

“病证结合”多主要终点评价法在中医药临床研究中的应用

胡晶¹ 刘卫红¹ 张会娜¹ 李博¹ 张晓雨² 商洪才²

(1 首都医科大学附属北京中医医院/北京市中医研究所流行病学室,北京,100010; 2 北京中医药大学东直门医院教育部中医内科学重点实验室,北京,100700)

摘要 中医药治疗效果需要科学的评价方法彰显。当前单指标评价法不足以全面体现中医药治疗优势,“病证结合”的多主要终点评价法为全面客观显示中医药疗效提供可能。尽管近年来多主要终点评价在统计学方法领域上有较大发展,但在其方法选择、样本量计算、统计软件使用等方面仍有很多挑战。因此本文对几种多主要终点评价方法进行介绍,以期指导中医药临床研究者正确选择和使用多主要终点评价法,为全面评价中医药特色与优势提供方法学支撑。

关键词 病证结合;多主要终点;随机对照试验;中医药

Application of Evaluation of Multiple Primary Endpoints Based on Combination of Diseases and Symptoms in TCM Clinical Trials

Hu Jing¹, Liu Weihong¹, Zhang Huina¹, Li Bo¹, Zhang Xiaoyu², Shang Hongcai²

(1 Department of Epidemiology, Beijing Hospital of Traditional Chinese Medicine, Capital Medical University/Beijing Institute of Traditional Chinese Medicine, Beijing 100010, China; 2 Key Laboratory for Internal Chinese Medicine of Ministry of Education, Dongzhimen Hospital, Beijing University of Chinese Medicine, Beijing 100700, China)

Abstract The development of a scientific evaluation tool is a prerequisite to reflect the efficacy of traditional Chinese medicine (TCM). At present, separate analysis of each outcome is the most commonly used method in randomized controlled trials (RCTs) of TCM, while this method has been reported to have some limitations. Evaluation of multiple primary endpoints which include diseases and symptoms outcomes can reflect the effectiveness of TCM comprehensively and objectively. In recent years, despite the great development of multiple primary endpoints in the field of statistical methods, there are still some challenges in choosing methods, calculating sample size and employing statistical software. In this article, we introduced several multiple primary endpoints evaluation methods, in the expectation of appropriate use of those methods by TCM researchers as well as methodological support for comprehensive assessment of the characteristics and advantages of TCM.

Key Words Combination of diseases and syndromes; Multiple primary endpoints; Randomized controlled trials; Traditional Chinese medicine

中图分类号:R241 文献标识码:A doi:10.3969/j.issn.1673-7202.2017.06.002

随机对照试验(Randomized Controlled Trials, RCTs)是目前被国际上公认为评价临床干预措施有效性的金标准。自1983年第1篇中医药RCT报告以来^[1],RCTs方法已广泛运用于评价中医药临床疗效,但随着RCTs在中医药领域的应用,其局限性也逐渐暴露出来,而缺乏能达成中西医共识的疗效评价方法更是成为制约中医药国际化进程的重要瓶颈。近年来围绕这一领域的研究已成为各方学者关注的热点问题^[2-4]。

1 单指标评价法难以从整体反映中医药的治疗优势
基于统计学的假设检验理论,ICH E9^[5]建议临

床试验的主要终点指标最好只设一个,因此,现阶段进行的中医药RCTs等临床试验设计中,疗效评价通常是将西医客观指标(例如病理、生化指标等)作为主要疗效指标,在次要指标中添加一些中医特色指标(例如中医证候、患者报告结局等),分析时对各指标单独评价并得出结论。

但由于中医和西医在临床治疗理论、药物研发模式等方面存在差异,中医更注重患者症状的改善,而西医更强调病理、生化等实验室指标的改变。如果采用现行的疗效评价方法,很可能出现在主要疗效指标上中药效果接近甚至劣于西药、而在次要

指标上优于西药的结果。

另外,中医药理论强调整体观,传统中医药就治病目的而言是多维度干预,既控制病情活动,又改善患者的生命质量;既改善患者主观感受的不适症状,又注重客观指标的恢复。由于临床试验中次要指标仅用于探索或参考目的,对疗效评价作用有限^[5]。因此,中医药 RCTs 若仅用西医疗效标准作为主要指标,无法全面客观地反映中医药的特色治疗优势,应把中医特色指标与西医常规指标放到同等位置对待,即将对“病”的疗效评价结果和对“证”的评价结果相结合进行分析。

2 “病证结合”的多主要终点为整体、客观评价中医药疗效提供可能

目前各方学者广泛认可的观点是,完善的中医药临床疗效评价指标体系应包括^[6-8]:1)对于疾病的常规西医疗效评价指标;2)反映中医证候的指标;3)生命质量评价指标。

鉴于此,本文提出体现中医特色的“病证结合”的多终点疗效评价法,即将以上3种不同属性的测量指标共同作为主要终点进行分析,建立既能让现代医学界认可、又能体现中医药特色治疗优势的评价方法。

尽管近年来多主要终点评价在统计技术上有较大发展,但至今在中医药领域仍然应用较少,可能原因:1)目前国内对几种多主要终点评价方法的理论和实际应用介绍较少,大多数学者还不熟悉该评价方法;2)多主要终点评价方法大多对统计理论和软件水平要求较高,研究者不易实现;3)多主要终点评价虽然方法较多,但目前尚没有公认的方法,几种方法各有利弊(表1),且各有其适用范围和应用条件^[9]。使得研究者在实际应用时无从选择。因此本文对几种多主要终点评价方法进行介绍,以期指导中医药临床研究者正确选择和使用多主要终点评价法,

为全面评价中医药特色治疗优势提供方法学支持。

3 多主要终点评价法的介绍

3.1 检验水准校正法

检验水准校正法(Alpha-adjustment Procedures)是对多主要终点中每个终点的检验水准进行调整的一类方法^[10],包括 Bonferroni 校正、Simes 检验、Hochberg 法等,其中 Bonferroni 法是最为经典也是最简单的方法^[11],其公式为 $\alpha^* = \alpha/K$ (K 为终点个数)。Hochberg 法的原理是先将各个终点检验的 $P(i)$ 值进行排序, $P(i) (K) \geq P(K-1) \geq \dots \geq P(1)$ (K 为终点个数),对应的检验水准 $\alpha(i)$ 分别为 $\alpha, \alpha/2, \dots, \alpha/k$,随后按 $P(i)$ 值由大到小的顺序逐步检验,直到出现第一个 $P(i) < \alpha(i)$ 的结论即停止,之后的所有检验均认为有统计学意义。

检验水准校正法优势:1)可以对单个终点的疗效进行分析,控制试验总 I 类错误率(Family-wise Type I Error Rate, FWER)膨胀。2)当干预措施对一个结局指标效应较强时具有较高统计检验效能^[12]。

检验水准校正法不足:1)无法对多个终点疗效进行整体评价,另外当多终点间结果不一甚至相反时,无法得出确切结论^[13]。2)当多终点间相关程度较高时结论过于保守。

3.2 多变量检验

多变量检验(Multivariate Test)是对多个终点的联合分布进行一次假设检验,对组间差别进行推断的方法。包括 Hotelling's T^2 检验、针对多组比较的多变量方差分析(MANOVA)等^[14]。多变量检验的计算比较繁琐,但用 SPSS 或 SAS 软件计算则非常简单。在大多数情况下,多变量检验结论与对 K 个多终点进行 K 次单变量假设检验的结论是一致的,即多变量假设检验拒绝 H_0 , K 次单变量假设检验至少有一次拒绝 H_0 。但单变量假设检验不能代替多变量假设检验,主要理由:1) K 次单变量假设检验增加假阳性错误的概率。2)单变量假设检验只说明某一变量在数轴分布上的组间

表1 多主要终点评价方法的比较

评价方法	常用模型	优点	缺点
检验水准校正法	Bonferroni 法、Simes 法、Hochberg 法	控制 FWER 膨胀;容易操作	缺乏多终点疗效整体评价;多终点间强相关时结论保守
多变量检验	Hotelling's T^2 检验、MANOVA	控制 FWER 膨胀;考虑多终点间相关性	缺乏多终点疗效整体评价;终点概率分布假设不易满足
多水平模型	多元混合效应模型、潜变量模型、贝叶斯模型	能分析不同数据类型;考虑多终点间相关性	模型假设无法验证;潜变量等不易理解
综合评价法	层次分析法、TOPSIS 法、模糊数学法	对多终点疗效整体评价;统计效能较高	指标权重不易确定;不同模型结论差异较大;综合评价指标不易解释
全局检验法	O'Brien 非参数法和最小二乘法	对多终点疗效整体评价;考虑多终点间相关性	缺乏单个终点疗效评价

差别,不能反映多个变量在平面或空间上的差别。有可能会出现 K 个多终点在进行单变量假设检验时均无统计学意义,但多变量假设检验却检验出有差异。

多变量检验优势:1)避免多次单变量检验增加假阳性错误的概率;2)在校正过程中充分考虑了多终点间的相关结构。

多变量检验不足:1)缺乏对多个终点疗效整体评价的结果;2)对多终点联合概率分布的数学假设前提要求较高^[15]。

3.3 多层统计分析模型 多层统计分析模型(Hierarchical Model)是同时对多个终点进行分析,包括多元混合效应模型、潜变量模型、贝叶斯模型等^[16]。多层统计分析模型利用添加一个虚拟 1 水平进行分析,将几个终点作为 1 水平上的观察单位,将受试者作为 2 水平单位。HLM 和 MLwin 是专门用于多层模型的分析软件。SAS 软件的 PROC MIXED 程序也广泛地应用于连续型结局变量的多层模型分析,PROC GLIMMIX 和 PROC NLMIXED 程序可用于分类结局变量和计数数据的多层模型分析。

多水平模型优势:1)能将不同类型的数据联合进行分析,包括连续变量和分类变量;2)考虑了多终点间相关性,同时可以有效处理缺失数据^[17]。

多水平模型不足:1)无法证实模型所采用假设的准确性,尤其是关于多终点间相互关系的假设;2)潜变量或超参数等含义不易被临床医生所理解。

3.4 综合评价法 综合评价(Comprehensive Evaluation Method)是对一个复杂系统多个指标进行总评价的特殊方法。目前已经逐渐应用于中医药临床研究中,例如层次分析法^[18]和 TOPSIS 法^[19]等。综合评价法的实施过程类似,包括筛选评价指标、根据指标重要性赋予权重、采用相应方法建立综合评价模型并进行指标合并等。

综合评价法的优势:1)可以将西医常规指标、中医证候、患者报告结局等多个指标组合成一个综合指标进行分析,对综合指标进行检验,适应中医疗效多维度的特点;2)统计检验效能较高。

综合评价法的不足:1)指标权重的确定或过于依赖主观判断,或把指标重要性同等化,造成权重系数不合理,导致最终结果的不确定性^[20];2)目前综合评价法虽多,但没有完美的方法,多方法间结论存在较大差异^[21];3)对综合评价指标的解释存在一定困难。

3.5 全局检验法 全局检验法(Global Statistical

Test, GST)是将多个终点综合为一个检验统计量进行分析的方法^[22]。在 GST 中,O'Brien 法是各方学者较为认可、应用较为广泛的一种方法^[23],可采用一般最小二乘法、广义最小二乘法、非参数法进行分析。其中非参数法可以针对非正态分布的多终点,适用范围较广。全局治疗效应(Global Treatment Effect, GTE)是 GST 中定量反映试验组间疗效差异的指标^[22]。GTE 的数值固定,不随终点测量方法的改变而变化,取值范围为“-1~1”,“0”代表试验组和对照组间疗效无差异,“1”代表试验组完全优于对照组,“-1”代表对照组完全优于试验组。GTE 可作为结局指标的效应值,采用 Splus 软件进行样本含量的估算。

全局检验法的优势:1)可以检验多个终点的全局治疗效应,将 GST 应用于中医药 RCTs 最大的优势是可以凸显中医药的整体治疗优势;2)在分析时充分考虑多终点间的相关结构;3)当多终点间结果不一致甚至相反时更利于结果的解释。

全局检验法的不足:只能发现试验组间整体疗效是否有差异,无法对单个终点的疗效得出结论^[24]。

针对多主要终点指标的样本量计算问题,有学者建议在样本量计算时可对每个主要终点分别计算,然后取最大值作为临床试验所需的样本量^[25]。Sozu 等探讨了用 SAS 软件对同时包含连续变量和二分类变量的多主要终点进行样本含量估计^[26]。杨卫娇对多终点指标的两阶段适应性设计的样本量再估计进行了探讨^[27]。

4 结语

中医药治疗效果需要科学的评价方法彰显,在中医药临床研究中,如何科学、客观的选择结局指标对于疗效评价至关重要。作为临床试验的统计学指导规范,虽然 ICH E9^[5]建议主要终点指标最好只设一个,但这种考量主要是基于统计学的假设检验理论,由于中医“整体观”的理论特点,在中医药临床试验中,一个主要终点难以全面体现中医药治疗优势,此时通常需要通过多主要终点评价来解决,其优势是可以将对于疾病的常规西医疗效评价指标以及反映中医证候的指标等共同作为主要终点进行分析,能全面、客观的体现中医药特色治疗优势。尽管多主要终点比单个终点在中医药疗效评价中体现出更多的优势,其随之而来的一系列统计学问题却不容忽视,统计的 2 类错误都需要进行控制。本文对几种多主要终点评价方法进行介绍,以期指导中医

药临床研究者了解和使用多主要终点评价法,为全面评价中医药特色治疗优势提供方法学支持。

参考文献

- [1] 单平,毛如宝,徐济民,等. 黄杨宁治疗冠心病一用双盲法观察110例临床疗效分析[J]. 中医杂志,1983,24(5):37-40.
- [2] Li Zhang, Junhua Zhang, Jing Chen, et al. Clinical research of traditional Chinese medicine needs to develop its own system of core outcome sets [J]. Evid Based Complement Alternat Med, 2013, 2013 (2013):202703.
- [3] 潘万旗,邓素玲,杨英豪,等. 关于中医药标准化的思考[J]. 中医学报,2015,30(6):795-797.
- [4] Pritzker S, Hui KK. Building an evidence-base for TCM and integrative east-west medicine: a review of recent developments in innovative research design [J]. J Tradit Complement Med, 2012, 2(3):158-163.
- [5] ICH Topic E9 Statistical Principles for Clinical Trials (CPMP/ICH/363/96), www.emea.europa.eu, 1998. Accessed 8th April 2013.
- [6] 王贤良,毛静远,侯雅竹. 病证结合、系统分段、多维指标中医临床效应评价方法建立初探[J]. 中国中西医结合杂志,2013,33(2):270-273.
- [7] 高凡珠,谢雁鸣,王永炎. 中医复杂干预与疗效综合评价[J]. 中国中医基础医学杂志,2010,16(6):527-529.
- [8] 李建生,余学庆. 病证结合模式下疗效评价指标体系建立的思考[J]. 中华中医药杂志,2011,26(8):1666-1670.
- [9] Huang P, Goetz CG, Woolson RF, et al. Using global statistical tests in long-term Parkinson's disease clinical trials [J]. Mov Disord, 2009, 24(12):1732-1739.
- [10] Sankoh AJ, SrdtrB, Huque MF. Efficacy endpoint selection and multiplicity adjustment methods in clinical trials with inherent multiple endpoint issues [J]. Stat Med, 2003, 22(20):3133-3150.
- [11] Leon AC, Heo M. A comparison of multiplicity adjustment strategies for correlated binary endpoints [J]. J Biopharm Stat, 2005, 15(5):839-855.
- [12] Dmitrienko A, D'Agostino RB Sr, Huque MF. Key multiplicity issues in clinical drug development [J]. Stat Med, 2013, 32(7):1079-1111.
- [13] Alosch M, Bretz F, Huque M. Advanced multiplicity adjustment methods in clinical trials [J]. Stat Med, 2014, 33(4):693-713.
- [14] Olson CL. Comparative robustness of six tests in multivariate analysis of variance [J]. J Am Stat Assoc, 1974, 69(348):894-908.
- [15] Johnson JL, Muller KE, Slaughter JC, et al. POWERLIB: SAS/IML Software for Computing Power in Multivariate Linear Models [J]. J Stat Softw, 2009, 30(5):1-27.
- [16] Leiby BE, Ten Have TR, Lynch KG, et al. Bayesian multivariate growth curve latent class models for mixed outcomes [J]. Stat Med, 2014, 33(20):3434-3452.
- [17] Teixeira-Pinto A, Mauri L. Statistical analysis of noncommensurate multiple outcomes [J]. Circ Cardiovasc Qual Outcomes, 2011, 4(6):650-656.
- [18] 李景. 糖尿病肾病中医药疗效综合评价体系研究 [D]. 北京:北京中医药大学, 2012.
- [19] 陈磊,梁伟雄,吕志平. 生脉胶囊治疗慢性充血性心力衰竭临床疗效的 TOPSIS 法综合评价 [J]. 南方医科大学学报, 2010, 30(4):820-822.
- [20] Freemantle N, Calvert M, Wood J, et al. Composite endpoints in randomized trials: greater precision but with greater uncertainty [J]. JAMA, 2003, 289(19):2554-2559.
- [21] Rauch G, Jahn-Eimermacher A, Brannath W, et al. Opportunities and challenges of combined effect measures based on prioritized outcomes [J]. Stat Med, 2014, 33(7):1104-1120.
- [22] Huang P, Woolson RF, O'Brien PC. A rank-based sample size method for multiple outcomes in clinical trials [J]. Stat Med, 2008, 27(16):3084-3104.
- [23] O'Brien PC. Procedures for comparing samples with multiple endpoints [J]. Biometrics, 1984, 40(4):1079-1089.
- [24] Baraniuk S, Seay R, Sinha AK, et al. Comparison of the global statistical test and composite outcome for secondary analyses of multiple coronary heart disease outcomes [J]. Prog Cardiovasc Dis, 2012, 54(4):357-361.
- [25] Cook RJ, Farewell VT. Guidelines for monitoring efficacy and toxicity responses in clinical trials [J]. Biometrics, 1994, 50(4):1146-1152.
- [26] Sozu T, Sugimoto T, Hamasaki T. Sample size determination in clinical trials with multiple co-primary endpoints including mixed continuous and binary variables [J]. Biom J, 2012, 54(5):716-729.
- [27] 杨卫娇. 多终点指标下两阶段适应性设计的样本量再估计 [D]. 上海:华东师范大学, 2011.

(2017-05-10 收稿 责任编辑:徐颖)