

熵聚堆方法在络气虚滞诊断标准研究中的应用

高怀林¹,张富生¹,吴以岭^{2*},陈静³,西广成³,贾振华¹,袁国强¹,吴相春¹,魏聪²,谷春华²,陈建新³

(1.河北医科大学附属以岭医院内科,石家庄 050017; 2.河北以岭医药研究院,石家庄 050035; 3.中国科学院自动化研究所复杂系统与智能科学重点实验室,北京 100080)

[摘要] **目的:**探索熵聚堆方法在络气虚滞诊断标准研究中的应用。**方法:**首先根据临床流行病学调查获得400例血管内皮功能障碍的四诊资料,以熵聚堆方法提取血管内皮功能障碍络气虚滞证的中医症状信息;然后计算络气虚滞证中各症状的关联度并确定血管内皮功能障碍络气虚滞证诊断阈值;最后根据临床流行病学诊断性试验研究方法,从灵敏度、特异度、符合率3个方面对此诊断标准进行回顾性和前瞻性检验。**结果:**建立的血管内皮功能障碍络气虚滞证的诊断标准有较好的诊断效能。**结论:**熵聚堆方法适用于中医证候诊断标准的研究,方法科学严谨,有临床实用价值。

[关键词] 络气虚滞;熵;证候;诊断标准

[中图分类号] R 256.25 **[文献标识码]** A **[文章编号]** 0258-879X(2007)07-0778-04

Application of entropy partition method for studying diagnostic criteria of Traditional Chinese Medicine syndrome

GAO Huai-lin¹, ZHANG Fu-sheng¹, WU Yi-ling^{2*}, CHEN Jing³, XI Guang-cheng³, JIA Zhen-hua¹, YUAN Guo-qiang¹, WU Xiang-chun¹, WEI Cong², GU Chun-hua², CHEN Jian-xin³ (1. Yiling Hospital, Hebei Medical University, Shijiazhuang 050017, China; 2. Hebei Yiling Pharmaceutical Research Institute, Shijiazhuang 050035; 3. Key Laboratory of Complex Systems and Intelligence Science, Institute of Automation, Chinese Academy of Science, Beijing 100080)

[ABSTRACT] **Objective:** To investigate the application of entropy partition method (EPM) for studying the diagnostic criteria of Traditional Chinese Medicine (TCM) syndrome. **Methods:** A clinical epidemiological survey was conducted in 400 patients with vascular endothelial dysfunction and the EPM was adopted to extract the TCM syndrome elements of "stagnancy of collateral-Qi in a deficiency condition", the correlation degree between each symptom of "stagnancy of collateral-Qi in a deficiency condition" was calculated and the diagnostic threshold of vascular endothelial dysfunction was established. The sensitivity, specificity and accuracy of the EPM-associated diagnostic criteria were assessed by comparing with those of clinical epidemiology. **Results:** The diagnostic criteria of "stagnancy of collateral-Qi in a deficiency condition" in vascular endothelial dysfunction established with EPM had a satisfactory diagnostic efficiency. **Conclusion:** EPM for complex system is suitable for studying the diagnostic criteria of TCM syndrome. EPM is a scientific and strict method and has great value in clinical practice.

[KEY WORDS] stagnancy of collateral-Qi in deficiency condition; entropy; symptom complex; diagnostic criteria

[Acad J Sec Mil Med Univ, 2007, 28(7): 778-781]

血管内皮(ET)功能障碍是冠心病、糖尿病等血管病变的始动因素及早期表现,在此阶段控制病情的进一步发展,对降低患者的致残、致死率,提高患者的生存质量有重要意义。众所周知,中医在防治血管病变方面有其独特的优势,因在血管内皮功能障碍阶段,病情较轻,未引起人们的足够重视,迄今未见关于血管内皮功能障碍的中医证候诊断标准。

本研究根据中医证候是“非线性的复杂巨系统”的理论^[1],采用复杂系统的分析方法^[2-3](熵聚堆方法)建立了血管内皮功能障碍络气虚滞证的中医证候量化诊断标准,以期对冠心病、糖尿病血管病变及其他血管性疾病的防治有所裨益,结果报告如

下。

1 对象和方法

1.1 对象

1.1.1 样本选择 纳入标准:(1)单纯高血压、高血脂、超重或肥胖、吸烟、代谢综合征人群,至少具备其中一项诊断;(2) NO<5 μmol/L 及 ET>53.38 pg/ml。排除标准:(1)心、脑、肾、神经、眼底等靶器官损

[基金项目] 国家重点基础研究发展计划(973计划)(2005CB523301)。Supported by National Program on Key Basic Research Projects (973 Program) (2005CB523301)。

[作者简介] 高怀林,博士生,副主任医师。

* Corresponding author. E-mail:jiatcm@163.com

伤者;(2)严重肝肾功能不全者;(3)妊娠或哺乳期妇女、精神病患者。

1.1.2 一般情况 病例为2006年6~9月河北医科大学附属以岭医院、河北省鹿泉市人民医院、河北师范大学医院符合上述标准的门诊体检者400例,将其随机分为2组,甲组300例,其中男127例,女173例,年龄25~82岁,平均年龄(55.17±10.38)岁,为运算组;乙组100例,其中男45例,女55例,年龄21~77岁,平均年龄(54.25±10.61)岁,为考核组。甲组用于建立络气虚滞证的量化诊断标准及对量化诊断标准进行回顾性检验,乙组用于量化诊断标准的前瞻性检验。

1.2 确定临床调查表 首先广泛查阅有关中医“肥胖”、“眩晕”、“头痛”、“消渴”等与内皮功能障碍有关疾病中的症状描述,按频数筛选症状,确定“血管内皮功能障碍中医证候信息条目池”;采用德尔菲分析法(Delphi法),从全国范围选择30名本专业专家对本条目池进行两轮专家问卷调查,综合两轮专家问卷调查结果确定进行临床调查的中医四诊信息条目;对上述最后确定的临床中医四诊信息条目参照文献^[4-5]进行中医四诊信息的量化分级,按无、轻、中、重四级分别计为0、1、2、3分;对不能分级的舌象、脉象分为无、有两类,分别计为0、1分,最后形成“内皮功能障碍中医证候临床调查表”。

1.3 数据分析 采用熵聚堆方法对数据进行处理分析,按症状聚堆结果,提取内皮功能障碍络气虚滞证的中医症状信息;计算络气虚滞证中各症状的关联度,并转化为症状对证候的贡献分值;调整参数,绘出积分与专家辨证(专家辨证指同时以3位副主任医师以上职称的专家对该部分患者进行辨证,以三者的一致性意见为专家辨证结果)符合率对应图,确定血管内皮功能障碍络气虚滞证诊断阈值;最后根据临床流行病学诊断性试验研究方法,进行量化诊断标准的回顾性和前瞻性检验。

以上数据采用 Matlab 工程软件分析处理。

2 结果

2.1 症状聚堆和症状抽提 采用熵聚堆方法,在不同参数调整下把有内在联系的变量(症状)聚在一起,得到多个由若干症状组成的四诊信息集合群(略);参照中医理论,把络气虚滞证的症状集合抽提并且合并,在此基础上,根据症状在证候中的关联度

系数确定络气虚滞证的最终症状。即得到络气虚滞证的症状集合:乏力、神疲、气短、懒言、自汗、舌淡、脉弱。

以内皮功能障碍甲组的病例300×57数据为例(300为病例数目,57为症状数目),我们应用熵聚堆方法在不同的参数调整下可以得到大量的症状集合群。根据充分证明了的“N级相关”思想,我们将多个症状之间的关联度值的计算转化为了任意两个症状之间关联度的计算(略)。在这些计算数值的基础上,根据亲友团思想和滤波思想,应用算法进行聚堆。

事实上,在此络气虚滞证集合中,每一症状同其他症状都有其关联度值。我们考虑两个症状之间的相关时必须考虑到其间关联强度的不对等性,即最终进入确定为络气虚滞证候的所有症状两两之间都必须强相关的。综合考虑我们确定了内皮功能障碍络气虚滞证中各症状,见表1。

表1 血管内皮功能障碍络气虚滞证各症状对证候关联度及贡献分值

症状	关联度	贡献分值
乏力	0.090 164	4
神疲	0.067 047	3
气短	0.078 747	3
懒言	0.054 997	2
自汗	0.023 065	1
舌淡	0.026 745	1
脉弱	0.018 606	1

2.2 计算各症状对络气虚滞证的关联度并转化为贡献分值 各基本证型即为由特定若干各症状组成的集合,而各个症状对各证候的贡献打分为其同该症状集合的关联度数值,反映的是该症状对于该症状集合(证候)的组成所作的贡献(重要程度)。根据具体数值,为使得临床应用时操作更为方便,我们将该贡献数值作适当的整数化处理。

以内皮功能障碍络气虚滞证型下的乏力症状为例,我们计算乏力在络气虚滞证型中的贡献值。该贡献值由乏力和该证型下其他症状之间的关联度系数整合得到,乏力:

$$g_i = \sum_{k=1}^k \frac{\mu(i,j) + \mu(j,i)}{2} = 0.090 164$$

其中, k 为络气虚滞证型下的症状数目, $\mu(i,j)$ 、 $\mu(j,i)$ 为乏力症状同其他症状之间的关联度系数。

在得到确定各个症状之后,我们将确定的用于诊断络气虚滞证的症状放一集合中,计算各症状对此症状集合(证候)的关联度及转化后的贡献分值(表1)。

2.3 确定诊断阈值 理论上的阈值如果太小或是太大,其应用于原始患者身上的符合率(与专家辨证结果比较)都会较低,也就是说,理论上存在着一个最优的阈值使得将该阈值方案应用于原始患者后的符合率为最高。我们假定预先完全不知道各证型中阈值的存在与否,试着从算法上去寻找这个最优点。结果表明,事实上存在着这样的阈值,该阈值方案的应用同临床也是符合的。见图1。据症状得分与符合率对应关系图,可看出5分为最优诊断阈值。

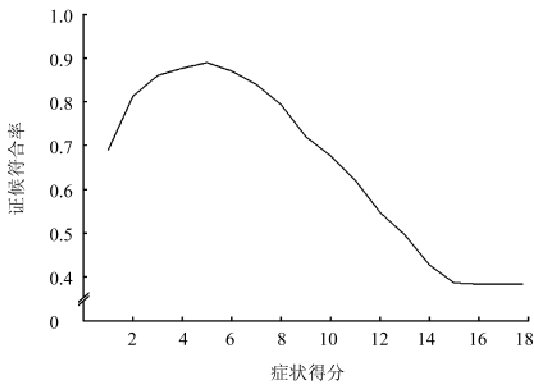


图1 内皮功能障碍络气虚滞证候符合率-症状得分曲线

2.4 对诊断阈值的检验 根据表2中的数据计算,敏感度、特异度、符合率分别为86.5%、93.0%、89.0%。根据表3中的数据计算,敏感度、特异度、符合率分别为89.7%、93.8%、91.0%。

表2 血管内皮功能障碍络气虚滞证患者量化诊断标准回顾性检验数据

标准辨证	专家辨证		合计
	阳性	阴性	
阳性	160(a)	8(b)	168
阴性	25(c)	107(d)	132
合计	185	115	300

灵敏度 = a/(a+c); 特异度 = d/(b+d); 符合率 = (a+d)/(a+b+c+d)

从以上回顾性和前瞻性检验结果来看,其灵敏度、特异度、符合率均大于85%,说明此诊断阈值有较好的诊断效能。

表3 血管内皮功能障碍络气虚滞证患者量化诊断标准前瞻性检验数据

标准辨证	专家辨证		合计
	阳性	阴性	
阳性	61(a)	2(b)	63
阴性	7(c)	30(d)	37
合计	68	32	100

灵敏度 = a/(a+c); 特异度 = d/(b+d); 符合率 = (a+d)/(a+b+c+d)

3 讨论

众所周知,血管内皮功能障碍是整个血管病变的始动因素与初始阶段,与血管病变的发生、发展有密切的联系,因此,人们提出了亚临床血管病变——微循环障碍的概念,其主要表现为血管内皮功能障碍[6]。因为微循环与中医之络脉高度相关,所以我们在血管内皮功能障碍的中医证候诊断标准研究中,把气虚证描述为络气虚滞证更贴近临床实际。

证候学是中医理论的核心,近50年来证候研究一直是中医研究的热点和难点,但始终没取得大的进展,成为了制约中医发展的瓶颈。在中医自身苦苦寻觅找不到好的方法的情况下,多学科的交叉渗透或许能为中医证候的研究寻找到一突破口。近年来,随着复杂系统科学的蓬勃发展,为中医证候的标准化、规范化研究提供了新思路和新方法。

我们进行症状聚堆和症状抽提都基于熵这个概念。熵的概念源于物理学的范畴,在公元1854年由德国物理学家克劳修斯(Clausius)提出。此后,这一概念广泛发展并应用于许多领域。简单地说,从宏观角度理解,熵增意味着能量的退化;从微观角度理解,熵是系统混乱程度的标志。熵在信息论中的引用为信息的定量描述引入了信息熵这个概念,为熵概念的进一步泛化奠定了基础。近年来,熵理论在生命科学的定性定量研究特别是中医领域中也有了长足的发展。以一个具体的疾病为模型,运用熵研究证候、症状、疾病、方剂的相关性,从而为中医证候量化诊断标准研究提供依据。

对于一个只了解描述其特征的若干变量的复杂系统S,可以由以下矢量表示:

$$S = (X_1, X_2, \dots, X_i, \dots, X_p)^T \quad (1)$$

其中 $X_i = (X_{ia}), (i=1, 2, \dots, p; a=1, 2, \dots, q)$ 是描述系统某个特征的变量。

我们的目标是从大量的 X_i 集里归结出具有相近特征的 X_i 子集。令 $Q_i (i=1, 2, \dots, p)$ 为 X_i 分类的集合, Q_i 的第 a 个元素 Q_{ia} , 则有 $Q_i = \{1, 2, \dots, a, \dots, k\}, k \leq q$, 并令 $n_a [\sum_{a=1}^k n_a = q]$ 为事件 X_i 属于 C_i 第 a 类的数量。则变量 X_i 的熵定义为:

$$H(X_i) = - \sum_{a=1}^k n_a / q \log n_a / q \quad (2)$$

X_i 和 X_j 的联合熵定义为:

$$H(X_i, X_j) = - \sum_{a,b} n_{ab} / q \log n_{ab} / q \quad (3)$$

其中 n_{ab} 表示事件 X_i 属于 C_i 的第 a 类同时 X_j 属于 C_j 的第 b 类的数量。

在上述熵定义的基础上, 下面给出关联度的定义。

定义 1: 假设 $X_i \cap X_j = \phi$, 则称熵:

$$\mu(X_i, X_j) = H(X_i) + H(X_j) - H(X_i, X_j) \quad (4)$$

为 X_i 和 X_j 之间的关联度。

定义 2: 假设对任意 $i, j (i \neq j)$, $X_i \cap X_j = \phi$, p 为任意正整数, 称

$$\mu(X_1, X_2, \dots, X_p) = \sum_{i=1}^p H(X_i) - H(\sum_{i=1}^p X_i) \quad (5)$$

为 X_1, X_2, \dots, X_p 之间的关联度^[6]。

以上关于变量之间关联度的定义完全适用于子系统之间关联度的定义。事实上, 变量本身作为系统的子集也是子系统。

定义 3: 假定系统 S 被划分为 n 个子系统 S_1, S_2, \dots, S_n , 对任意子系统 $S_i = (X_i^1, X_i^2, \dots, X_i^{n_i})$ 和 $S_j = (X_j^1, X_j^2, \dots, X_j^{n_j}) (i \neq j)$ 而言, $S_i \cap S_j = \phi, S = \sum_{i=1}^n S_i$, 定义

$$\mu(S_1, S_2, \dots, S_n) = \sum_{i=1}^n H(S_i) - H(\sum_{i=1}^n S_i) \quad (6)$$

为 S_1, S_2, \dots, S_n 之间的关联度。

将复杂系统分析方法用于中医证候的研究, 是近几年中医证候研究的热点和趋势, 尤其在证候的诊断标准研究中取得了较好的结果。熵聚堆方法^[2-3,7]是信息论中熵方法和熵语言在非线性相关模式识别领域的具体应用, 该方法遵照数据的内在联系, 不对数据作刚性先行分割, 依据数据内在关联进行自主聚类, 可以无监督地处理多变量、多层次上的

复杂数据, 对于提取中医证候相关症状并分析症状之间的非线性相关关系具有重要应用价值。它不仅适用于处理线性数据, 而且适用于处理非线性数据。熵聚堆方法在中医证候研究中的应用, 是通过信息熵的关联度实现的, 它可以度量变量之间的任意统计相关性, 对变量的分布类型没有任何特殊要求。它把密切相关的症状置于同一个集合中, 而这样的症状集合正是中医理论下证候的内涵体现。王阶等^[8]通过多中心流行病学调查, 收集包括 5 种病的 696 例病例, 采用信息熵的关联度和多元对应分析对中医证候中的血瘀证及亚型进行研究, 综合以上 2 种算法, 说明复杂算法更适合中医证候规范化研究。我们在自己的课题“脉络-血管系统病”^[9]的中医证候量化诊断标准研究中, 也引入了复杂系统分析方法, 并取得了较满意的结果。通过对内皮功能障碍络气虚滞证的诊断阈值的确定, 诊断标准的建立并检验, 可看出, 熵聚堆方法科学严谨, 结果可靠, 可用于中医证候的诊断标准的研究, 有其实用价值。

[参考文献]

- [1] 郭 蕾, 张启明, 王永炎, 等. 证候规范化研究的思路和方法探讨[J]. 中国中西医结合杂志, 2006, 26: 258-261.
- [2] 西广成. 复杂系统分划的熵方法[J]. 自动化学报, 1984, 13: 216-220.
- [3] 西广成. 生态经济区划的熵方法[J]. 自动化学报, 1990, 16: 170-173.
- [4] 徐迪华, 徐剑秋. 中医量化诊断[M]. 南京: 江苏科学技术出版社, 1997: 35-84.
- [5] 郑筱萸. 中药新药临床研究指导原则[M]. 北京: 中国医药科技出版社, 2002: 378-388.
- [6] 王 红, 张向阳. 血管内皮功能障碍的研究进展[J]. 新疆医科大学学报, 2005, 28: 180-182.
- [7] 陈 静, 西广成, 易建强, 等. 中医复杂系统中两类分划方法的比较[C]. 中国自动化协会第 21 届青年学术年会. 2006. 中国烟台.
- [8] 王 阶, 李海霞, 孙占全, 等. 基于复杂算法的中医证候研究[J]. 北京中医药大学学报, 2006, 29: 581-585.
- [9] 吴以岭. “脉络-血管系统”相关性研究[J]. 中医杂志, 2007, 48: 5-8.

[收稿日期] 2007-05-20

[修回日期] 2007-06-20

[本文编辑] 尹 茶