

· 现场调查 ·

山东省中西部农村居民高血压危险因素
分类树分析

刘甲野 马吉祥 徐爱强 付振涛 贺桂顺 贾崇奇 于洋

【摘要】 目的 探讨山东省中西部地区 25 岁以上农村常住居民高血压的危险因素及高危人群。方法 采用多阶段分层随机抽样的方法,对该地区调查对象进行问卷调查、体格检查以及实验室检测。应用分类树分析高血压人群的危险因素及高危人群。结果 高血压的主要危险因素为年龄、腹型肥胖、超重或肥胖、高血压家族史、血糖值较高等。高血压的高危人群主要为高年龄者;年龄较大且血糖值较高者;年龄较大且腹型肥胖或超重者;有高血压家族史且年龄较大者;有高血压家族史、腹型肥胖且年龄较大者。分类树模型学习集的灵敏度为 71.87%,特异度为 66.38%,总正确率为 68.79%;检验集的灵敏度为 70.70%,特异度为 65.84%,总正确率为 67.97%。结论 控制体重和血糖为一般人群的预防措施,不同高危人群应根据所具有的危险因素采取不同的防治措施。高血压的人群防治应对一般人群的普遍预防和对高危人群的重点预防相结合。

【关键词】 高血压;危险因素;高危人群;分类树

Study on the risk factors of hypertension among rural residents in mid-west areas of Shandong province, using the classification tree analysis methodology LIU Jia-ye*, MA Ji-xiang, XU Ai-qiang, FU Zhen-tao, HE Gui-shun, JIA Chong-qi, YU Yang. *Institute of Biostatistics and Epidemiology, School of Public Health, Shandong University, Jinan 250012, China
Corresponding author: XU Ai-qiang, Email: aqxuepi@163.com

【Abstract】 Objective To explore the risk factors of hypertension and risk population for adults aged ≥ 25 in the mid-western rural areas of Shandong province and to provide evidence for development of intervention measure. **Methods** Subjects aged ≥ 25 were selected by multi-stage stratified random sampling method. All participants were interviewed with a standard questionnaire and physically examined on height, weight, waist circumference, blood pressure and fasting plasma glucose (FPG). Classification tree analysis was employed to determine the risk factors of hypertension and high risk populations related to it. **Results** The major risk factors of hypertension would include age, abdominal obesity, overweight or obesity, family history and high blood sugar. The major populations at high risk would include populations as: a) being elderly, b) at middle-age but with: high blood sugar or with abdominal obesity/overweight, or with family history, c) people at middle-age but with family history and abdominal obesity. Through classification tree analysis, sensitivity, specificity and overall correct rates were 71.87%, 66.38% and 68.79%, respectively on 'learning sample' while 70.70%, 65.84% and 67.97% respectively on 'testing sample'. **Conclusion** Efforts on both weight and blood sugar reduction were common prevention measures for general population. Different kinds of prevention and control measures should be taken according to different risk factors existed in the targeted high-risk population of hypertension. Community-based prevention and control for hypertension measures should be integrated when targeting the population at high risk.

【Key words】 Hypertension; Risk factors; High risk population; Classification tree

目前有关高血压的调查主要根据一些基本临床体征对病例和对照进行分组,忽略了不同病例间生物学特征和表型的多样性,组内同质性较差,利用传统的 logistic 回归等归一化模型分析时效果往往

不很理想,分类回归树(classification and regression trees, CART)通过递归分型能使局部病例的内部同质性得到有效改善。因此,本研究利用分类树分析方法探讨山东省中西部地区农村居民高血压相关影响因素。

对象与方法

1. 对象:采用多阶段分层随机抽样方法,根据社会经济发展和地理位置,在山东省西部抽取商河县、陵县、宁阳县和莱城区 4 个县(市、区),在中部抽取寿光

基金项目:卢森堡-世界卫生组织-中国山东省农村卫生人员培训与慢病控制资助项目(WP/2006/CHN/NCD/2.4/001)

作者单位:250012 济南,山东大学公共卫生学院流行病与卫生统计学研究所(刘甲野、贺桂顺、贾崇奇、于洋);山东省疾病预防控制中心(马吉祥、徐爱强、付振涛)

通讯作者:徐爱强,Email: aqxuepi@163.com

市、广饶县、沂源县和苍山县 4 个县(市、区),每县(市、区)按照随机抽样原则,随机抽取 2 个乡镇,每个乡镇随机抽取 2 个村,每村按户主花名册进行系统抽样,被抽取家庭内所有 25 岁以上常住人口均为调查对象。对抽中的居民进行问卷调查、体格检查和实验室检测。每个村计划调查 300 户(约 600 人),户数不足 300 的村,从临近村仍按照系统抽样方法补足。

2. 调查方法及标准:对所有调查对象采用入户面对面询问调查的方法进行问卷调查,并于次日进行体格检查,包括血压、身高、体重、腰围和血糖。血压的测量按照“2004 年中国高血压防治指南”确定的方法^[1]。血糖的测定采用葡萄糖氧化酶法,受试对象空腹 >10 h 抽取静脉血。高血压及其分级的诊断参考 1999 年 WHO/ISH 推荐的标准^[2]。肥胖的判定采用国际生命科学学会中国办事处中国肥胖问题工作组的研究结果^[3]。

3. 模型拟合方法:利用 CART 二叉树分支特征和阈值把节点处的病例分为 2 组,产生 2 个子结点,选择 Gini 不纯度指数作为分支标准。采用交叉有效性检验方法来确定学习集和检验集,其过程是首先将数据分成 10 份,每个子集中的数据随机选择得到,保证每个子集所包含的样本数大致相同,且包含的各类数据的个数也大致相同;然后,用样本的 90% 作为学习集生成树,通过修剪方法生成 1 个子树序列;最后用剩余样本的 10% 作为检验集,每次用样本的不同部分做检验集对同一子树重复测试 10 次,将得到的错误率的期望值作为子树的分类性能,比较各棵子树的性能,将期望错误率最低子树作为最优修剪后的子树。

4. 质量控制:调查前对调查员及体检员统一培训,统一检查仪器。体格检查每天随机抽取 30 名调查对象复检,与测量员进行对比。血糖测定每 50 个

样品做一套质控系列。数据录入采用双录入方法,并对关键变量设置逻辑核对。

5. 统计学分析:利用 CART 6.0 软件对高血压的危险因素进行分类树分析。患病率的标化采用 2000 年山东省第 5 次人口普查中的农村人口数据。

结 果

1. 一般情况:实际调查 40 个村 20 087 人,共有 16 388 人完成问卷调查和体格检查,实际测量血压者 16 371 人,其中男性 7006 人(42.80%),女性 9365 人(57.20%)。共有高血压患者 7173 人,高血压粗患病率为 43.80%,标化患病率为 37.80%。

2. 高血压危险因素及高危人群分析:表 1 为高血压各危险因素的相对重要性,以年龄为参照,将各危险因素按相对重要性从大到小排列。高血压的主要危险因素为年龄、腰围、BMI、性别、高血压家族史及血糖等。表 2 显示,共有 12 个终结点,其中第 6、7、9、11、12 终结点内高血压患病率均较高,都在 55% 以上,这 5 个终结点内的病例数占总病例数的 64.76%。第 12 终结点为年龄 >59.5 岁,节点内患病率为 66.57%,点内病例数占高血压患病总人数的 40.40%;第 11 终结点为 48.5 岁 < 年龄 ≤ 59.5 岁、腰围 >84.5 cm 者,节点内患病率为 62.76%,点内病例数占高血压总人数的 15.22%。

表1 高血压分类树分析预测因子的相对重要性

变量	积分
年龄	100.00
腰围	36.21
BMI	31.04
性别	5.08
高血压家族史	4.84
血糖	4.70
静态活动时间	2.42
吸烟	1.93
被动吸烟	1.22

表2 山东省中西部农村居民高血压危险因素分类树分析

终结点	节点内规则	点内总人数	点内病例数	点内患病率 (%)	点内病例数占总病例数 (%)	点内病例数占总病例数累计 (%)
12	年龄 >59.5 岁	4353	2898	66.57	40.40	40.40
9	48.5 岁 < 年龄 ≤ 59.5 岁;腰围 ≤ 84.5 cm; BMI ≤ 24.5 kg/m ² ; 血糖 >5.9 mmol/L	39	25	64.10	0.35	40.75
11	48.5 岁 < 年龄 ≤ 59.5 岁;腰围 >84.5 cm	1740	1092	62.76	15.22	55.97
6	腰围 >93.5 cm; 34.5 岁 < 年龄 ≤ 48.5 岁	520	306	58.85	4.27	60.24
7	有高血压家族史; 48.5 岁 < 年龄 ≤ 59.5 岁; 腰围 ≤ 84.5 cm	565	324	57.35	4.52	64.76
3	有高血压家族史; 腰围 >79.7 cm; 39.5 岁 < 年龄 ≤ 48.5 岁	355	182	51.27	2.54	67.29
10	无高血压家族史; 48.5 岁 < 年龄 ≤ 59.5 岁; 腰围 ≤ 84.5 cm; BMI >24.5 kg/m ²	675	328	48.59	4.57	71.87
4	无高血压家族史; 79.7 cm < 腰围 ≤ 93.5 cm; 39.5 岁 < 年龄 ≤ 48.5 岁	1012	364	35.97	5.07	76.94
5	腰围 >93.5 cm; 年龄 < 34.5 岁	140	50	35.71	0.70	77.64
8	48.5 岁 < 年龄 ≤ 59.5 岁; 腰围 ≤ 84.5 cm; BMI ≤ 24.5 kg/m ² ; 血糖 ≤ 5.9 mmol/L	2047	675	32.98	9.41	87.05
2	79.7 cm < 腰围 ≤ 93.5 cm; 年龄 < 39.5 岁	1235	317	25.67	4.42	91.47
1	年龄 ≤ 48.5 岁; 腰围 >79.7 cm	3690	612	16.59	8.53	100.00

高血压的高危人群主要集中在第 6、7、9、11、12、3、10 等高血压患病率较高的终结点内。高血压的高危人群主要为: 高龄者; 年龄较大且血糖值较高者; 年龄较大且腹型肥胖或超重者; 有高血压家族史且年龄较大者; 有高血压家族史、腹型肥胖且年龄较大者。上述高危人群内的病例数占高血压总病例数的 71.87%。

3. 分类树模型拟合情况: 表 3 显示, 高血压危险因素分类树模型中学习集和检验集的灵敏度、特异度均较高, 模型拟合效果较好。

表 3 高血压危险因素分类树模型拟合情况

评价指标	学习集	检验集
灵敏度 (%)	71.87	70.70
特异度 (%)	66.38	65.84
总正确率 (%)	68.79	67.97

讨 论

随着农村居民生活水平的提高和现代生活方式的普及, 高血压已经成为影响农村人群健康的重要公共卫生问题, 近年来农村居民高血压患病率增长速度明显高于城市^[4]。本研究显示山东省中西部地区 25 岁以上农村居民高血压患病率较高, 同为山东省的青岛地区 15 岁以上农村居民高血压患病率为 33.1%^[5], 另一项研究对辽宁省彰武县 15 岁以上农村居民的调查显示, 高血压患病率为 35.0%^[6]。尽管研究对象略有不同, 但这些调查结果都揭示了高血压控制对保障农村居民健康的重要性, 农村疾病控制工作面临着严重的挑战。

本次研究显示高血压的主要危险因素为年龄、腹型肥胖、超重或肥胖、性别、高血压家族史、血糖值较高等, 与多数研究结果一致^[7,8]。有研究显示腹型肥胖对高血压的影响作用大于全身性肥胖^[9], 与本研究结果一致。本研究中腰围和 BMI 是除了年龄之外对高血压影响作用最大的两个因素, 同时也是两个可改变的因素, 对于 BMI 达到超重或是肥胖者以及腹型肥胖者应采取以科学运动、健康膳食等为主要内容的控制措施。

本研究确定高血压高危人群的方法与“2004 年中国高血压防治指南”中危险分层的方法具有相似性, 并计算出不同高危人群中高血压的患病率, 高血压的人群预防可根据不同人群的血压水平和所具有的危险因素确定针对性的预防措施, 高血压患者可根据所处的危险分层参照高血压防治指南的建议采取相应的治疗措施。高血压是一种多因子疾病, 危险因素在个体水平上存在一定的聚集性, 因此, 在制定人群预防措施的基础上还应注重个体化预防措施的作用。

本研究与其他同类研究在对象选择、抽样方法和模型构建等方面都存在着一定差异, 故在进行灵敏度、特异度比较时可比性不是很好, 但经比较, 仍然发现该研究结果与贾崇奇等^[10]所进行的类似研究具有一致性。

CART 分析是一种新的非参数多因素分析方法, 结果直观, 易于理解, 近年来在国外开始广泛应用^[11]。它通过递归分型将人群分为不同的亚群, 能有效改善局部样本的内部同质性; CART 的理论模型不要求因变量和自变量具有某种特定的分布, 并能有效解决变量之间的共线性问题; CART 的树型分析过程与公共卫生人员的思维十分相似, 即先抓住疾病的主要影响因素, 后抓次要因素。有研究利用 CART 和 logistic 回归分析两种方法对同一数据进行对比分析, 结果显示 CART 的分析效果明显好于 logistic 回归分析^[12]。但 CART 也存在缺点, 如 CART 本身是一种大样本统计分析方法, 样本量较小时模型不稳定; 对于内部同质性较好的数据, CART 分析的结果与其他分析方法得到的结果基本一致。尽管如此, CART 在公共卫生领域仍然有较好的应用前景。本次调查虽采取多种方法尽量提高调查对象的参加意愿, 但由于调查地区农村居民外出打工等原因, 被调查者仍存在 18.5% 的无应答者, 对结果会产生一定的偏倚。

(本次调查承蒙山东省卫生厅和有关单位的大力支持, 在此一并志谢)

参 考 文 献

- [1] 刘力生. 2004 年中国高血压防治指南(实用本). 中华心血管病杂志, 2004, 32(12): 1060-1064.
- [2] WHO ISH Hypertension Guidelines Committee. 1999 World Health Organization international society of hypertension guidelines for management of hypertension. Hypertensions, 1999, 17(2): 151-183.
- [3] 中国肥胖问题工作组数据汇总分析协作组. 我国成人体重指数和腰围对相关疾病危险因素异常的预测价值: 适宜体重指数和腰围切点的研究. 中华流行病学杂志, 2002, 23(1): 5-10.
- [4] 唐新华, 徐小玲, 俞志红, 等. 浙江省高血压流行趋势及影响因素分析. 中国慢性病预防与控制, 2000, 8(5): 201-203.
- [5] 李善鹏. 青岛市居民高血压患病率情况分析. 中国健康教育, 2007, 23(8): 593-594, 597.
- [6] 叶景璞, 王海龙, 李辉, 等. 辽宁省彰武县农村人群高血压患病率调查. 中华流行病学杂志, 2003, 24(7): 547-550.
- [7] 刘伟佳, 陈维清, 罗不凡, 等. 广州市城乡居民高血压相关影响因素分析. 中国公共卫生, 2007, 23(1): 34-36.
- [8] 罗雷, 荣荣生, 袁祥, 等. 中国居民高血压病主要危险因素的 Meta 分析. 中华流行病学杂志, 2003, 24(1): 50-53.
- [9] 尹小菁, 林绍强, 李秀敏, 等. 自然人群腰围及体质指数与血压的关系调查. 广东医学, 2005, 26(7): 986-988.
- [10] 贾崇奇, 赵仲堂, 王立华, 等. 高血压危险因素分类树分析. 中国公共卫生, 2003, 19(6): 684-685.
- [11] Lewis RJ. An introduction to classification and regression tree (CART) analysis. The 2000 Annual Meeting of the Society for Academic Emergency Medicine in San Francisco, California, 2000.
- [12] 赵一鸣. 分类与回归树——一种适用于临床研究的统计方法. 北京大学学报(医学版), 2001, 33(6): 562-565.

(收稿日期: 2008-04-17)

(本文编辑: 张林东)