

DOI: 10.13288/j.11-2166/r.2016.11.014

证候 · 诊断

缺血性中风证候要素诊断量表不完善金标准验证结果的校正

刘强¹, 王肖南², 王少卿³, 高颖^{3*}, 周晓华⁴

(1. 中国中医药国际合作中心, 北京市朝阳区小营路 19 号, 100101; 2. 中国人民大学统计学院; 3. 北京中医药大学东直门医院; 4. 美国华盛顿大学公共卫生学院)

摘要 目的 探讨缺乏辨证金标准的缺血性中风证候要素诊断量表灵敏度与特异度的校正方法。方法 收集全国 9 个省市 10 家医院的缺血性中风患者 204 例, 每例患者分别采用《缺血性中风证候要素诊断量表》和《中风病辨证诊断标准》完成证候要素判断, 同时由 3 位中医专家对患者证候要素进行独立的临床经验辨证, 2 位及以上专家的共同意见作为证候要素的专家经验判断结果, 完成灵敏度与特异度的校正。构造似然函数时, 假设专家经验综合辨证方法独立于量表辨证, 而两种量表辨证结果彼此不独立。采用 MCMC 方法, 根据贝叶斯参数估计原理, 利用抽样方法进行后验密度估计。结果 与《中风病辨证诊断标准》相比, 《缺血性中风证候要素诊断量表》内风证、血瘀证的敏感度有显著提升, 痰湿证和气虚证方面敏感度显著降低 ($P < 0.05$)。量表和专家经验综合判断都体现出较好的特异度。结论 采用多诊断方法用于同一总体时估计灵敏度与特异度的校正方法, 可应用于缺乏金标准的缺血性中风证候要素量表灵敏度与特异度做出校正后的估计。

关键词 缺血性中风; 诊断标准; 诊断量表; 证候要素

证候诊断量表作为一种诊断工具其验证应该通过诊断试验来进行, 然而诊断学试验中所依赖的金标准在证候研究中往往无从获得。目前证候量表的研制和评价往往以临床医师或专家判断作为参考标准, 但临床医师或专家判断个体和中心差异较大, 可靠性难以保证^[1]。

中风病是证候诊断标准化探索较早的领域。1994 年在国家八五科技攻关项目支持下, 国家中医药管理局脑病急症科项目组制定并发布了《中风病辨证诊断标准》^[2] (以下简称“1994 标准”), 这一量化诊断标准在后续的中风病临床研究中被广泛使用。在此基础上, 国家重点基础研究发展计划 (“973”计划) “缺血性中风病证结合的诊断标准与疗效评价体系研究”课题组进一步对证候诊断量表的理论基础和方法学进行了深入研究, 并于 2011 年制定了《缺血性中风证候要素诊断量表》

(以下简称“2011 量表”)^[3]。量表验证结果表明, 缺血性中风证候要素诊断量表在特异性、敏感性和判断准确率上明显优于 1994 标准^[4]。2011 量表在验证过程中采用了 3 个专家同时独立判断证候, 取其中 2 名及以上的一致意见作为参考标准来计算诊断量表的敏感度和特异度。这一方法在一定程度上避免了证候判断的个性化差异, 但 3 个专家同时独立进行证候判断在临床上操作较为困难, 同时这种“3+1”的验证模式还是无法解决证候缺乏完善金标准的问题。

诊断学试验中, 当研究者采用不完善标准作为“金标准”时, 可能低估或者高估被测诊断方法的敏感度和特异度, 这种情况产生的偏倚被称为不完善金标准偏倚 (imperfect gold standard bias)^[5]。为了进一步探讨 2011 量表的临床验证, 我们参考多诊断方法同时用于来自一个总体的随机样本时估计灵敏度与特异度的校正方法^[5-6], 针对证候量表的不完善金标准情况下 3 个诊断试验评价方法展开如下探讨。

基金项目: 国家科技重大专项“重大新药创制”项目 (2009ZX09502-028)

* 通讯作者: gaoying973@126.com, (010) 84013209

1 资料与方法

1.1 临床资料

本研究使用的临床数据来自于 9 个省市 10 家医院 2008 年 2 - 11 月的 204 例缺血性中风患者, 其中女性 73 例, 男性 131 例; 年龄 41 ~ 84 岁, 平均年龄 (65.2 ± 10.5) 岁。研究中进行临床经验判断的专家要求从事中医脑病临床工作经验 5 年以上、有主治医师以上职称者, 10 家医院共计 57 位专家参与。研究病例及专家均来自北京中医药大学东直门医院、广东省中医院、安徽中医学院第一附属医院、天津中医药大学第二附属医院、河南中医学院第一附属医院、辽宁省血栓病中西医结合医疗中心、湖南省中医药研究院附属医院、浙江省中医院、南京市中医院、佛山市中医院。

1.2 不完善金标准情况下的诊断学评价

每例患者分别采用 1994 标准和 2011 量表完成证候要素判断, 同时由 3 位中医专家对患者证候要素进行独立的临床经验辨证, 2 位及以上专家的共同意见作为证候要素的专家经验判断结果。

2011 量表验证时进行的三种不同证候判断方法, 由于量表研发时直接参考了 1994 标准中的条目, 因此, 两者不服从条件独立假设。但同时专家经验辨证及其决策模式独立于量表判断进行, 因此, 在本研究中采用 2 种诊断方法之间不独立, 而第 3 种诊断方法独立的假设, 对敏感度和特异度的估计方法进行校正。基于本研究针对的问题, 由于参数个数大于自由度, 因此, 相对于经典统计方法, 贝叶斯方法更适合于分析本研究中的数据。贝叶斯方法的基本思想是: 假定对所研究的对象在抽样前已有一定的认识, 常用先验分布来描述这种认识, 然后基于抽取的样本再对先验认识作修正, 得到后验分布, 而各种统计推断都基于后验分布进行。在本研究中, 后验分布没有显著的表达式, 因此, 我们选用了 markov chain monte carlo (MCMC) 方法来生成数据逼近后验分布。

1.3 数据分析方法

在本研究中, 将 1994 标准、2011 量表和专家经验辨证视为针对同一缺血性中风人群的 3 种不同证候要素诊断方法。为实现多个试验同时用于同一整体的敏感度和特异度的校正, 采用基于贝叶斯估计的 MCMC 算法完成量表和专家经验综合判断的敏感度、特异度估计。在数据计算过程中, 选取 beta 分布作为参数的先验分布^[7-8]。为了确定 beta

分布的超参数 α 和 β 的值, 灵敏度和特异度的众数选为 0.70, 95% 先验概率区间设定为 0.50 ~ 1.00。

数据分析采用 Winbugs 软件(版本号: 1.4.3)完成。将 MCMC 抽样迭代 100 000 次, 采用 10 000 次迭代退火(Burn-in)后的迭代结果作为 markov 链进行参数估计, 采用后验分布的中位数做为参数的估计值, 得到参数的后验估计及 95% 的贝叶斯可信区间(CI95)。

2 结果

本研究在缺乏完善金标准的情况下, 对两种量表和专家经验综合判断, 对 6 种缺血性中风证候要素的敏感度和特异度进行了估计, 具体见表 1。数据分析结果显示, 与 1994 标准相比, 2011 量表内风证、血瘀证的敏感度有显著提升, 但痰湿证和气虚证敏感度较专家经验判断结果显著偏低, 量表和专家经验综合判断都体现出较好的特异度。该结果与将专家综合判断做为金标准的评价结果有着一致的趋势。在该估计结果中, 专家经验综合判断表现出比量表更好的敏感度和特异度。

3 讨论

在缺乏中医证候完善金标准的情况下, 多种证候诊断方法用于同一总体时估计灵敏度与特异度的校正, 可以对缺乏完善金标准的中医证候要素量表灵敏度与特异度做出校正后的估计。校正后的分析结果显示, 专家经验判断结果与量表分析结果相比较优, 而 2011 量表与 1994 标准相比, 内风、血瘀证的敏感度有显著提升。

辨证论治是中医在临床实践中的重要特征, 证候判断是基于对患者的体格检查和望、闻、问、切获得的客观信息, 但辨证的决策过程主要依赖医师的个人经验, 具有一定的主观性。辨证论治的这一特征决定了证候诊断不可能存在完善而被公认的“金标准”, 同时经验辨证本身的准确性也受到质疑。基于计量诊断思想研制的中医证候诊断量表, 广泛应用于中医临床科研, 其验证过程中辨证敏感度和特异度的计算往往只能参考经验辨证结果。

2011 量表在验证中采用了 3 位专家同时独立判断证候, 取其中 2 位及以上的一致意见作为参考标准的方式, 相较于单个医师的经验辨证结果有着更好的可靠性, 但无法给出数据支持。本研究将缺乏完善金标准情况下多种诊断方法用于同一总体时估计灵敏度与特异度的校正方法, 应用于缺血性中

表 1 缺血性中风证候要素诊断量表敏感度、特异度校正结果

证候要素	项目	诊断方法	专家金标准法		校正后结果	
			估计值	CI95	估计值	CI95
内风证	敏感度	专家综合判断			0.813	0.742 ~ 0.880
		1994 标准	0.673	0.587 ~ 0.759	0.687*	0.605 ~ 0.765
		2011 量表	0.894	0.837 ~ 0.951	0.884 [△]	0.815 ~ 0.938
	特异度	专家综合判断			0.922	0.831 ~ 0.981
		1994 标准	0.659	0.562 ~ 0.756	0.776	0.652 ~ 0.885
		2011 量表	0.692	0.597 ~ 0.787	0.875	0.739 ~ 0.968
内火证	敏感度	专家综合判断			0.840	0.764 ~ 0.903
		1994 标准	0.810	0.723 ~ 0.897	0.827	0.728 ~ 0.908
		2011 量表	0.734	0.637 ~ 0.831	0.751	0.646 ~ 0.844
	特异度	专家综合判断			0.937	0.870 ~ 0.983
		1994 标准	0.888	0.833 ~ 0.943	0.932	0.864 ~ 0.979
		2011 量表	0.872	0.813 ~ 0.931	0.905	0.836 ~ 0.958
血瘀证	敏感度	专家综合判断			0.877	0.816 ~ 0.927
		1994 标准	0.408	0.322 ~ 0.494	0.479*	0.386 ~ 0.585
		2011 量表	0.744	0.667 ~ 0.821	0.818 [△]	0.717 ~ 0.917
	特异度	专家综合判断			0.816	0.678 ~ 0.947
		1994 标准	0.949	0.900 ~ 0.998	0.954	0.886 ~ 0.990
		2011 量表	0.861	0.785 ~ 0.937	0.909	0.814 ~ 0.972
痰湿证	敏感度	专家综合判断			0.871	0.813 ~ 0.920
		1994 标准	0.738	0.665 ~ 0.811	0.765	0.680 ~ 0.858
		2011 量表	0.716	0.642 ~ 0.790	0.743*	0.657 ~ 0.837
	特异度	专家综合判断			0.827	0.652 ~ 0.957
		1994 标准	0.794	0.694 ~ 0.894	0.892	0.777 ~ 0.970
		2011 量表	0.810	0.713 ~ 0.907	0.903	0.790 ~ 0.975
气虚证	敏感度	专家综合判断			0.837	0.753 ~ 0.905
		1994 标准	0.388	0.271 ~ 0.505	0.620*	0.472 ~ 0.761
		2011 量表	0.493	0.373 ~ 0.613	0.751*	0.589 ~ 0.878
	特异度	专家综合判断			0.833	0.760 ~ 0.909
		1994 标准	0.939	0.887 ~ 0.976	0.939	0.887 ~ 0.976
		2011 量表	0.965	0.918 ~ 0.996	0.965	0.918 ~ 0.996
阴虚证	敏感度	专家综合判断			0.825	0.734 ~ 0.898
		1994 标准	0.630	0.490 ~ 0.770	0.759	0.619 ~ 0.883
		2011 量表	0.587	0.445 ~ 0.729	0.717	0.581 ~ 0.844
	特异度	专家综合判断			0.925	0.868 ~ 0.973
		1994 标准	0.962	0.933 ~ 0.991	0.966	0.924 ~ 0.992
		2011 量表	0.930	0.890 ~ 0.970	0.934	0.885 ~ 0.969

注: 1994 标准 《中风病辨证诊断标准》; 2011 量表 《缺血性中风证候要素诊断量表》; * 与专家综合判断比较, $P < 0.05$; [△] 与 1994 标准比较, $P < 0.05$

风证候要素诊断量表的再验证, 对量表的敏感度、特异度进行了校正。分析结果给出的专家经验综合辨证的敏感度、特异度估计值显示, 专家经验综合辨证敏感度、特异度大多能保持在 0.75 ~ 0.96 之间, 体现出较好的准确性。但由于这一判断结果是基于 3 位专家中的 2 位及以上意见, 尚不能因此认为单个专家独立辨证也能保证较好的敏感度和特异度。研究中得到的量表校正后的数据部分体现出明显差异趋势, 但缺乏统计学意义, 这一情况可能和样本量大小有关。本研究参考的多诊断方法用于同一总体时估计灵敏度与特异度的校正方法, 其样本

量要求未见文献报道, 有待进一步研究。

中风病证候量表研究起于 20 世纪 80 年代, 1994 标准的研究奠定了中风病证候量表研究的基础, 但其他中医疾病领域证候量表研究刚刚起步, 大多缺乏可供参照的早期量表和先验信息数据, 这可能给本研究报告的方法在其他疾病证候量表研究过程中的推广应用带来困难。缺乏完善金标准情况下的诊断学数据校正方法是诊断医学统计学中重要的研究领域, 已有多种校正方法发表, 这些方法在中医证候量表的研究中有着广泛的应用空间, 值得进一步关注和探讨。

参考文献

- [1] 张宇龙, 刘强, 高颖, 等. 辨证一致性问题 [J]. 辽宁中医杂志, 2007, 34(10): 1384-1386.
- [2] 国家中医药管理局脑病急症科研组. 中风病辨证诊断标准(试行) [J]. 北京中医药大学学报, 1994, 17(3): 64-66.
- [3] 高颖, 马斌, 刘强, 等. 缺血性中风证候要素诊断量表编制及方法学探讨 [J]. 中医杂志, 2011, 52(24): 2097-2101.
- [4] 高颖, 马斌, 刘强, 等. 《缺血性中风证候要素诊断量表》临床验证 [J]. 中医杂志, 2012, 53(1): 2097-2101.
- [5] HUI SL, ZHOU XH. Evaluation of diagnostic tests without gold standard [J]. Stat Methods Med Res, 1998, 7(4): 354-370.
- [6] ZHOU XH, OBUCHOWSKI NA, MCCLISH DK. Statistical Methods in Diagnostic Medicine [M]. New York: John Wiley Sons, 2011.
- [7] JOSEPH L, GYORKOS T, COUPAL L. Bayesian estimation of disease prevalence and parameters of diagnostic tests in the absence of a gold standard [J]. Am J Epidemiol, 1995, 141(3): 263-272.
- [8] DENDUKURI N, JOSEPH L. Bayesian approaches to modeling conditional dependence between multiple diagnostic tests [J]. Biometrics, 2001, 57(1): 158-167.

Correction of Imperfect Golden Standard Validation Result of Diagnostic Scale of Syndrome Elements in Ischemic Stroke

LIU Qiang¹, WANG Xiaonan², WANG Shaoqing³, GAO Ying³, ZHOU Xiaohua⁴

(1. China Center for Chinese Medicine International Cooperation, Beijing, 100101; 2. School of Statistics, Renmin University of China; 3. Dongzhimen Hospital, Beijing University of Chinese Medicine; 4. School of Public Health, University of Washington, U. S. A.)

ABSTRACT Objective To explore the correction method of sensitivity and specificity of Diagnostic Scale of Syndrome Elements in Ischemic Stroke lacking of golden standard of syndrome differentiation. **Methods** Totally 204 patients with ischemic stroke in 10 hospitals of 9 provinces and cities around China were collected. Each patient took syndrome elements discrimination adopting Diagnostic Scale of Syndrome Elements in Ischemic Stroke and Criteria for Diagnosis of Stroke respectively. At the same time, 3 Chinese medicine experts made independent clinical experience differentiation on patients' syndrome elements. Consensus of two or more experts was the expert clinical experience discrimination result of syndrome elements. The correction of sensitivity and specificity was finished. When constructing likelihood function, the authors supposed that the expert experience comprehensive syndrome differentiation method was independent from scale syndrome differentiation, while syndrome differentiation results of two scales were not independent from each other. By Markov chain Monte Carlo (MCMC) method, according to Bayesian Parameter Estimation principle, posterior density estimation was carried out adopting sampling technique. **Results** Compared with Criteria for Diagnosis of Stroke, the sensitivity of the syndrome of endogenous wind and the syndrome of blood stasis in Diagnostic Scale of Syndrome Elements in Ischemic Stroke increased significantly, while the sensitivity of the syndrome of phlegm-dampness and the syndrome of Qi deficiency decreased ($P < 0.05$). The specificity showed well both in scales and expert experience comprehensive discrimination. **Conclusion** This correction method, estimating sensitivity and specificity adopting multiple diagnostic methods on the same population, could be applied for estimating sensitivity and specificity of Diagnostic Scale of Syndrome Elements in Ischemic Stroke lacking of golden standard after correction.

Keywords ischemic stroke; diagnostic criterion; diagnostic scale; syndrome element

(收稿日期: 2015-06-02; 修回日期: 2015-09-10)

编辑: 崔京艳

欢迎邮购 《中医杂志》 光盘合订本

1951-1986 年(4 张光盘) 480 元; 1987-1994 年(1 张光盘) 260 元; 1995-2004 年(1 张光盘) 280 元; 2005-2013 年(1 张光盘) 300 元。

光盘合订本为计算机用数据光盘, 收入了同期出版的全部《中医杂志》的内容, 可对所有文章进行任意字检索, 方便实用。

汇款地址: 北京东直门内南小街 16 号中医杂志社北京医海林音像书店, 邮政编码: 100700

电话: (010) 64035632; 网址: <http://www.jtcm.net.cn>