

类风湿关节炎证候诊断因子优化组合研究

白云静¹ 申洪波² 姜德训¹ 孟庆刚³

(1 北京军区总医院,北京市东城区东直门内大街 5 号,100700; 2 北京大学第三医院; 3 北京中医药大学)

摘要 目的:建立类风湿关节炎(RA)证候诊断因子优化组合,实现对 RA 证候诊断信息进行降维研究的目的。方法:采用聚类分析、Logistic 逐步回归分析、逐步判别分析、因子分析等多元统计分析方法,对一组 RA 临床证候资料进行统计分析。结果:建立了 RA 证候诊断因子优化组合。结论:RA 证候诊断因子优化组合具有较高的诊断价值;将聚类分析、Logistic 逐步回归分析、逐步判别分析、因子分析等多元统计分析方法联合应用的证候因子优化组合方法具有很好的降维处理能力。

关键词 类风湿关节炎/证候研究

Optimization of Diagnostic Factors of Rheumatoid Arthritis Syndromes

Bai Yunjing¹, Shen Hongbo², Meng Qinggang³

(1 The Military General Hospital of Beijing, Add.: No. 5, Nanmencang, Dongsi, Dongcheng District, Beijing, Post code: 100700; 2. Third Hospital of Beijing University; 3. Beijing University of Chinese Medicine)

Abstract Objective: To optimize the diagnostic factors of rheumatoid arthritis (RA) syndromes to achieve dimension reduction research of RA diagnostic information. **Methods:** To analyze clinical syndrome information of RA using such multiple statistical analyses as cluster analysis, logistic gradual regressive analysis, stepwise discriminate analysis, factor analysis, etc. **Results:** We have established optimal combination of the diagnostic factors of RA. **Conclusion:** The abovementioned combination has high diagnosis value for RA syndromes and the multiple statistical analyses applied have good dimension reducing capacity.

Key Words Rheumatoid arthritis/ Syndrome research

证候是一个多维多阶多变量的复杂系统,证候诊断信息具有典型的高维性特征,这给证候诊断的规范化研究带来相当的难度,为此,我们开展了类风湿关节炎(rheumatoid arthritis,简称 RA)证候因子优化组合的研究,以期进一步探讨对证候诊断信息进行降维处理的可行性方法。

1 资料来源

采用望京医院对全国 8 个地区三级以上中医院的住院患者所做的证候流行病学调查资料,共 765 例。参考《中药新药临床研究指导原则》^[1]、文献整理结果及专家经验,共观察 183 个症状、体征,每个症状按无、轻、中、重分别记为 0、1、2、3 分。并同时给出 10 个基本证型的诊断。

2 研究方法

2.1 初步筛选类风湿关节炎的中医四诊信息 首先统计 765 例患者共 183 个症状和体征的出现频率及症状得分情况,保留出现频率 $\geq 5\%$ 的 86 个症状,暂时保留症状出现频率在 3%~5%的 14 个症状和症状出现频率虽然 $< 3\%$ 但根据专业知识判断对 10 个假设证型可能有诊断意义的 3 个症状,舍弃其余 80 个症状。然后将粗筛后暂时保留的 17 个症状分别进行单变量方

差分析。结果显示:其中 6 个症状的 P 值均 < 0.05 ,提示对各证型的贡献度有显著性差异,应予以保留。经过第一步粗筛和第二步单变量统计分析进行指标的初步筛选后,183 个指标共保留 92 个。

2.2 初选指标的聚类分析——确定证类及证型的相关程度 将上述 92 个指标作为变量,进行变量聚类分析,在聚为 10 类时,合计能解释的方差为 57.41304,占合计总方差的 62.41%,在此部位截取。其中第六类包含的症状主要为:胸闷、气短、心慌,根据中医理论与临床经验,该类无明显证候特征,可淘汰该类。其余 9 类分别与临床诊断的证型相对应,只是无其中的脾肾阳虚证,分析其原因,765 例患者中脾肾阳虚证共有 10 例,全部有兼加证存在,说明脾肾阳虚证与他证兼加情况较明显,且病例数少,无法进行分析,可剔除。根据聚类分析结果,最终确定 765 例观察病例的 92 个指标集中反映了寒湿阻络证、肝肾不足证、瘀瘀阻络证、气血两虚证、湿热阻络证、瘀血阻络证、寒热错杂证、阴虚内热证、热毒蕴结证等 9 个证型的诊断信息,故确定本资料的中医证型为上述 9 个。另外,系统聚类分析在给出证候分类的同时,还给出了典型的证候群,还可以根据各分类之间的类间相关值以及指标聚类树形图粗略地判断各分类(证型)之间的相关程度。

2.3 初选指标的 Logistic 回归分析——优化指标 用

基金项目:国家自然科学基金资助(NO:90409005),教育部博士点基金资助(NO:20050026011)

Logistic 回归分析的逐步法、前进法或后退法从 92 个初选指标中筛选 9 个证型的诊断敏感性指标。通过分析,将上述 9 种证型的诊断优化指标由 92 个降为 52 个。(见表 1)

表 1 logistic 回归分析筛选的优化指标

证型	症状
湿热阻络证	脉弦滑、苔黄腻、小便黄、关节肿痛、关节热、心烦、关节红
寒湿阻络证	关节喜暖、脉细、恶风寒、关节重着、舌苔白、肢冷
寒热错杂证	关节喜暖、小便黄、关节热、脉滑数、舌质红、喜暖、脉弦滑
肝肾不足证	腰背酸痛、腰膝酸软、肌肉萎缩、足跟痛、关节强直、倦怠乏力、关节畸形
气血两虚证	语声低微、爪甲色淡、肌肉酸痛、心慌、面色不华、倦怠乏力
阴虚内热证	手足心热、盗汗、多梦、苔少津、大便干、腰膝酸软、眼干、关节酸痛
瘀血阻络证	口唇紫暗、舌有瘀斑、舌底紫暗、舌质暗、面色暗、舌底迂曲、痛如针刺
痰瘀阻络证	色素沉着、边有齿痕、脉沉细、关节重着、舌质淡暗、苔白腻、口唇淡暗
热毒蕴结证	关节红、关节热、高热、口干欲饮、多汗、小便黄、脉弦细数

2.4 初选指标的逐步判别分析——补充优化指标
逐步判别分析要求病例的证型诊断没有兼加证的存在。本组 765 例病例中有 398 例只有单证诊断,分别为:湿热阻络证 163 例、寒湿阻络证 96 例、寒热错杂证 9 例、肝肾不足证 42 例、气血两虚证 32 例、阴虚内热证 5 例、瘀血阻络证 12 例、痰瘀阻络证 27 例、热毒蕴结证 12 例。脾肾阳虚证 10 例患者均有兼加证的存在,故舍弃该证型。对上述 398 份病例初步筛选后的 92 个指标 9 个证型作逐步判别分析以筛选变量。最终保留优化指标 46 个。(见表 2)

表 2 逐步判别分析筛选的指标

优化指标					
疼痛固定	关节热	倦怠乏力	舌苔白	寐差	阴雨加重
关节喜暖	遇冷痛甚	口干不饮	苔黄	恶风寒	舌质红
多汗	盗汗	喜暖	舌底紫暗	脉弦细	腰膝酸软
口干欲饮	舌质淡暗	肢冷	面色不华	小便黄	关节强直
舌体胖大	肢体麻木	关节重着	边有齿痕	舌底迂曲	多梦
脉弦滑	脉弦	肌肉酸痛	手足心热	耳鸣	痛如针刺
心烦	语声低微	风湿结节	关节酸痛	眼干	脉细
口唇紫暗	爪甲色淡	足跟痛	高热		

2.5 汇总优化指标 将 logistic 回归分析筛选的 52 个优化指标和逐步判别分析筛选的 46 个优化指标合并,并参考聚类分析 9 类证型所对应的证候群,确定 RA9 种证候诊断的优化指标共 64 个。(结果略)。

2.6 优化指标的因子分析——确定主、次症

2.6.1 用因子分析法进一步筛选指标 由 logistic 回归分析和逐步判别分析可得出各证型的优化指标共

64 个。首先对 64 个指标采用主成分分析法提取公因子,分析矩阵为协方差阵,用最大方差旋转法对因子轴进行旋转,对上述因子载荷阵中因子载荷小于 0.2 的变量剔除,并重新按上述方法进行分析,再剔除因子载荷小于 0.2 的变量,如此循环,直至变量的因子载荷均大于 0.2 为止。经过上述分析,最后纳入因子分析的变量数为 58 个。(结果略)

2.6.2 用因子分析法确定主、次症 对最终确定的 58 个指标再次进行因子分析,根据因子载荷阵的数值确定基本证型诊断的主、次症。首先用主成分分析方法提取公因子。从累积贡献率来看,取前 9 个特征值时,累积贡献率为 0.7929,根据最小特征值标准保留 9 个公因子,这 9 个公因子对 58 项指标的全部信息的累积贡献率为 79.29%。

然后对 58 个症状进行因子分析,采用最大方差旋转法,保留因子载荷大于 0.20 的指标,对因子载荷小于 0.2 的变量剔除。根据上述因子载荷阵的结果,取症状对证型贡献度为前三位的为主症,其余为次症,列出 RA9 种证型的证候因子优化组合。(见表 3)。

表 3 RA 基本证型证候因子优化组合

证型	主症	次症
湿热阻络证	关节热、苔黄腻、关节肿痛	小便黄、关节红、脉弦滑、舌质红、多汗、心烦
寒湿阻络证	关节喜暖、恶风寒、遇冷痛甚	关节重着、舌苔白、肢冷、阴雨天加重、边有齿痕、脉细、舌体胖大
寒热错杂证	关节喜暖、小便黄、关节热	脉滑数、舌质红、关节肿痛、喜暖
肝肾不足证	关节酸痛、腰膝酸软、肌肉萎缩	腰背酸痛、足跟痛、关节强直、关节畸形、耳鸣、寐差
气血两虚证	语声低微、爪甲色淡、倦怠乏力	面色不华、肌肉酸痛、关节酸痛
阴虚内热证	手足心热、盗汗、苔少津	大便干、眼干、多梦、舌质红
瘀血阻络证	口唇紫暗、舌有瘀斑、痛如针刺	舌底紫暗、舌质暗、疼痛固定、面色暗、舌底迂曲、脉沉细
痰瘀阻络证	色素沉着、风湿结节、舌质淡暗	关节畸形、脉沉细、苔白腻
热毒蕴结证	关节红、关节热、高热	关节肿痛、口干欲饮、小便黄、脉弦细数、苔黄

2.7 判别分析——检验证候因子优化组合的诊断价值 判别分析只适用于单类判别的情况,为了检验优化指标的可靠性,我们从 765 例样本中筛选出 398 例只有单证诊断的病例进行判别分析。其中湿热阻络证 163 例,寒湿阻络证 96 例,寒热错杂证 9 例,肝肾不足证 42 例,气血两虚证 32 例,阴虚内热证 5 例,瘀血阻络证 12 例,痰瘀阻络证 27 例,热毒蕴结证 12 例。

判别分析结果显示:用优化出的 58 个指标建立的

判别函数模型诊断准确率为 86.18%。说明 RA 的证候因子优化组合对 9 种基本证型的诊断非常有意义。从总的分类矩阵表中可见,第一组判对 139 名占 85.28%,判错 24 名占 14.72%;第二组判对 87 名占 90.63%,判错 9 名占 9.37%;第三组判对 7 名占 77.78%,判错 2 名占 22.22%;第四组判对 38 名占 90.48%,判错 4 名占 9.52%;第五组判对 23 名占 71.88%,判错 9 名占 28.12%;第六组判对 4 名占 80%,判错 1 名占 20%;第七组判对 10 名占 83.33%,判错 2 名占 16.67%;第八组判对 23 名占 85.19%,判错 4 名占 14.81%;第九组判对 12 名占 100%。总体上该模型共判对 343 名占 86.18%,判错 55 名占 13.82%。说明该模型的诊断准确率为 86.18%。

3 结果

我们首先对 765 例类风湿关节炎患者的临床观察指标共 183 个进行初步筛选;然后再对初步筛选后的指标进行聚类分析;在聚类分析的基础上进行 logistic 回归分析优化指标;同时对初步筛选后的指标用逐步判别分析进行筛选;将 logistic 回归分析和逐步判别分析筛选的指标合理合并,且参考最初聚类分析指标筛选的结果,确定证候诊断的优化指标群;在此基础上进行因子分析,最终确定了 RA 证候的诊断指标优化组合。

经过上述分析,得出如下结果:1)确立了 RA 的基本证型为湿热阻络证、肝肾不足证、寒湿阻络证、瘀血阻络证、气血两虚证、痰瘀阻络证、阴虚内热证、热毒蕴结证和寒热错杂证等 9 个证型。2)确立了 RA 的证候因子优化组合(详见表 3)。3)用上述证候因子优化组合建立的证候线性判别函数模型具有很好的判别能力。

4 讨论

4.1 证候诊断指标的降维研究 中医证候具有高维性特征,在此,“维”有多方面的含义:包括证候诊断指标的高维性、证候构成要素的高维性、证候诊断方法的高维性、证候动态时相的高维性等等。我们仅探讨对证候诊断指标高维性的降维处理方法。

在适当的范围内,维度越小,越容易掌握,使用者的可操作性也就越大。但是,中医证候的四诊信息繁多而冗余,这给证候的诊断带来相当的难度,尤其不利于对证候诊断的规范化研究。因此,在保持证候特色的前提下,对证候诊断资料的高维性进行降维处理,就成为证候规范化研究的当务之急。

聚类分析、主成分分析、因子分析、回归分析、典型相关分析、多维尺度分析等多元统计方法是常见的降

维方法。本研究我们选择聚类分析、因子分析、回归分析等方法对 RA 的证候表征信息进行降维处理,通过多种统计方法的联合应用,最终建立了 RA 的证候因子优化组合。

4.2 初步筛选指标的必要性 首先,我们对资料所包含的全部 183 个中医四诊信息进行了初步筛选。多元统计分析都是建立在大样本基础上的,一般在用 DME 方法取得大样本资料后,都要先进行自变量(观察指标)的初筛,这是因为聚类分析、因子分析等没有筛选变量的过程,有多少变量进入分析,就有多少变量最终进入模型,这样不仅使分析过程繁琐复杂、因子模型过于冗长,而且还夹杂了许多不必要的混杂因素,影响了模型的敏感性;另外,主成分分析、因子分析等是基于奇异值分解原理的多元统计分析方法,在分析时如果变量过多,可能会导致原始资料收敛性不良,表现为入选主因子较多,累积贡献率却偏低。因此,在进行多元分析之前,我们先用频率筛选法和单变量统计分析对指标进行初步筛选,最终保留 92 个指标进入多元统计分析。

4.3 多种统计方法联合应用的意义 本试验虽然在聚类分析中没有得到脾肾阳虚证的类别,但临床并不能否认 RA 脾肾阳虚证的存在,我们观察的 765 例患者中有 10 例脾肾阳虚证患者,因为他们都有兼加证的存在,所以聚类分析对此无能为力。这暴露了聚类分析的局限性,聚类分析只能“硬性”分裂变量,使变量的归属非此即彼,一个症状仅能隶属于一个证型,这不符合中医的证候学特征。也就是说,聚类分析无法处理证候、症状的多重共线性问题。因此,聚类分析的结果也就大打折扣,尤其是由聚类分析得到的各证型的相应证候群并不能直接作为证候的优化指标,仅能作为筛选优化指标的参考。

针对上述各种统计方法的优缺点,我们选择了将上述方法有机结合的方式进行研究。在聚类分析的基础上,我们对 92 个指标的 765 份病例进行了 Logistic 回归分析。用回归分析进行变量筛选的方法很多,但最常用的是逐步分析法,因此,我们选择逐步法进行变量的筛选,遇到病例数较少,筛选结果不理想时,我们再用前进法和后退法进行补充,以便于筛选出更有意义的指标。Logistic 回归分析应有病例组与正常组的对照,在没有正常对照资料的情况下,我们根据聚类分析类相关值及聚类树形图的结果,选择与所分析证型距离较远(相关程度较小)证型的相应病例作为对照,从而保证了分析结果的可靠性。

4.4 降维效果的检验 用多元统计方法进行类风湿

卡培他滨联合草酸铂加健脾散结中药治疗晚期复发食管癌的临床观察

陈春永 朱迪盈 林少峰

(广东省中医院内科, 广东省广州市大德路 111 号, 510120)

摘要 目的: 通过健脾散结中药加上卡培他滨联合草酸铂的方案, 比较卡培他滨联合草酸铂的方案治疗晚期复发食管癌的疗效和不良反应。方法: 将既往接受过 PF 方案化疗的复发的晚期食管癌 76 例, 随机分为中药组与对照组各 38 例。对照组进行化疗同时配合口服大剂量的维生素 B6, 中药组在化疗同时配合健脾散结中药内服。结果: 中药组治疗晚期复发食管癌的有效率 (RR) 71.1% 明显高于对照组 44.7% ($P < 0.05$), 而在毒副反应方面, 譬如恶心呕吐、腹泻、贫血、血小板减少、肝功能损害、白细胞减少、手足综合征、脱发、神经毒性, 中药组发生的几率也比对照组明显降低。结论: 相对于卡培他滨联合草酸铂的方案, 健脾散结中药加上卡培他滨联合草酸铂治疗晚期复发食管癌具有良好的减毒增效的作用, 值得进一步的深入研究。

关键词 晚期复发食管癌/中医药疗法; 健脾理气; 卡培他滨; 草酸铂

Effects of Spleen Invigorating and Mass Resolving TCM Medicines on Chemotherapy for Advanced Esophageal Cancer Relapse

Chen Chunyong, Zhu Diying, Lin Shaofeng

(Dept. of Internal Medicine, Guangdong Hospital of Traditional Chinese Medicine, Add.: No. 111, Dade Road, Guangzhou, Guangdong Province, Post code: 510120)

Abstract Objective: To compare efficacy and adverse effect of a chemo therapy (Capecitabine + Oxaliplatin) and the therapy plus spleen invigorating and mass resolving TCM medicines for advanced esophageal cancer relapse. **Methods:** Seventy six patients with advanced esophageal cancer relapse and past PF chemo treatment, were randomly divided to the TCM group (38 cases) and control group (38 cases) to receive the TCM medicine, or Vitamin B6 respectively in addition to the chemo treatment. **Results:** The effective rate of the TCM group (71.1%) was significantly higher than that of the control group (44.7%, $P < 0.05$). Such adverse effect as vomiting, diarrhea, anaemia, plate - reduction, liver impairment, aleucocytosis, hand - foot syndrome, alopecia; neurotoxicity were also significantly reduced in the TCM group compared to the control group. **Conclusion:** Spleen invigorating and mass resolving TCM medicine can enhance efficacy and reduce adverse effect of the chemo treatment for advanced esophageal cancer relapse.

Key Words Advanced esophageal cancer relapse/ Chinese medical therapy; Spleen invigorating and qi regulating; Capecitabine; Oxaliplatin

食管癌是临床上最常见的恶性肿瘤之一, 50% 以上的食管癌患者诊断时已属晚期, 对于晚期食管癌的姑息治疗及辅助治疗, 化疗具有重要地位, 已逐步成为食管癌综合治疗的重要部分^[1]。卡培他滨 (capecitabine, Xeloda, 希罗达) 联合草酸铂 (Oxaliplatin, L -

OHP, 奥沙利铂) 的抗肿瘤作用已经得到肯定, 现广泛运用于晚期食管癌、胃癌、大肠癌的临床治疗中^[2-3]。但在使用该方案的化疗过程中出现的毒副反应, 如手足综合征、恶心、腹泻、呕吐、神经毒性等^[2,4-5], 严重影响治疗效果和患者的生活质量, 已经引起了广泛的关

关节炎证候诊断指标的降维处理后, 降维后的效果如何还需要进行必要的验证。理论上我们可以用 Logistic 回归分析方程、因子得分模型估计的因子得分值等进行验证, 但由于进入模型的指标太多, 上述方程式过于繁琐, 因此, 进行验证的工作量非常大。我们仅对具有单证诊断的 398 份病例进行了 58 个优化指标 9 个证型的判别分析验证。结果显示: 总体上该模型共判对 343 名占 86.18%, 判错 55 名占 13.82%, 诊断准确率为 86.18%。说明有上述 58 个指标对 9 个证型的判别函数具有很好的判别能力, 我们筛选的证候因子

优化组合有很高的诊断价值。

但是, 用判别分析进行检验属于原始判别, 即原始分类和信息同时进入模型, 以原始分类与判别结果比对, 看错判多少例。这种判别因为事先输入了原始分类, 所以结果并不十分可信。另一方面, 由于判别分析所建立的判别函数是线性模型, 而证候是非线性复杂系统, 四诊信息是证候的各层级结构逐级涌现的结果, 证候与症状之间的关系是非线性的, 显然, 这也说明判别分析结果的可靠性值得商榷。

(2009-09-28 收稿) □