刘建平¹ 陈可冀²

循证医学(evidence-based medicine, EBM)是近 20 余年来在临床医学实践中发展起来的一门新兴学科,即“遵循证据的医学”,被誉为是医学界的一次革命。证据是 EBM 的核心,最好的研究证据主要来自临床的人体研究。但是,由于临床研究数量大,种类繁多,且质量良莠不齐,医疗决策者及临床医生很难花费大量的精力和时间从浩瀚的信息海洋中筛选出真实而适用的证据。于是,证据分级体系应运而生,其目的在于对不同来源的证据进行质量分级,临床医生只需充分利用研究人员预先确立的证据分级标准和推荐意见使用各种高质量证据即可,从而达到正确、合理使用证据的目的。

1 国内对于证据分级体系的引入和应用

1996 年,国内在一份动脉瘤性蛛网膜下腔出血处理准则的翻译版中首次出现了“证据分级”的概念,该准则中使用了 1986 年 David Sackett 推荐的证据分级系统来提供推荐意见^[1]。1998 年,唐镜波在药品监督指导的文章中介绍了美国卫生保健政策与研究管理局(Agency for Health Care Policy and Research, AHCPR)制订的证据等级评定标准^[2]。1999 年,廖静秋等^[3]在介绍系统综述时,提到了需要根据证据等级对证据进行综合,表明了证据分级也可用于系统综述的制作中。1999 年 10 月,丹麦 Erik Magid 博士在全球作 EBM 循环讲座上海站中正式向国内学者说明了证据质量分级的必要性及重要性,并举例介绍了美国预防服务工作组制定的证据分级标准^[4]。自此开始,国内学者开始关注证据分级体系,积极向国内介绍并引进国外证据分级系统,包括对牛津循证医学中心(Oxford Centre for Evidence-based Medicine, OCEBM)证据分级系统修订的翻译^[5],以及对国

外证据分级标准进行了系统性综述^[6],并列举了“美国证据金字塔、新西兰证据之箭、加拿大定期体检特别工作组(Canadian Task Force on the Periodic Health Examination, CTFPHE)分级系统、AHCPR 标准、苏格兰院际指南网络(Scottish Intercollegiate Guidelines Network, SIGN)标准、OCEBM 标准”等具有代表性的证据分级标准。

同时,随着国外制订临床实践指南的兴起,国内学者也渐渐意识到制订临床实践指南的重要性,以及证据分级对其的重要意义。2004 年,杨振华提出国内应当重视临床实践指南,尤其是循证临床实践指南的制订,并建议在制订指南过程中应用好证据的分级系统^[7]。但是,早期国内制订指南时,应用证据分级系统的并不多,其中应用较多的证据分级为刘建平 2007 年在《中国中西医结合杂志》上发表的基于国际多个证据分级体系充分考虑传统医学特点的证据分级系统(表 1)^[8-11],其次是 OCEBM 证据分级^[12]。

表 1 基于“证据体”的临床研究证据分级

证据级别	描述
I a	由随机对照试验、队列研究、病例对照研究、病例系列这 4 种研究中至少 2 种不同类型的研究构成的证据体,且不同研究结果的效应一致
I b	具有足够把握度的单个随机对照试验
II a	半随机对照试验或队列研究
II b	病例对照研究
III a	历史性对照的病例系列
III b	自身前后对照的病例系列
IV	长期在临床上广泛运用的病例报告和史料记载的疗法
V	未经系统研究验证的专家观点和临床经验,以及没有长期在临床上广泛运用的病例报告和史料记载的疗法

2004 年,国际“推荐分级的评价、制定与评估(Grades of Recommendations Assessment, Development and Evaluation, GRADE)”工作组正式推出了 GRADE 证据质量分级和推荐强度系统(以下简称 GRADE 系统)^[13],成为证据发展史上的里程碑事件。GRADE 系统规范了国际上证据等级分级标准不一的现象,成为国际上公认的被纵多国际组织机构

基金项目:国家自然科学基金青年基金项目(No. 81603451)

作者单位:1. 北京中医药大学循证医学中心(北京 100029); 2. 中国中医科学院西苑医院心血管病中心(北京 100091)

通讯作者:刘建平, Tel: 010-64286760, E-mail: Liujp@bucm.edu.cn

DOI: 10. 7661/j. cjim. 20171009. 233

所采纳的证据分级系统。随着 GRADE 的推广及中国 GRADE 中心的建立,国内学者对 GRADE 系统亦开始重视起来。2011 年,刘为民等^[14]中国中医科学院和中国 EBM 中心的一批专家肯定了 GRADE 对中医药的价值,认为它将有利于中医临床疗效评价,有利于发扬中医的特色,有助于中医学走向现代化和国际化。但 GRADE 系统在国内的应用仍然十分有限。笔者检索了截止到 2017 年 7 月发布的数百篇中医药相关临床指南,发现仅有 5 篇使用了 GRADE 系统进行证据分级,另有 20 余篇指南使用了 GRADE 的推荐强度。

尽管 GRADE 系统在国内受到越来越多的关注,但是在应用过程中仍然遇到了许多问题,尤其是在评价中医药相关的证据时。首先,GRADE 系统无法将中医药研究中主观性的经验、建议和意见进行规范化,并充分利用其价值。现代医学的临床证据主要来自于最新的临床研究,特别重视大规模、多中心的随机对照试验,并不在意经验或共识的评价。而中医学是中华民族历代社会实践基础上,不断总结保障人体健康经验基础上逐渐形成的一门传统医学。古代医家经验是数千年中医学知识积累的集中体现,许多古代医家经验尤其是名医大家、经典著作的论述,至今仍是现代临床医生诊治疾病的准绳,在指导临床实践和临床研究中起着巨大作用。因此,中医学的临床证据不应该只关注现代中医临床研究的证据,也应当重视历代传承的古今医家经验。如果参照目前 GRADE 分级系统,中医临床专家经验和中医经典古代文献不在其证据分级体系之内,其结果而言可能会出现临床确有疗效的诊疗技术难以得到推荐的现象,从而影响了中医临床实践指南的权威性、可信性和实用性。此外,中医学兼具人文和自然学科的双重特点,医学最佳决策的证据来源不应该仅限于一系列人体试验中获得的客观化数据,而且应从其操作者,即研究者和各个环节中的医疗人员的主观世界里获得有价值的信息。因此,对中医学临床证据进行分级和评价时,应当考虑经典的定量研究和关联分析与社会学的定性研究方法等多种研究方法相结合的证据。

2 国内中医药证据分级体系的提出

GRADE 系统在国内应用面临的问题,主要还是来源于与中医学的特色不相吻合。如何在证据分级中体现中医学的整体观念和特色,国内部分学者提出了自己的想法。

2007 年,卢笑晖等^[15]提出,中医证据应由病因、病机、症状、证候等构成,并认为以临床试验及其系统综述作为中医临床证据是不合理的,传承了千百年的

中医典籍才应当是中医学的最高水平证据,以专家经验为主要内容的个案报告或病例前后自身对照研究也可作为中医临床证据的重要组成部分,现代临床研究也是一个重要证据来源。此论述虽不是一个证据分级体系,但也为中医临床证据的构成提供了思路。也有学者基于自己对中医特点的理解以及 EBM 的应用,提出了各具特色的证据分级系统。

2007 年,北京中医药大学循证医学中心的刘建平^[16]提出了“基于证据体的传统医学证据分级”,该分级体系考虑到中医证据的构成,将效应一致的随机对照试验、队列研究、病例对照研究、病例系列这 4 种研究中至少 2 种不同类型的研究构成的证据体列为最高级证据,且纳入了临床上运用广泛的病例报告和史料记载的疗法作为低级别证据(表 1)。此标准自提出便得到了国内指南制定者的认可,该文章目前已被引用 138 次,下载 1 277 次。2007 年中国中医科学院与 WHO 西太区合作开发的第一套中医药循证临床实践指南,涉及 28 种有中医诊疗特色和优势的疾病,其中就有 26 个疾病的指南使用了这一证据分级体系。是目前国内中医药证据分级体系中应用最广泛的体系。

2009 年,李敬华^[17]提出,中医临床区别于现代医学的一大特色就是辨证论治,故提出了以“辨证论治”作为评价第一要素,将辨证论治的研究列为 I 级证据,其中按照研究设计类型和质量再分为 4 等,非严格辨证论治的研究为 II 级证据,非辨证论治的研究为 III 级证据(表 2)。并且提出了详细的质量评分方法。

表 2 中医治疗性研究文献证据等级体系

证据级别	描述
I 级	辨证论治的研究 I a 级 辨证论治的随机对照试验,设计及完成质量良好 I b 级 辨证论证的其他类型临床研究,设计及完成质量良好 I c 级 辨证论治经验的专家经验总结,记录者和专家层次高 I d 级 辨证论治的个案,资料完整,疗效肯定
II 级	非严格辨证论治的研究
III 级	非辨证论治的研究

2010 年,李可建^[18]欲寻找有效解决病机假说的方法,分析了病机假说检验标准,提出中医药临床疗效试验研究是检验病机假说的重要临床证据,并根据证据强度将临床疗效试验分为一至四级。一级包括随机对照试验、半随机对照试验、组群随机对照试验、单个病例的随机对照试验、交叉试验,其中以设计良好、执行严格的随机对照试验为最高证据;二级包括队列研究设计及前后对照试验;三级包括设有对照组的回顾性研究或病例对照研究;四级包括病例系列、个案总结

及专家述评。此标准基本与国际以“研究设计类型”为中心的
证据分级体系一致,只是将其扩展应用于中医病机假说中。

2010 年,王阶等^[19]采用调查问卷的方式对全国中医领域 384 位专家进行了两轮采访,调查结果显示,“目前仍在使用的四大经典医籍”、“目前仍在使用的国家标准及行业制定的标准”、“所有随机对照试验的系统综述”、“经过系统整理的名老中医经验”是中医临床最高证据,“无对照的病例观察”和“医案医话”是最低证据(表 3)。

表 3 中医临床证据分级与评分体系

证据级别	描述	评分
I a	目前仍在使用的四大经典医籍	每项 10 分
I b	目前仍在使用的国家标准及行业制定的标准	每项 9 分
	多个随机对照试验的系统综述	每项 9 分
	经过系统整理的名老中医经验 (以国家中医药管理局确认的名老中医为准)	每项 9 分
II a	单个的正确设计的随机对照试验的结果	每项 7 分
II b	目前仍在使用的国家统编教材	每项 6 分
	设计良好的半随机对照试验	每项 6 分
III	目前仍在使用的其他古代经典医籍	每项 5 分
IV	无对照的病例观察	每项 2 分
	医案医话	每项 2 分

2012 年,任玉兰等^[20]提出基于 EBM 针灸临床研究证据评价体系,并构建了古籍载录证据、专家经验证据、随机对照试验、非随机对照试验、病例序列研究及个案报告的质量评价表,每个质量评价表均有详细的评价项目、指标和分值,通过累加,分值较高的为高质量证据,最后基于以上评分的证据建立证据分级及推荐强度标准(表 4)。

表 4 研究证据强度及推荐级别标准

证据级别/ 推荐级别	描述
I a	符合二者之一:(1) 高质量古籍载录证据 + 高质量随机对照试验研究;(2) 高质量专家经验证据 + 高质量随机对照试验研究
I b	符合四者之一:(1) 高质量古籍载录证据 + 高质量非随机临床对照试验研究;(2) 高质量专家经验证据 + 高质量非随机临床对照试验研究;(3) 古代文献记载证据 + 高质量随机对照试验研究;(4) 专家经验证据 + 高质量随机对照试验研究
I c	符合三者之一:(1) 高质量古籍载录证据;(2) 高质量专家经验证据;(3) 高质量随机对照试验研究
II	符合三者之一:(1) 古代文献载录证据 + 高质量非随机临床对照试验研究;(2) 名老专家经验证据 + 高质量非随机临床对照试验研究;(3) 随机对照试验研究
III	符合二者之一:(1) 高质量非随机临床对照试验研究;(2) 随机对照试验 + 高质量病例序列研究
IV	符合二者之一:(1) 高质量病例序列研究;(2) 高质量个案报道

2013 年,汪受传等^[21]提出了基于 Delphi 法的中医文献证据分级标准,其中纳入的“基于古今文献的中医专家共识”,指古代医籍记载、历代沿用至今、当代专家调查意见达成共识者,“当代中医专家共识”是指当代专家调查意见达成共识者,“专家意见”仅指个别专家意见(表 5)。并已在《中医儿科常见病诊疗指南》^[22]中应用。

表 5 基于 Delphi 法的中医文献证据分级体系 (汪受传等)

证据级别	描述	推荐级别	描述
I	大样本,随机研究,结果清晰,假阳性或假阴性的错误很低	A	至少有 2 项 I 级研究结果支持
II	小样本,随机研究,结果不确定,假阳性或假阴性的错误较高	B	仅有 1 项 I 级研究结果支持
III	非随机,同期对照研究和基于古今文献的中医专家共识	C	仅有 II 级研究结果支持
IV	非随机,历史对照和当代中医专家共识	D	至少有 1 项 III 级研究结果支持
V	病例报道,非对照研究和专家意见	E	仅有 IV 级或 V 级研究结果支持

2016 年,袁敬柏^[23]在对当前中医诊疗指南证据评价方法进行梳理分析的基础上,结合中医学理论及临床特色,提出将中医诊疗证据分为理论证据和研究证据两类,分别进行证据分级,其中研究证据采用牛津循证医学中心 2013 版证据分级标准^[24],然后综合形成推荐意见。并且该分级标准涉及了诊断和治疗干预两方面(表 6、7)。

表 6 理论证据分类、分级与评价

证据级别	诊断措施	干预措施
I a	《黄帝内经》、《伤寒杂病论》、《黄帝八十一难经》等东汉及东汉之前的著述支持	《黄帝内经》、《伤寒杂病论》、《黄帝八十一难经》等东汉及东汉之前的著述支持
I b	晋到清代医家相关的论述,具有较好的传承	晋到清代医家相关的论述,具有较好的传承
II	晋到清代医家相关的论述,传承存在不一致	晋到清代医家相关的论述,传承存在不一致
III a	近现代(民国 - 当代)名中医的著述中明确阐述	近现代(民国 - 当代)名中医的著述中明确阐述
III b	近现代(民国 - 当代)名中医医案能体现	近现代(民国 - 当代)名中医医案能体现
IV a	教材,行业规范性文件	教材,行业规范性文件,专著
IV b	专著	一定数量的单个病例报道

3 国内中医药证据分级体系的挑战

由以上证据分级体系可以看出,国内中医药的证据分级体系尚不健全,有的基于“辨证论治”理念,有的

表 7 理论证据用于诊断与干预措施推荐分级

推荐级别	描述
A	全部 4 类或 1 类证据加 2~4 类证据的 2 类或 2~4 中的 3 类,具有一致结论,今后研究不可能形成否定性证据
B	2~4 类证据中 2 类,具有比较一致结论,今后研究形成否定性证据可能性小
C	2~4 类证据中的 1 类
D	无证据支持,或观点分歧

重视“四大经典”的古籍文献,主要思路可以概括为传承千百年的文献古籍和临床经验不能丢,可以考虑作为高等级证据。此方法确实可以解决目前中医临床证据不足的问题,但也存在一些问题:(1)多数分级系统并不包含推荐意见的形成。推荐意见是制定临床实践指南的核心,一般需要考虑证据数量、质量、一致性、实用性、临床可推广性等问题,还要根据证据级别、患者偏好、卫生经济学评价等对推荐强度做出说明。(2)国内证据分级体系均由专家学者个人提出,尚未形成标准或规范。

笔者认为,建立国内中医药证据分级系统,主要有以下关键问题:(1)要注重证据来源的多样性,在评价时不可仅关注随机对照试验而忽视古籍文献的报道,需要建立对古籍文献的科学评价方法;(2)对中医药研究中主观性的经验、建议和意见进行规范化,并建立其科学的评价方法;(3)应考虑经典的定量研究和关联分析与社会学的定性研究方法等多种研究方法相结合的评价方法;(4)针对不同的中医药方法(如中药和针灸),探讨其各自关键的中医药特色,并制作不同的证据分级体系;(5)重视方法学的规范性及制定专家来源的多学科性。在初期阶段,可以考虑由目前应用最广泛的中医证据分级体系作为基础,按照标准制定的规范进行深入研究和补充完善,然后在行业内推广使用。

总之,制定统一且适宜的中医证据分级体系是提高中医药临床实践指南质量的关键,期待能有更多专家学者提供自己的思路,开拓创新,也希望决策者提高认识付诸实践。

参 考 文 献

[1] 动脉瘤性蛛网膜下腔出血处理准则[J]. 高血压杂志, 1996, 4(3): 78-80.

[2] 唐镜波. 药品管供用的监督指导一体化[J]. 中国药房, 1998, 9(5): 198-201.

[3] 廖静秋, 赵亚群, 傅鹰. 系统性综述简介[J]. 药物流行病学杂志, 1999, 8(1): 40-45.

[4] Erik Magid, 胡晓波, 吴文俊, 等. 循证医学[J]. 上海医学检验杂志, 2000, 15(6): 325-328.

[5] 刘合利, 杨连粤. 证据分级和建议等级于 1999 年 11 月 18 日最新修订[J]. 腹部外科, 2001, 14(6): 335-336.

[6] 管红珍, 彭智聪, 傅鹰. 循证医学中文献证据等级标准的系统性综述[J]. 药物流行病学杂志, 2002, 11(3): 145-148, 169.

[7] 杨振华. 指南与循证医学[J]. 中国循证医学杂志, 2004, 4(11): 747-749.

[8] 中国中西医结合学会肝病专业委员会. 肝纤维化中西医结合诊疗指南[J]. 中国中西医结合杂志, 2006, 26(11): 1052-1056.

[9] 中华医学会感染病学分会肝衰竭与人工肝学组, 中华医学会肝病学会重型肝病与人工肝学组. 肝衰竭诊疗指南[J]. 中华肝脏病杂志, 2006, 14(9): 643-646.

[10] 中华医学会肝脏病学分会脂肪肝和酒精性肝病学组. 非酒精性脂肪性肝病诊疗指南[J]. 中华肝脏病杂志, 2006, 14(3): 161-163.

[11] 中华医学会肝病学会脂肪肝和酒精性肝病学组. 酒精性肝病诊疗指南[J]. 中华肝脏病杂志, 2006, 14(3): 164-166.

[12] 蒋朱明, 江华, 詹文华, 等. 制定肠外肠内营养指南和规范的“指南”: 方法学、推荐意见分级与通过程序[J]. 中国临床营养杂志, 2006, 14(5): 283-288.

[13] Atkins D, Best D, Briss PA, et al. Grading quality of evidence and strength of recommendations[J]. BMJ, 2004, 328(7454): 1490.

[14] 刘为民, 刘保延, 何丽云, 等. 引入 GRADE 体系制定中医药国际标准的思考与实践探析[J]. 亚太传统医药, 2011, 7(1): 1-5.

[15] 卢笑晖, 王阶, 袁敬柏. 中医临床证据研究[J]. 中医杂志, 2007, 48(11): 965-967.

[16] 刘建平. 传统医学证据体的构成及证据分级的建议[J]. 中国中西医结合杂志, 2007, 27(12): 1061-1065.

[17] 李敬华. 中医疗疗文献质量评价方法及示范应用研究[D]. 北京: 中国中医科学院, 2009.

[18] 李可建. 检验病机假说的临床证据分级研究[J]. 山东中医药大学学报, 2010, 34(4): 294-295.

[19] 王阶, 何庆勇, 姚魁武, 等. 中医临床证据分级与评分体系研究[A]. 经济发展方式转变与自主创新——第十二届中国科学技术协会年会论文集(第 3 卷)[C]. 福州: 中国科学技术协会学会学术部, 2010: 1300-1304.

[20] 任玉兰, 吴曦, 梁繁荣. 基于循证医学针灸临床研究证据评价体系的初步构建[J]. 辽宁中医杂志, 2012, 39(2): 205-209.

[21] 汪受传, 陈争光, 徐珊, 等. 建立循证中医临床实践指南证据分级体系的构想[J]. 世界科学技术-中医药现代化, 2013, 15(7): 1488-1492.

[22] 汪受传, 李辉, 徐玲. 中医儿科临床诊疗指南·小儿鼻鼈[J]. 中华中医药杂志, 2016, 31(4): 1352-1355.

[23] 袁敬柏. 建立适合中医临床诊疗证据评价方法的建议[J]. 中华中医药杂志, 2016, 31(4): 1146-1148.

[24] Center for Evidence-Based Medicine. OCEBM Levels of Evidence[EB/OL]. <http://www.cebm.net/ocebm-levels-of-evidence/>. 2017.

(收稿: 2017-08-25 修回: 2017-09-20)
责任编辑: 汤 静