

专题——中医临床证据评价方法与转化应用

中医药领域的证据与推荐:理论、方法与实践

陈耀龙^{1,2} 张晓雨¹ 商洪才¹

(1 北京中医药大学东直门医院中医内科学教育部重点实验室,北京,100700; 2 兰州大学基础医学院循证医学研究所/GRADE 中国中心,兰州,730000)

摘要 近年来中医药在我国发展越发受到重视,但在西医系统认可度以及高质量疗效研究方面尚需加强。文章以循证医学的视角对中医药领域研究偏倚和证据进行探讨,从证据质量与推荐强度的分级以及各自的影响因素展开分析,以期通过原始研究、系统评价、实践指南等途径改善中医药证据的生产和应用。

关键词 中医药发展;循证医学;证据;推荐

Evidence and Recommendation on Traditional Chinese Medicine: Theory, Methods and PracticeChen Yaolong^{1,2}, Zhang Xiaoyu¹, Shang Hongcai¹

(1 Key Laboratory of Chinese Internal Medicine of Ministry of Education, Dongzhimen Hospital, Beijing University of Chinese Medicine, Beijing 100700, China; 2 Evidence-Based Medicine Center, School of Basic Medical Sciences Lanzhou University/Chinese GRADE Center, Lanzhou 730000, China)

Abstract In recent years, the development of Traditional Chinese Medicine (TCM) has become increasingly important, but it still needs to be strengthened in the aspects of the recognition of Western Medicine and high-quality treatment research. This paper discussed research bias and evidence of TCM in the perspective of Evidence-based Medicine, presented the grading factors of the quality of evidence and the strength of recommendation in area of TCM. It is expected to improve the production and application of evidence in TCM through the primary studies, systematic reviews and practice guidelines.

Key Words Traditional Chinese Medicine; Evidence-based Medicine; Evidence; Recommendation

中图分类号:R2-03 文献标识码:A doi:10.3969/j.issn.1673-7202.2017.06.001

1 背景

尽管中医药已经历了上千年的历史,且近年来无论从国家政府层面还是学术与医疗实践层面,不仅得到越来越多的关注和重视,也获得了更多的支持和资源投入^[1-4]。然而,在高质量科学研究证实其疗效,以及获得西医系统和西方国家主流医学广泛认可方面,仍然需要进行大量的工作。Cochrane 协作网的创始人 Iain Chalmers 爵士在其经典著作《治疗的真相》中指出^[5],面对任何一种可能有效的诊断或治疗方案,不管是西医还是传统或补充替代医学,都需要回答一个问题:它的科学依据是什么?这种依据到底有多可靠?为发现真实的疗效,我们应该如何排除安慰剂效应、疾病的自愈、向均数回归等现象,以及研究人员或医药企业对疗效的刻意误导?循证医学的产生及其方法学的出现,对中医药的发

展和解决以上问题提供了重要的思路^[6-8],中医药领域越来越注重科学研究证据在临床决策中的作用,本文将从证据质量和推荐强度角度,探讨其对中医药科学化和国际化发展的重要性。

2 偏倚、研究和证据

David Sackett 指出医学研究中的偏倚超过 100 多种^[9],比如随机对照试验可能的偏倚见表 1^[10-11],对于补充替代医学而言,偏倚同样存在。为减少可能的偏倚,发现中医药的真实疗效,早在北宋年间,中国的药师们就进行了积极的探索。据 1061 年成书的《本草图经》记载,为了验证人参的作用,需要设置同期的外部对照:“相传欲试上党人参者,当使二人同走,一与人参含之,一不与,度走三、五里许,其不含人参者,必大喘,含者气息自如者,其人参乃真也”。这说明,中医药先驱们很早就意识到严谨的

基金项目:国家自然科学基金青年项目(编号:81503459)

作者简介:陈耀龙(1981—),男,博士后,研究方向为证据分级与知识转化,E-mail:chenyaolong21@163.com

通信作者:商洪才,男,博士,研究员,博士研究生导师,重点实验室主任,研究方向:临床证据的评价与转化方法学及其机制研究,Tel: (010) 84012510, E-mail: shanghongcai@foxmail.com

科研方法的重要性。同样,随机分组在中医药研究中的重要性也不言而喻。如果不能够实现参与对照的2组人员的基线均衡,具有同等的机会获得人参的话,就算设置了对照,也很难发现真实疗效。比如,有意让体质较强的人,而非较弱的那个人服用人参,就算没有服用人参的那位气喘,也不能说明是没有服用人参造成的。同时,为了避免人为因素影响随机分组造成选择性偏倚,需要采用合理的方法对受试人群和实施人员隐藏随机分组的情况。因为在纳入受试对象的过程中,如果分组情况被暴露,相关人员很可能自觉或不自觉地将特殊人群有意从某组剔除或倾向性地分到某一组,破坏随机分配,从而导致治疗结果被夸大或减小^[12-13]。同样,为避免安慰剂效应或观察者的主观误导,如果再用一种淀粉或不含有有效成分的植物,同时也给另外一个人服用,通过对最终结果的观察,那么就能够对人参的效果更有信心。所以,无论是何种中医药治疗方案,随机、对照和安慰剂,其实是能够帮助,而非阻碍验证中医药的疗效^[14]。另一方面,试验完成产出结果后,需要透明的报告和发表,因为无论何种原因的选择性报告都会造成研究结果的不完整,从而带来偏倚^[15-16]。

此外,如果当时仅进行了这样一项研究,其实还是很难说明人参的效果,因为它的样本量太小,在人群的代表性和结果的精确性方面,可能仍然会存在偏离真实效果的情况。样本量太小,结果显示的效果就有可能是偶然出现的情况^[17]。所以,我们需要系统收集验证人参疗效的所有研究,对其进行评价和综合,在这种情况下,我们才能够获得有关该干预

措施的证据体,不至于被单个研究的结果所误导。即,针对某一个或某一类具体问题,尽可能全面收集有关该问题的全部原始研究,进行严格评价、综合、分析、总结后所得出的综合结论,是对多个原始研究再加工后得到的证据。以上综合证据的方法可分为3大类,即系统评价(Systematic Review,SR)/Meta分析,卫生技术评估(Health Technology Assessment,HTA)和实践指南(Practice Guideline),其共同点为:均基于原始研究,对其进行系统检索、严格评价和综合分析;均可进行证据质量分级;均可作为决策的最佳依据。

3 证据质量的分级及其影响因素

这种获取针对某个治疗措施全部证据的过程,就是系统评价方法的核心,而系统评价最终纳入的符合标准的研究,就构成了循证医学中的证据体。下一步需要完成的,是对该证据体的可信度进行判断——在此引出了证据质量的定义:即在获得证据体的前提下,我们对观察值接近真实值的确信程度^[18]。因为真实值是无法被直接发现的,所以我们只能依靠观察值来对其进行推断。如果我们非常确信观察到的结果就是真实值,那么此时我们将系统评价获得的证据体定义为高质量。比如有设计严谨的若干大样本的随机对照试验或观察性研究发现某种中医疗法(可以是草药、针灸或推拿等)有效,在现有的科学认知水平下,没有或很少发现存在影响真实值的偏倚,未来进一步的研究也很难改变当前的结果,那么该种疗法的证据质量为高。反之,如果我们对观察到的结果究竟能否反映真实值完全不确定,则将其定义为低质量,比如仅有几例病例报告显

表 1 随机试验的研究局限性

序号	
1	无隐蔽分组
	招募受试者的人知道下一位受试者将被分到哪一组(或交叉试验中的哪一时期)(按星期几、出生日期或图表编号等来分配的“假”或“半”随机试验的主要问题)
2	未设盲
	患者、照护者、记录结果者、裁定结果者或数据分析者,知道患者分配到哪一组(或交叉试验中目前正在接受的药物治疗)
3	不完整报告患者和结局事件
	优效试验中的失访和未遵从意向性治疗原则;或非劣效试验中的失访和未同时进行2种分析;仅分析坚持治疗者和分析所有可得结果数据的患者
4	选择性结果报告偏移
	不完整报告或不报告某些结果及基于结果的其他内容
5	其他局限性
	· 因早期获益而终止试验
	· 使用未经验证的结果测量方法(如患者报告的结果)
	· 交叉试验中的延滞效应
	· 整群随机试验中的招募偏移

示某种祖传药方治疗某种疾病有特效,我们发现了大量潜在影响其结果真实性的偏倚,未来的研究很可能改变当前的结果,则证据质量为低或极低。

同样,正如我们无法获得真实值一样,我们无法直接给出证据的质量。我们对证据质量的判断,依赖于对证据产生的全过程中可能存在的各种偏倚的判断。在 GRADE(The Grading of Recommendations, Assessment, Development and Evaluations)系统中,有五个方面来帮助评估者判断:1)组成证据体的各个证据单体的偏倚风险,主要存在于每个研究的设计、实施和评价环节。比如随机方法、分配方案的隐藏、盲法、失访的情况、对结局指标的完整报告等。2)组成证据体的各个证据单体之间结果的一致性。理想情况下,单体之间可能不完全一致,但也不应该出现较大分歧甚至自相矛盾。如果出现了,当我们无法合理解释这种异质性时,我们对其证据体的确信程度要降低。3)不精确性。当汇总了证据体的结果后,其预测真实值的观察值范围(95%的可信区间)不应该过大,试想如果某种中医药疗法在降低舒张压幅度方面,其点估计值为 5 mmHg(该值也是判断临床有效降压效果的切点),区间估计值为 2 ~ 8 mmHg,那么就算其有统计学意义,我们也很难确定其应用到临床后到底有无实际效果,这对医师是否采纳该疗法降压的决策造成了不确定性,所以需要更大的样本量去将其可信区间变窄,无论其变为 2 ~ 4 mmHg 还是 6 ~ 8 mmHg,临床决策都将变得清晰和容易,前者可能不会推荐使用该药物,后者可能会推荐。4)间接性。回到人参的例子,如果当时参与试验的都是青壮年男性,那么人参的作用在儿童、老人和女性当中的效果,可能会有不一样。同样,受试者所含的人参如果来自某个特定的产地(假设不同的产地效果不完全一样),那么我们也只能得出该产地人参的作用,而非所有人参的作用,所以当研究者下结论时如果没有区分,则我们对结论的可靠度也可能存在疑虑。5)发表偏倚。假设当时其实做了 3 个试验,被《本草图经》记录是只是有效的那个,另外 2 个因为无效而未被载入,那么我们对人参的作用也可能存在进一步的不确定性。

4 推荐强度的分级及其影响因素

确定证据质量的目的是为下一步做临床推荐奠定基础^[19-20]。假设我们将人参作为治疗某种疾病的药物来推荐给患者,除了确定人参的有效性之外,还需进一步了解其不良反应,平衡效果和不良反应对患者的实际影响,以及其成本和患者对服用人参

的偏好。如果效果好(证据质量同时为高),不良反应很小,价格也便宜,服用方便,那么很可能在临床上会得到一个强推荐使用(即临床医师非常确定患者服用人参后利大于弊),在这种情况下,除非特别说明,否则医生都应该给患者服用人参,患者也都应该接受人参的治疗,卫生政策制定者应该将其纳入医保。反之,虽然有证据显示人参有效,但其证据质量较低,或不良反应明显,或价格昂贵,或多数患者对服用人参的偏好差异性很大,那么临床有可能会形成一个弱推荐使用,即临床医师认为患者服用人参可能利大于弊,但需要进一步跟患者讨论后共同决策。此外,针对具体的实施环境,推荐或不推荐人参还要考虑其是否可以获取到(可及性)以及是否方便实施(可实施性)。

5 跨越从知到行的鸿沟

针对中医药领域证据的分级,国内学者在引入循证医学理念的同时,结合中医辨证施治个体化的特点,对中医证据评价体系作了许多突破性的尝试^[21]。2007 年刘建平结合中医药临床特点和研究现状,提出针对中医药的临床证据分级方法,为指南制订和临床决策提供了重要参考和依据(表 2)。2009 年李敬华等指出中医临床与现代医学的一个明显区别是其坚持辨证论治,其以“辨证论治”作为评价中医临床治疗文献质量等级的第一要素,并将辨证论治的研究列为 I 级证据,提出了中医治疗文献证据等级体系(表 3)^[22]。何庆勇等在 2010 年基于大量文献分析、系统回顾国内外证据以及 2 轮专家咨询,制订了中医特色的临床证据分级与评分体系(表 4)^[23]。2016 年,袁敬柏在对当前循证诊疗指南及中医诊疗指南有关证据评价方法进行系统分析的基础上,结合中医理论与临床特色,提出了中医临床诊疗证据评价方法的建议(表 5)及推荐级别(表 6)^[24]。2016 李慧等结合中医(中西医结合)的实践特点介绍了 GRADE 系统的应用方法,以及将其引入中医药领域面临的挑战与发展,其认为 GRADE 在制订中医和中西医结合指南的应用中将逐渐成熟和普及^[25]。本文作者之一(陈耀龙)目前承担的国家自然科学基金资助的《中西医结合领域在线证据质量分级系统的研发》项目,基于 GRADE 分级,结合中医药特色,探索如何制订中西医结合领域的证据质量分级方法。

另一方面,除非能找到一套适用于评价中医药疗效和安全性的公认的方法和体系^[26],否则,从减少偏倚的途径入手,采用严谨的设计、规范的实施和

表 2 中医临床研究证据分级参考建议

证据级别	分级依据
I a	由 RCT、队列研究、病例对照研究、案例系列这 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 RCT
II a	半随机对照试验或队列研究
II b	病例对照研究
III a	历史性对照的病案系列
III b	自身前后对照的病案系列
IV	长期在临床上广泛运用的病案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案报告和史料记载的疗法

注:表中的 RCT,包括随机双盲安慰剂的 RCT、实用型的 RCT、等待名单的 RCT、单个病例的 RCT、集团 RCT 及技能型 RCT 等各种随机研究的方法。

表 3 中医治疗文献证据等级体系

分级	依据
I 级	辨证论治的研究
I a 级	辨证论治的 RCT,设计及完成质量良好
I b 级	辨证论治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

表 4 中医临床证据与评分体系

证据级别	分级证据	评分
I 类	I a	目前仍在使用的四大经典医籍 每项 10 分
		目前仍在使用的国家标准及行业制定的标准 每项 9 分
	I b	多个 RCT 的系统评价 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 类	II a	单个正确设计的 RCT 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 类		目前仍在使用的其他古代经典医籍 每项 5 分
IV 类		无对照的病例观察 每项 2 分
		医案医话 每项 2 分

表 5 理论传承证据分类、分级与评价

类别	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)名中医的医案能体现	近现代(民国-当代)名中医的医案能体现
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

表 6 理论传承证据用于诊断与干预措施推荐分级

推荐级别	推荐依据
A	全部 4 类,或 I 类证据加 II ~ IV 类证据的 2 类,或 II ~ IV 类中的 3 类,具有一致结论,今后研究中不可能形成否定性证据
B	II ~ IV 类中的 2 类,具有比较一致结论,今后研究中形成否定性证据可能性小
C	II ~ IV 类证据中的 1 类
D	无证据支持或观点分歧

充分的报告来生产中医药领域的证据,仍然不失为目前让中医药走出困境,走向世界的一条重要途径^[27]。通过研究中医药自身理论及证据特点,在突出中医特色的同时,借鉴循证医学方法学和国际证据分级及推荐强度标准,形成符合中医药特色的证据评价和推荐强度标准,从而使中医领域临床实践指南的制订既体现并适用于中医特点,又符合国际标准和循证医学的基本原则^[21]。同时,通过系统评价这个枢纽,建立证据体,结合 GRADE 分级系统,梳理出当前中医药研究的现状,一方面为中医药指南制订提供依据,架起从研究到实践的桥梁;另一方面为中医药临床研究提供选题,开展更多高质量研究(不管是随机对照试验还是观察性研究),从根本上解决中医药领域缺乏高质量证据的问题。

循证医学诞生 25 年来,发展突飞猛进^[28-29],中医药领域如何充分利用其方法学优势,从证据的生产、评价、报告和推荐各个环节,为我所用,是下一步需要努力的方向,也是中医药未来发展的重要道路之一。

参考文献

- [1] 政府工作报告. http://www.gov.cn/premier/2017-03/16/content_5177940.htm.
- [2] 商洪才,李幼平,张伯礼,等. 中医药临床疗效个体化评价方法初探——循证目标成就量表法的提出[J]. 中国循证医学杂志, 2007,7(7):537-541.
- [3] 胡学军,商洪才,张伯礼,等. 生存质量及其量表在中医药疗效评价中的应用[J]. 天津中医药,2004,21(3):191-193.
- [4] 王永炎,黄启福,鲁兆麟,等. 中医药学学科方向的变革与创新[J]. 北京中医药大学学报,2011,34(1):5-11.
- [5] Evans I, Thornton H, Chalmers I. Testing Treatments: Better Research for Better Healthcare, 1st edn[M]. London: British Library, 2006.
- [6] 王永炎,刘保延,谢雁鸣. 应用循证医学方法构建中医临床评价体系[J]. 中国中医基础医学杂志,2003,9(3):17-23.
- [7] 李华伟. 从循证医学的角度看中医发展的机遇与挑战[J]. 中医研究,2005,18(11):1-3.
- [8] 陈可冀,宋军. 循证医学的提出对中西医结合的启发[J]. 中国中西医结合杂志,1999,19(11):643-644.
- [9] Sackett D L. Bias in analytic research[J]. Journal of Chronic Diseases, 1979,32(1-2):51-63.
- [10] Gordon H. Guyatt, Andrew D. Oxman, Gunn Vist, 等. GRADE 指南: IV. 证据质量分级——研究的局限性(偏倚风险)[J]. 中国循证医学杂志,2011,11(4):456-463.
- [11] Guyatt G H, Oxman A D, Vist G, et al. GRADE guidelines: 4. Rating the quality of evidence—study limitations (risk of bias)[J]. Chinese Journal of Evidence-Based Medicine, 2011,64(4):407-415.
- [12] Chalmers TC, Celano P, Sacks HS, et al. Bias in treatment assign-

ment in controlled clinical trials[J]. New England Journal of Medicine, 1983,309(22):1359-1361.

- [13] Schulz KF, Chalmers I, Hayes RJ, et al. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials[J]. Jama the Journal of the American Medical Association, 1995,273(5):408-412.
- [14] 张俊华,商洪才,戴国华,等. 谈中医药临床治疗性研究的方案设计[C]//中国. 天津国际针灸学术交流会, 2006.
- [15] Chan AW, Hrobjartsson A, Haahr MT, Gotzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles[J]. Jama the Journal of the American Medical Association, 2004,291(20):2457-2465.
- [16] Ghersi D, Clarke M, Berlin J, et al. Reporting the findings of clinical trials: a discussion paper[J]. Bulletin of the World Health Organization, 2008,86(6):492-493.
- [17] Faber J, Fonseca L M. How sample size influences research outcomes[J]. Dental Press Journal of Orthodontics, 2014,19(4):27-29.
- [18] 陈昊,王艳,陈耀龙,等. 从 GRADE 系统方法学看中医药临床证据体系的构建[J]. 中华中医药杂志,2017,32(1):25-28.
- [19] Guyatt G H, Oxman A D, Kunz R, 等. GRADE: 从证据到推荐[J]. 中国循证医学杂志,2009,9(3):257-259.
- [20] 陈耀龙,李幼平,杜亮,等. 医学研究中证据分级和推荐强度的演进[J]. 中国循证医学杂志,2008,8(2):127-133.
- [21] 刘欣娟,郭海玲,田润溪,等. 中医临床实践指南制定过程中证据分级与推荐强度的研究进展[J]. 中国护理管理,2016,16(12):1613-1618.
- [22] 李敬华. 中医治疗文献质量评价方法及应用研究[D]. 北京: 中国中医科学院, 2009.
- [23] 何庆勇,王阶,王师菡,等. 中医临床证据分级与评分体系研究[C]. //中华中医药学会心病分会第十二次学术年会论文, 2010:8-12.
- [24] 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.
- [25] 李慧,陈耀龙,谢秀丽,等. 中医(中西医结合)临床实践指南制订方法-证据质量分级[J]. 中华中医药杂志,2016,31(7):2652-2654.
- [26] 商洪才,李幼平,张伯礼. 中医药临床疗效评价实践中的几点思考[J]. 天津中医药,2007,24(4):275-277.
- [27] 商洪才,张伯礼,王永炎. 中医药临床研究待“突破”[J]. 天津中医药,2013,30(1):4-5.
- [28] Djulbegovic B, Guyatt G H. Progress in evidence-based medicine: a quarter century on[J]. Lancet, 2017.
- [29] 陈耀龙,杨克虎,田金徽. 循证实践指南的制定: 国际经验与中国实践[J]. 兰州大学学报: 医学版, 2016,42(1):29-35.

(2017-05-10 收稿 责任编辑:徐颖)