

基于LASSO回归的塔城地区人群冠状动脉狭窄程度 预测模型构建及验证

徐以康¹,刘蕾²,刘丽敏¹,马晶茹¹,王嘉钰²,马军³,甄紫伊⁴

(1. 沈阳医学院附属第二医院心内科,沈阳 110001; 2. 辽宁中医药大学护理学院社区护理教研室,沈阳 110033; 3. 塔城市人民医院心内科,新疆 塔城 834700; 4. 沈阳医学院公共卫生学院流行病与卫生统计学教研室,沈阳 110034)

摘要 **目的** 分析新疆塔城地区人群冠状动脉中重度狭窄的危险因素,构建并验证冠状动脉狭窄程度列线图风险预测模型。**方法** 回顾性收集2021年1月至2023年6月于塔城市人民医院心内科住院治疗的629例患者的临床资料。用R语言软件将所有患者的临床资料纳入LASSO回归进行危险因素初筛。按7 : 3的比例将629例患者随机分为训练组(440例)和验证组(189例)。训练组数据用于模型构建,以冠状动脉狭窄程度为因变量,将LASSO回归筛选出的变量作为自变量纳入logistic回归建模。验证组用于模型验证。基于logistic分析结果,用R语言软件构建冠状动脉狭窄程度预测的可视化列线图。应用曲线下面积(AUC)、临床决策曲线分析(DCA)及校准曲线评价模型的区分度、临床效用和校准度。**结果** 年龄、非汉族、高血压、高脂血症、脑血管病史是发生冠状动脉中重度狭窄的危险因素,纳入风险预测模型。训练组和验证组AUC分别为0.905(95%CI:0.790~0.863)和0.864(95%CI:0.744~0.861),校准曲线预测值与实际值一致度较高(训练组和验证组Brier得分分别为0.03和0.14),模型的预测性能好,DCA结果提示本模型具有临床净获益。**结论** 本研究所构建塔城地区人群冠状动脉狭窄程度风险预测模型具有良好的预测性能,可为筛查冠状动脉中重度狭窄患者提供简便易行、经济、易推广的评估工具。

关键词 冠状动脉狭窄程度;列线图;预测模型;LASSO回归

中图分类号 R543.3 文献标志码 A 文章编号 0258-4646(2025)02-0139-06

网络出版地址 <https://link.cnki.net/urlid/21.1227.R.20250115.1125.018>

DOI:10.12007/j.issn.0258-4646.2025.02.008

Construction and validation of a prediction model for coronary artery stenosis based on LASSO regression

XU Yikang¹, LIU Lei², LIU Limin¹, MA Jingru¹, WANG Jiayu², MA Jun³, ZHEN Ziyi⁴

(1. Department of Cardiology, The Second Affiliated Hospital of Shenyang Medical College, Shenyang 110001, China; 2. Department of Community Nursing, School of Nursing, Liaoning University of Traditional Chinese Medicine, Shenyang 110033, China; 3. Department of Cardiology, Tacheng People's Hospital, Tacheng 834700, China; 4. Department of Epidemiology and Health Statistics, School of Public Health, Shenyang Medical College, Shenyang 110034, China)

Abstract **Objective** To analyze the risk factors for moderate-to-severe coronary artery stenosis in the population of Tacheng, Xinjiang Uygur Autonomous Region, and to construct and verify a nomogram prediction model for the degree of coronary artery stenosis. **Methods** We retrospectively selected 629 patients who were hospitalized in the Cardiovascular Department of Tacheng People's Hospital from January 2021 to June 2023. Using R language software, the sociodemographic data, disease-related data, and various laboratory indicators of the 629 patients were included in the initial screening of risk factors for use in the LASSO regression analysis using a random number table method. The 629 patients were divided into a training group ($n = 440$) and a validation group ($n = 189$) in a 7 : 3 ratio. Data from the training group were used for model construction, with the degree of coronary artery stenosis as the dependent variable, and the variables selected by LASSO regression as independent variables in the logistic regression model. The validation group was used for model validation. Based on the results of the logistic regression analysis, a visual nomogram for predicting the degree of coronary artery stenosis was constructed using R language software. The discriminability, calibration, and clinical utility of the model were evaluated using the area under the receiver operating characteristic curve (AUC), a calibration curve, and decision curve analysis (DCA). **Results** Age, non-Han ethnicity, hypertension, hyperlipidemia, and a history of cerebrovascular disease were risk factors for mode-

基金项目:辽宁省科学技术计划科技援疆、援藏医疗专项(2023-MS-19);

辽宁省教育厅基本科研项目(LJKMZ20221798)

作者简介:徐以康(1982-),男,副主任医师,硕士.

通信作者:刘蕾, E-mail: liulei0428@sina.com

收稿日期:2024-05-06

网络出版时间:2025-01-15 17:15:29

rate-to-severe coronary artery stenosis and were included in the risk prediction model. The AUC of the training group and the validation group were 0.905 (95%CI:0.790–0.863) and 0.864 (95%CI:0.744–0.861), respectively. The predicted values of the calibration curve were consistent with the actual values (Brier scores of the training and validation group:0.03 and 0.14, respectively). The predictive performance of the model was good, and the DCA results indicated that the model had net clinical benefits. **Conclusion** The risk prediction model for coronary artery stenosis in the population of the Tacheng area constructed in this study has good predictive performance and can provide a simple, feasible, economical, and easy-to-promote evaluation tool for medical personnel to screen patients with moderate-to-severe coronary artery stenosis.

Keywords degree of coronary artery stenosis; nomogram; prediction model; LASSO regression

全球缺血性心脏病负担报告^[2]显示,全球2019年约有1.97亿冠状动脉粥样硬化性心脏病(coronary heart disease, CHD)患者,估计914万人因患CHD死亡。《中国心血管健康与疾病报告2022概要》^[3]的数据表明,我国CHD患者约1 139万,且其患病率、死亡率呈上升趋势^[4]。最新的研究^[5]显示,新疆地区居民的缺血性心脏病疾病负担是上海等地区的4倍,而西北地区的心血管疾病医疗资源及医院疾病防治能力却相对落后^[6-7]。研究^[8-10]显示,尽早对高危人群进行评估和筛查,消除人群的不健康行为,可有效降低CHD患病率。本研究拟探讨塔城市人民医院住院治疗患者的冠状动脉中重度狭窄(狭窄程度>50%)的危险因素,应用logistic回归建立风险预测模型,以期评估CHD风险、选择治疗方案、优化医疗资源提供科学依据。

1 材料与方法

1.1 研究对象

回顾性选取2021年1月至2023年6月在塔城市人民医院心内科住院并行冠状动脉造影检查的629例冠状动脉中重度狭窄患者作为研究对象。依据随机数字表法,按7 : 3比例随机分为训练组($n = 440$)和验证组($n = 189$)。纳入标准:年龄 ≥ 18 岁;依据冠状动脉造影结果确定冠状动脉中重度狭窄,即至少有1根主要冠状动脉的狭窄程度 $>50\%$ ^[11]。排除标准:患有风湿性心脏病、严重心脏瓣膜病、肺功能衰竭、凝血功能障碍及恶性肿瘤等,临床资料信息不全。本研究已获得沈阳医学院附属第二医院伦理委员会批准(2023-48)。所有患者知情同意。

1.2 资料收集

查阅患者的电子病历,收集其一般资料、疾病相关资料和实验室及辅助检查结果。社会人口学资料包括性别、年龄、民族、饮酒史、吸烟史、腰围、体重指数(body mass index, BMI)、婚姻状况。疾病相关资料包括入院时的心率和血压,糖尿病、慢性肾功能不全、高脂血症、心房颤动/心房扑动、高血压、心力衰竭、卒中或心肌梗死病史。实验室检查指标包括血脂、心钠素、血肌酐、血离子、超敏C反应蛋白、尿酸、尿素氮。辅助检查有颈动脉超声及选择性冠状动脉造影。

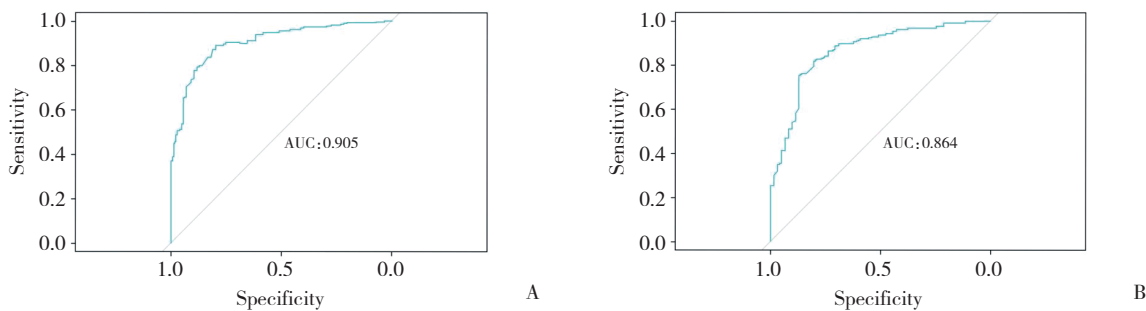
1.3 统计学分析

采用SPSS 26.0软件和R软件(4.0.5)进行统计分析。计量资料用 $\bar{x} \pm s$ 表示,组间比较符合正态分布者使用 t 检验,不符合正态分布者使用秩和检验。计数资料用构成比(%)表示,组间比较使用 χ^2 检验。 $P < 0.05$ 为差异有统计学意义。

采用LASSO回归精准筛选出危险因子,训练组使用logistic回归构建预测模型,以列线图可视化形式直观呈现,用验证组进行内部验证。应用受试者操作特征(receiver operating characteristic, ROC)曲线下面积(area under the curve, AUC)来评估模型的预测效能。AUC $>0.5\sim 0.7$,模型预测性能一般;AUC $>0.7\sim 0.9$,模型预测性能良好;AUC >0.9 ,模型预测性能优秀^[12]。应用Brier评分评价模型校准度,Brier评分越接近0,代表模型的校准度越精准^[13]。应用临床决策曲线分析(decision curve analysis, DCA)验证模型在临床实践中的实用性。

2 结果

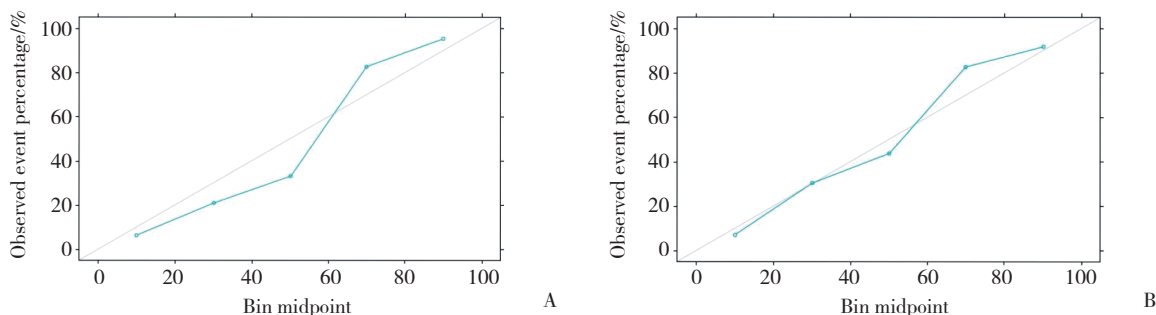
2.1 基本情况



A, training group; B, validation group.

图3 训练组与验证组预测模型ROC曲线

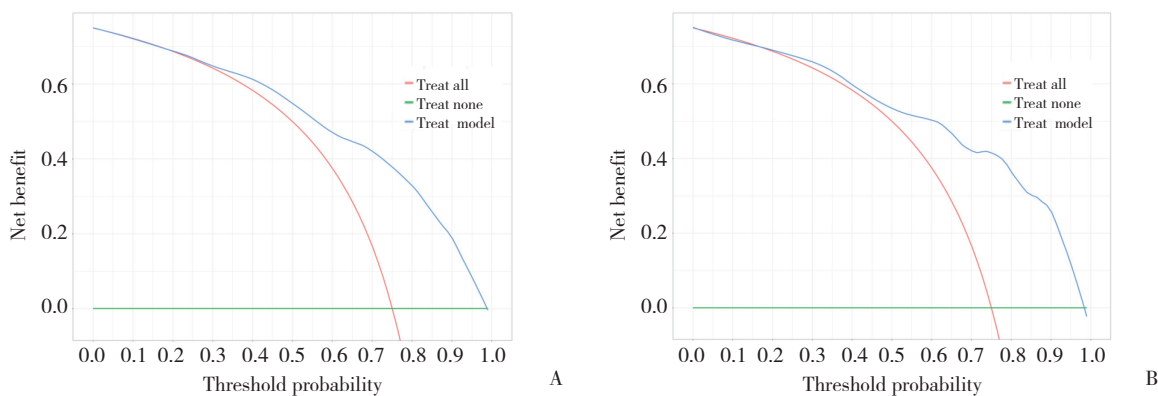
Fig.3 ROC curves of the prediction model in the training and validation groups



A, training group; B, validation group.

图4 训练组与验证组预测模型校准曲线

Fig.4 Calibration curves of the prediction model in the training and validation



A, training group; B, validation group.

图5 训练组与验证组DCA曲线

Fig.5 DCA curves of the prediction model in the training and validation groups

3 讨论

本研究发现,年龄大、非汉族、脑血管病史、高血压、高脂血症为塔城地区人群冠状动脉中重度狭窄的危险因素。美国心血管疾病研究机构数据^[15]显示,冠状动脉发生狭窄的风险随年龄增加而增加,呈正相关。我国居民的CHD患病风险也随着年龄增长而上升^[16]。随着血管老化,动脉血管内膜发生不

同程度的损伤,动脉管壁内脂质含量增加,均可导致进展性的动脉粥样硬化^[17]。研究^[19-21]显示,新疆非汉族居民的CHD患病率高于全国的平均患病水平,非汉族居民可能存在更多CHD危险因素,更容易罹患心血管方面疾病,本研究结果也证实了以上结果。塔城地区少数民族喜食油脂含量较高的食物,如腌肉和奶茶,易发生冠状动脉狭窄^[22-23]。研究^[24]显示,高血压是罹患CHD的重要危险因素,其

次是高脂血症,尤其是低密度脂蛋白胆固醇(low-density lipoprotein cholesterol, LDL-C)高。高血压患者常合并高脂血症,二者并存能显著增加CHD的发生风险^[25]。患者长期处于血压升高的状态,冠状动

脉内皮细胞增生,管腔内径缩小,从而引发心肌供血不足^[26]。血管中LDL-C积聚,内皮细胞被活化,血管壁产生炎症、坏死、增殖和钙化,最终产生动脉粥样硬化斑块^[27-28],冠状动脉管腔变狭窄。

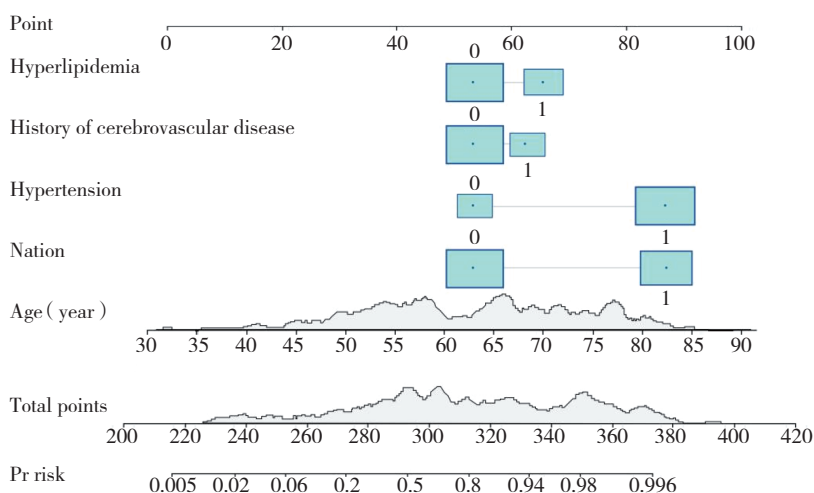


图6 冠状动脉狭窄程度预测列线图

Fig.6 Nomograms for predicting the degree of coronary artery stenosis

本研究构建的新疆塔城地区人群冠状动脉狭窄程度预测模型中,训练组和验证组的AUC值分别为0.905 (95%CI:0.790~0.863)和0.864 (95%CI:0.744~0.861),提示模型具有良好的区分度,校准曲线与理想曲线接近,模型校准度表现良好,训练组和验证组Brier得分分别为0.03和0.14,可见模型预测值与真实值具有高度一致性,提示本研究构建的预测模型预测性能良好。训练组与验证组DCA结果均显示,本研究构建的预测模型能够为实际临床应用带来净获益,可为新疆塔城地区人群冠状动脉狭窄风险的评估提供有力工具。

本研究的局限性在于为单中心研究,且样本量较小,未来可进一步拓展样本规模,涵盖更多中心,纳入更全面的危险因素,并对模型进行外部验证,以提高模型的精准度和通用性。

综上所述,本研究构建了新疆塔城地区人群冠状动脉狭窄程度列线图预测模型,采用科学的赋值方法对各危险因素进行量化评估,可高效评估冠状动脉狭窄程度,有助于早期识别本地区存在冠状动脉中重度狭窄危险因素的人群,从而及早实施干预措施,降低不良结局的发生率。

参考文献:

- [1] 中华医学会,中华医学会杂志社,中华医学会全科医学分会,等. 稳定性冠心病基层诊疗指南(2020年)[J]. 中华全科医师杂志, 2021, 20(3): 265-273. DOI: 10.3760/cma.j.cn114798-20210120-00079.
- [2] SAFIRI S, KARAMZAD N, SINGH K, et al. Burden of ischemic heart disease and its attributable risk factors in 204 countries and territories, 1990–2019 [J]. Eur J Prev Cardiol, 2022, 29 (2): 420-431. DOI: 10.1093/eurjpc/zwab213.
- [3] 中国心血管健康与疾病报告编写组. 中国心血管健康与疾病报告2022概要[J]. 中国循环杂志, 2023, 38 (6): 583-612. DOI: 10.3969/j.issn.1000-3614.2023.06.001.
- [4] 国家卫生健康委员会. 中国卫生健康统计年鉴—2020[M]. 北京: 中国协和医科大学出版社, 2020.
- [5] 李镒冲, 刘世伟, 曾新颖, 等. 1990—2016年中国及省级行政区心血管病疾病负担报告[J]. 中国循环杂志, 2019, 34 (8): 729-740. DOI: 10.3969/j.issn.1000-3614.2019.08.001.
- [6] 张斌, 张岩, 赵冬, 等. 我国分省心血管病救治的现状评价: 中国心血管健康指数解析[J]. 疾病监测, 2020, 35 (5): 401-405. DOI: 10.3784/j.issn.1003-9961.2020.05.008.
- [7] 姜莹莹, 毛凡, 张岩, 等. 我国各省心血管病相关健康资源分布评价[J]. 疾病监测, 2020, 35 (5): 387-393. DOI: 10.3784/j.issn.1003-9961.2020.05.006.
- [8] 应旭华, 李松涛, 陈芸, 等. 玉环市心血管病高危人群干预效果调查研究[J]. 中国预防医学杂志, 2020, 21 (11): 1210-1215. DOI: 10.16506/j.1009-6639.2020.11.011.
- [9] ABBARA S, BLANKE P, MAROULES CD, et al. SCCT guidelines for the performance and acquisition of coronary computed tomographic