

老年男性人群缺血性心脑血管病 预测模型的建立

陈金宏 吴海云 何昆仑 何耀 秦银河

【摘要】 目的 建立和验证老年人群缺血性心脑血管病(ICVD)预测模型。方法 统计分析来自某保健医院 2003 年 5 月的体检资料、历年住院资料、问卷调查资料和电话回访资料。按照 4:1 的比例随机抽取基线人群,生成建模组和验证组。将验证人群基线资料代入建模人群的回归模型生成预测值。用 ROC 曲线下面积(AUC)检验预测模型的判别能力;用 Hosmer-Lemeshow 检验比较预测率每十分位分组的平均值和实际率来判断预测的准确性;将预测的 6 年 ICVD 发病风险的人群均值与实际观察到的 6 年累计患病率进行比较,计算误差率,验证预测模型群体水平的预测能力。结果 分析样本为 2271 名 >65 岁男性老年人,建模人群 1817 人,验证人群 454 人。把年龄分为两层(≥ 75 岁高龄组; < 75 岁老龄组)建立分层 Cox 比例风险回归模型。老龄组有统计学意义的危险因素是年龄、SBP、血清肌酐(Scr)、空腹血糖(FBG),保护因素是高密度脂蛋白胆固醇(HDL-C);高龄组有统计学意义的危险因素是 BMI、SBP、TC、Scr、FBG,保护因素是 HDL-C。ROC 的 AUC 及其 95%CI 为 0.723(0.687~0.759),将个体按预测 ICVD 累计患病率与实际患病率进行 Hosmer-Lemeshow 检验: $\chi^2=1.43, P=0.786$,模型群体水平预测误差率为 -2.23%,能力较好。结论建立的老年男性人群 ICVD 预测模型判别能力较好,个体预测能力和群体预测能力较为满意。

【关键词】 缺血性心脑血管病;老年人群;预测模型

Establishment of the prediction model for ischemic cardiovascular disease of elderly male population under current health care program CHEN Jin-hong^{1,2}, WU Hai-yun², HE Kun-lun², HE Yao³, QIN Yin-he². 1 The General Hospital of CAPF, Beijing 100039, China; 2 The General Hospital of PLA Corresponding author: HE Yao, Email: yhe301@sina.com; QIN Yin-he, Email: cjh007@126.com This work was supported by grants from the Health Care Fund in the Ministry of Health of China (No. 06H050), Research Grant of the Ministry of Science and Technology of China (No. 2009BAI86B01), Medical Research Grant of the Ministry of Health of PLA in China (No. 06L037) and Beijing Medical Scientific Development Foundation (No. 2007-2039).

【Abstract】 **Objective** To establish and verify the prediction model for ischemic cardiovascular disease (ICVD) among the elderly population who were under the current health care programs. **Methods** Statistical analysis on data from physical examination, hospitalization of the past years, from questionnaire and telephone interview was carried out in May, 2003. Data was from a hospital which implementing a health care program. Baseline population with a proportion of 4:1 was randomly selected to generate both module group and verification group. Baseline data was induced to make the verification group into regression model of module group and to generate the predictive value. Distinguished ability with area under ROC curve and the predictive veracity were verified through comparing the predictive incidence rate and actual incidence rate of every deciles group by Hosmer-Lemeshow test. Predictive veracity of the prediction model at population level was verified through comparing the predictive 6-year incidence rates of ICVD with actual 6-year accumulative incidence rates of ICVD with error rate calculated. **Results** The samples included 2271 males over the age of 65 with 1817 people for modeling population and 454 for verified population. All of the samples were stratified into two layers to establish hierarchical Cox proportional hazard regression model, including one advanced age group (greater than or equal to 75 years old), and another elderly group (less than 75 years old). Data from the statically analysis showed that the risk factors in aged group were age, systolic blood pressure, serum creatinine level, fasting blood glucose

DOI: 10.3760/cma.j.issn.0254-6450.2010.10.021

基金项目:中央保健科研课题(06H050);国家科技支撑计划(2009BAI86B01);军队“十一五”课题(06L037);首都医学发展基金重点项目(2007-2039)

作者单位:100039 北京,武警总医院医务部(陈金宏);解放军总医院老年医学研究所(陈金宏、吴海云、何昆仑、何耀、秦银河)

通信作者:何耀, Email: yhe301@sina.com; 秦银河, Email: cjh007@126.com

level, while protective factor was high density lipoprotein; in advanced age group, the risk factors were body weight index, systolic blood pressure, serum total cholesterol level, serum creatinine level, fasting blood glucose level, while protective factor was HDL-C. The area under the ROC curve (AUC) and 95% CI were 0.723 and 0.687-0.759 respectively. Discriminating power was good. All individual predictive ICVD cumulative incidence and actual incidence were analyzed using Hosmer-Lemshow test, $\chi^2=1.43$, $P=0.786$, showing that the predictive veracity was good. **Conclusion** The stratified Cox Hazards Regression model was used to establish prediction model of the aged male population under a certain health care program. The common prediction factor of the two age groups were: systolic blood pressure, serum creatinine level, fasting blood glucose level and HDL-C. The area under the ROC curve of the verification group was 0.723, showing that the distinguished ability was good and the predict ability at the individual level and at the group level were also satisfactory. It was feasible to using Cox Proportional Hazards Regression Model for predicting the population groups.

【Key words】 Ischemic cardiovascular disease; Aged population; Prediction model

据中国冠心病政策模型预测^[1], 2030 年中国人口将达到 14.6 亿, 其中老年人的比例将达到 16%, 同时, 以冠心病和缺血性脑卒中为代表的缺血性心脑血管病 (ICVD) 的发病及死亡将快速增长。目前少有老年人群 ICVD 预测模型报道, 本研究通过对无 ICVD 疾病史的老年人群进行 6 年的前瞻性研究, 利用分层 Cox 比例风险回归模型初步建立了老年人群 ICVD 的预测模型。

资料与方法

1. 研究队列和数据来源: 基线人群为 2003 年 5 月某保健医院数据库记录在案, 出生于 1938 年 1 月 1 日前 (即基线年龄 ≥ 65 岁) 的所有男性保健对象。数据主要来自该医院 2003 年 5 月的体检资料、1985 年 1 月 1 日至 2009 年 10 月 1 日前的所有住院资料、问卷调查资料、电话回访资料, 数据采集采取查询电子病历和纸质病例相结合、客观指标测量与问卷调查相结合、现场调查与电话回访相结合的方式。基线数据中剔除已有 ICVD 者。

2. 质量控制: 体检采用系统误差较小、准确性较好的仪器进行测定, 同一指标连续测量 3 次后取平均值。由固定人员负责检测。依据仪器操作规则对仪器进行校正。根据专家意见和预调查反馈, 对问卷进行反复设计修改, 在预调查和正式调查前, 编制调查手册, 对目的、方法、重要性及问题进行说明, 尤其对易混淆的问题。对调查员进行规范化培训, 由专人负责考核, 合格后方可参与调查。正式调查阶段建立规范性的操作流程, 尽可能把数据收集齐全, 不明确者及时电话回访。

3. ICVD 界定和基线危险因素确立: 依照我国“十五”科技攻关课题“冠心病、脑卒中综合危险度评估及干预方案的研究”的方法^[2], 本研究将 ICVD 定义为冠心病和缺血性脑卒中的总和。ICVD 事件是指因 ICVD 发病或死亡。观察人员的随访主要依据

医院诊断, 根据 WHO-MONICA 计划的建议及我国“十五”科技攻关课题的研究结果^[3], 包括①确诊的急性心肌梗死; ②经硝酸甘油等药物治疗后症状缓解但未确诊者, 不含无症状的心肌梗死; ③大脑前动脉栓塞导致的脑梗死及缺血性脑梗死; ④出现相应临床症状但未确诊者, 不含短暂性脑缺血发作、血液病导致的脑梗死、脑肿瘤或创伤导致的脑梗死; ⑤由上述疾病导致死亡者。

危险因素包括基线时年龄、体重指数 (BMI)、SBP、血总胆固醇 (TC)、血甘油三酯 (TG)、高密度脂蛋白胆固醇 (HDL-C)、血清肌酐 (Scr)、血载脂蛋白 A1 (ApoA1)、糖尿病、吸烟。

4. 模型的建立和验证: 以人年作为观察时间, 使用 Cox 比例风险回归模型进行多因素分析。各危险因素的回归系数为 β_i , 在控制其他危险因素不变的情况下, 危险度 $HR=e^{\beta_i}$ 。如 $\beta_i > 0$, 则说明该变量是 ICVD 事件的危险因素, 反之则可认为该变量是 ICVD 事件的保护因素, $\alpha=0.05$ 。

5. 统计学分析: 利用建模人群的拟合结果预测验证人群 6 年 ICVD 发病风险, 并与实际观察的累积患病率进行比较后绘制 ROC 曲线, 采用文献 [4] 方法计算 ROC 曲线下面积 (AUC); 将个体预测的 10 年 ICVD 发病危险的十分位数进行分组, 比较每组预测概率的均值和实际累积患病率, 并进行 Hosmer-Lemshow 检验判断其预测准确性^[5]。将预测的 6 年 ICVD 发病风险的人群均值与实际观察到的 6 年累积患病率进行比较, 计算误差率, 误差率 = [(人群预测率均值 - 实际累积患病率) / 实际累积患病率] $\times 100\%$, 用来验证预测模型在群体水平的预测能力。采用 EpiData 3.1 软件录入系统, 采用 Stata 9/SE 软件内置的 Cox 比例分析模块进行数据分析和建模。

结 果

1. 数据整理: 2003 年 5 月某保健医院进入基线

观察的老年保健对象共计 4060 人,为减少偏性,剔除女性、已患有 ICVD 者,共有 2271 人进入研究队列,累计观察 12 852.8 人年。因该老年保健人群的依从性好且联系方便,本研究无失访数据。在观察期内,ICVD 住院 523 人(其中冠心病住院 289 人,累计患病率为 12.73%;缺血性脑卒中住院 234 人,累计发病率为 10.30%),因其他原因住院 1499 人(66.01%),未住院 249 人(10.96%);在观察期内,因 ICVD 死亡 81 人(其中冠心病死亡 63 人,累计患病率为 2.73%;缺血性脑卒中死亡 28 人,累计患病率为 1.23%),因其他原因死亡 370 人(16.29%)。共发生 ICVD 事件 535 例。将 2271 人按照 4:1 的比例随机抽取生成建模人群和验证人群,其中建模人群 1817 人,验证人群 454 人。

2. 危险因素基线特征:危险因素都呈正偏态分布。除年龄分布接近标准正态峰,其余指标都是呈尖峭峰分布(表 1)。

表 1 老年人群 ICVD 危险因素的基线特征

指标	极小值	极大值	\bar{x}	M	s	偏度系数	峰度系数
年龄(岁)	66.00	100.00	75.90	75.00	5.90	0.59	3.12
BMI(kg/m ²)	15.05	37.76	23.36	23.11	2.63	0.64	4.58
SBP(mm Hg)	87.00	220.00	138.60	136.15	14.54	0.75	5.79
TC(mmol/L)	1.80	24.37	4.79	4.64	1.58	3.57	31.34
TG(mmol/L)	0.49	12.88	1.54	1.37	0.98	4.18	33.47
HDL-C(mmol/L)	0.33	7.77	1.89	1.33	1.45	1.81	4.97
Scr(μ mol/L)	40.03	177.90	75.28	75.06	16.22	1.64	10.90
Apoa1(μ mol/L)	5.51	88.15	52.81	54.51	10.62	-1.20	7.41
餐前血糖(mmol/L)	3.70	15.93	6.64	5.48	2.74	1.56	4.58

注:基线人群有 580 人吸烟,吸烟率为 25.54%; 1 mm Hg=0.133 kPa

3. 建模组和验证组的可比性:建模组 1817 人,在观察期内有 415 人(22.84%)发生 ICVD 事件;测试组 454 人,在观察期内有 415 人(26.43%)发生 ICVD 事件($\chi^2=2.60, P=0.107$)。除 Apoa1 外,各危险因素在两组内的分布较为均匀,因连续性变量呈偏态分布,故采用非参数检验 Kruskal-Wallis test,分类变量采用 χ^2 检验(表 2)。

4. 建模组 Cox 比例风险回归模型:危险因素均

表 2 建模组和验证组的可比性

指标	测试组			验证组			P 值
	\bar{x}	M	偏度系数	\bar{x}	M	偏度系数	
年龄(岁)	75.9	75.00	0.592	75.90	75.00	0.566	0.981
BMI(kg/m ²)	23.36	23.11	0.665	23.31	23.11	0.541	0.952
SBP(mm Hg)	138.00	136.00	0.829	139.00	136.00	0.423	0.216
TC(mmol/L)	4.69	4.55	6.469	4.79	4.53	6.434	0.780
TG(mmol/L)	1.56	1.37	7.621	1.52	1.35	3.136	0.656
HDL-C(mmol/L)	5.39	1.33	9.811	5.59	1.34	8.554	0.769
Scr(μ mol/L)	72.96	75.06	5.233	74.55	75.06	3.858	0.318
Apoa1(μ mol/L)	72.96	75.06	5.233	74.55	75.06	3.858	0.036
餐前血糖(mmol/L)	7.28	5.43	6.43	7.43	5.53	5.553	0.262
吸烟	-	-	-	-	-	-	0.552
观察时间(年)	5.21	6.42	-1.352	5.09	6.42	-1.142	0.249

注:吸烟率测试组为 25.81%,验证组为 24.45%

呈正偏态分布。除年龄分布接近标准正态峰,其余指标都呈尖峭峰分布。经多因素 PH 检验,当取原始值作为自变量时,TC、TG、Scr、Apoa1、吸烟满足 PH 假定,而将除年龄之外其余不符合正态分布的连续变量仿照 Framingham 研究取自然对数值后进行 PH 检验^[6],只有年龄不符合 PH 假定,若再把年龄分为两层(≥ 75 岁高龄组和 < 75 岁老龄组),每层的自变量均满足 PH 假定,因此本研究拟按年龄段作为分层因素建立分层 Cox 比例风险回归模型。结果显示:老龄组有统计学意义的危险因素是年龄、SBP、Scr、FBG,保护因素是 HDL-C;高龄组有统计学意义的危险因素是 BMI、SBP、TC、Scr、FBG,保护因素是 HDL-C(表 3)。

5. 模型的判别能力和预测准确性:将验证组基线资料代入分层 Cox 比例风险回归模型预测该人群 6 年间 ICVD 的发病风险,并与实际的 6 年发病率比较,绘制 ROC 曲线(图 1)。ROC 的 AUC 及 95%CI 为 0.723(0.687 ~ 0.759),判别能力较好。

将个体按预测 ICVD 累计患病率从小到大排序,并按十分位数分组,比较每一组预测概率的平均值和实际的累计患病率(图 2),Hosmer-Lemshow 检验: $\chi^2=1.43, P=0.786$ 。除第 10 组预测率高于实际率外,其他组都是预测率略低于实际率。整体上看,

表 3 按年龄段分组 Cox 比例风险回归模型结果

危险因素	老龄组(< 75 岁)				高龄组(≥ 75 岁)			
	β	s_e	HR 值(95%CI)	P 值	β	s_e	HR 值(95%CI)	P 值
年龄	0.130	0.038	1.139(1.057 ~ 1.228)	0.001	-0.028	0.018	0.972(0.937 ~ 1.008)	0.057
ln(BMI)	0.625	0.708	1.868(0.466 ~ 7.478)	0.377	1.728	0.549	5.630(1.918 ~ 16.525)	0.002
ln(SBP)	0.035	0.004	1.035(1.027 ~ 1.043)	0.000	0.031	0.003	1.032(1.025 ~ 1.038)	0.000
ln(TC)	0.155	0.285	1.167(0.668 ~ 2.042)	0.587	0.526	0.253	1.693(1.031 ~ 2.778)	0.037
ln(TG)	-0.001	0.154	0.999(0.739 ~ 1.352)	0.996	-0.003	0.134	0.997(0.734 ~ 1.260)	0.887
ln(HDL-C)	-1.088	0.213	0.337(0.222 ~ 0.511)	0.000	-0.792	0.153	0.453(0.336 ~ 0.611)	0.000
ln(Scr)	0.358	0.116	1.430(1.140 ~ 1.794)	0.002	0.491	0.122	1.634(1.286 ~ 2.076)	0.000
ln(Apoa1)	0.176	0.294	1.192(0.670 ~ 2.121)	0.550	0.375	0.277	1.455(0.846 ~ 2.504)	0.175
ln(FBG)	0.870	0.169	2.387(1.713 ~ 3.325)	0.000	0.698	0.122	2.011(1.584 ~ 2.552)	0.000
吸烟	-0.260	0.194	0.771(0.528 ~ 1.127)	0.180	-0.047	0.145	0.954(0.718 ~ 1.268)	0.746

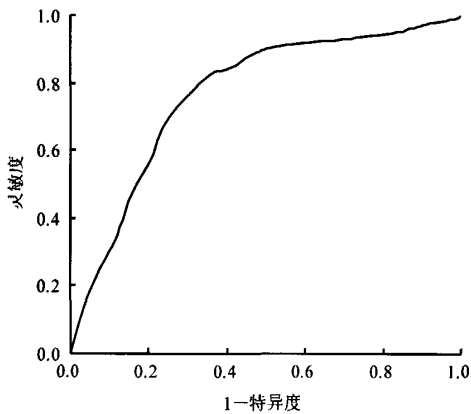


图1 测试人群预测结果的ROC曲线

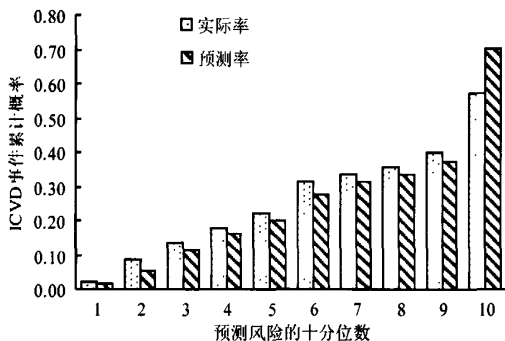


图2 测试人群预测ICVD累计患病率的实际率和预测率比较

模型预测比较准确。

6. 群体水平的预测能力: 预测人群累计患病率为26.43%, 预测患病率的平均值为25.84%, 误差率为-2.23%, 群体预测能力较好。

讨论

目前, 国内外已有大量成熟的心脑血管病预测模型, 其设计思路都是通过前瞻性研究, 根据基线危险因素对未来若干年内的发病风险进行预测, 其中最为成熟的是弗明汉研究。我国“十五”科技攻关课题“冠心病、脑卒中综合危险度评估及干预方案的研究”仿照弗明汉研究, 初步建立了适合我国人群疾病特点的缺血性心血管病10年发病危险评估方法和简易评估工具^[2]。

已有的国内外研究其特点是样本人群年龄29~75岁^[3]; 绝大部分研究排除基线已患有心脑血管疾病者; 前瞻性队列研究, 观察时间大部分为10年; 使用的模型大部分为Cox比例风险回归模型; 模型的外推能力普遍较弱, 存在“本地优势”(home advantage)。由于老年人身体基础较差, 原有疾病

多, 预测难度较大, 建立老年人ICVD的预测模型较为困难, 弗明汉模型在老年人群中的应用很不理想^[1]。为了减少偏性, 本研究在基线时剔除了有既往ICVD病史者, 通过分层Cox比例风险回归, 获得老龄组和高龄组共同的危险因素是SBP、FBG、Scr, 共同的保护因素是HDL-C, 提示控制血压、血糖、Scr, 提高血清HDL-C浓度是预防未来ICVD发病的重要对策。

本研究通过转换自变量, 将验证组基线资料代入基于建模组建立的分层Cox比例风险回归模型, 预测该人群6年间ICVD的发病风险, 其ROC的AUC为0.723, 低于文献[8]的0.796, 但也达到基本的判别能力。另外, 验证组基线资料回代, 说明模型预测的准确性和群体水平的预测能力较好。

本研究的局限是从某医院所有保健人群中划分出建模组和验证组, 难以避免“本地优势”, 同时也说明采用既定模型预测同类人群的可行性。为进一步验证该模型的科学性和外推作用, 需要寻找其他类似保健人群进行前瞻性观察和预测。

参考文献

- [1] Moran A, Zhao D, Gu DF, et al. The future impact of population growth and aging on coronary heart disease in China: projections from the coronary heart disease policy model-China. BMC Public Health, 2008, 8:394.
- [2] The Collaborative Research Group of the National 10th Five Year Plan Project "A Study on Evaluation and Intervention of the Coronary Heart Disease and Stroke Integrated Risk". A study on evaluation of the risk of ischemic cardiovascular diseases in Chinese and the development of simplified tools for the evaluation. Chin J Cardiol, 2003, 31(12):893-901. (in Chinese) 国家“十五”科技攻关“冠心病、脑卒中综合危险度评估及干预方案的研究”课题组. 国人缺血性心血管病的评估方法及简易评估工具的开发研究. 中华心血管病杂志, 2003, 31(12):893-901.
- [3] Voss R, Cullen P, Schulte H, et al. Prediction of risk of coronary events in middle-aged men in the Prospective Cardiovascular Munster Study (PROCAM) using neural networks. Int J Epidemiol, 2002, 31:1253-1262.
- [4] Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristics curves derived from the same cases. Radiology, 1983, 19:839-843.
- [5] Hosmer DW, Hosmer T, Le Cessie S, et al. A comparison of goodness-of-fit tests for the logistic regression model. Stat Med, 1997, 16(9):965-980.
- [6] Anderson KM, Odell PM, Wilson PWF, et al. Cardiovascular disease risk profiles. Am Heart J, 1991, 121:293-298.
- [7] D'Agostino RB, Ramachandran SV, Pencina MJ, et al. General cardiovascular risk profile for use in primary care: The Framingham Heart Study. Circulation, 2008, 117:743-753.
- [8] Wu Y, Liu X, Li X, et al. Estimation of 10-year risk of fatal and nonfatal ischemic cardiovascular disease in Chinese adults. Circulation, 2006, 114:2217-2225.

(收稿日期:2010-04-07)

(本文编辑:张林东)