

糖尿病心血管并发症预测模型的构建



孙 源^{1,2}, 刘 括^{1,2}, 李冰潇^{1,2}, 温馥源^{1,2}, 李盼弟^{1,2}, 杨晓俊^{1,2}, 屈艾彬^{1,2}, 张 玲^{1,2}

1. 首都医科大学公共卫生学院 (北京 100069)
2. 临床流行病学北京市重点实验室 (北京 100069)

【摘要】目的 筛选糖尿病心血管并发症的危险因素并构建预测模型, 为早期预防和减缓进展提供依据。**方法** 基于京津冀社区自然人群慢性病队列, 选取基线 (2017—2019 年) 糖尿病患者作为研究对象, 根据自报患病时间先后顺序判定是否发生心血管并发症作为结局指标。按 7:3 随机分成训练集和测试集分别构建和验证模型。使用 Lasso 结合 Logistic 回归模型方法筛选危险因素, 将其纳入多因素 Logistic 回归模型构建糖尿病心血管并发症发生风险的预测模型, 并绘制列线图进行可视化。绘制受试者工作特征曲线并计算曲线下面积, Hosmer-Lemeshow 拟合优度检验和绘制校准曲线评估校准度, 对预测模型进行评价和验证。**结果** 共纳入 813 例 2 型糖尿病患者, 平均年龄为 (62.6 ± 10.4) 岁, 其中训练集 569 例, 测试集 244 例, 两组患者除超敏 C 反应蛋白水平具有差异 ($P=0.028$) 外, 其他基本特征差异均无统计学意义。多因素 Logistic 回归分析显示, 糖尿病心血管并发症的危险因素包括年龄 [OR=1.040, 95%CI (1.010, 1.073), $P=0.010$]、高血压 [OR=2.211, 95%CI (1.263, 3.975), $P=0.006$]、糖尿病病程 [OR=1.063, 95%CI (1.028, 1.099), $P < 0.001$]、空腹血糖水平 [OR=1.186, 95%CI (1.075, 1.309), $P=0.001$]、血脂异常 [OR=2.051, 95%CI (1.167, 3.583), $P=0.012$]、心血管病家族史 [OR=2.794, 95%CI (1.650, 4.774), $P < 0.001$] 和吸烟 [OR=1.975, 95%CI (1.133, 3.462), $P=0.017$]; 而保护因素为血清胆红素 [OR=0.940, 95%CI (0.889, 0.991), $P=0.027$]。列线图显示 2 型糖尿病患者可根据模型中 8 个预测因素的动态变化计算出发生糖尿病心血管并发症的概率。训练集和测试集 ROC 曲线下面积分别为 0.803 和 0.820, Hosmer-Lemeshow 检验 P 值分别为 0.776 和 0.554, 校准曲线与理想曲线的平均绝对误差为 0.013, 表明预测模型区分度和校准度均较好。**结论** 本研究基于社区自然人群构建的糖尿病心血管并发症风险预测模型效果较好, 为糖尿病患者心血管并发症的早期预测预警提供了便利可行的工具。

【关键词】 糖尿病; 心血管并发症; Lasso 回归; 预测模型

Establishment of predictive model for cardiovascular complication of diabetes mellitus

SUN Yuan^{1,2}, LIU Kuo^{1,2}, LI Bingxiao^{1,2}, WEN Fuyuan^{1,2}, LI Pandi^{1,2}, YANG Xiaojun^{1,2}, QU Aibin^{1,2}, ZHANG Ling^{1,2}

1. School of Public Health, Capital Medical University, Beijing 100069, China
2. Beijing Key Laboratory of Clinical Epidemiology, Beijing 100069, China

DOI: 10.12173/j.issn.1004-5511.202310094

基金项目: 国家重点研发计划“精准医学研究”重点专项 (2016YFC0900603); 国家自然科学基金青年科学基金项目 (81602908); 北京市教委科研计划科技一般项目 (KM202210025028)

通信作者: 张玲, 博士, 教授, 博士研究生导师, Email: zlllyepi@ccmu.edu.cn

Corresponding author: ZHANG Ling, Email: zlllyepi@ccmu.edu.cn

【Abstract】Objective To screen the risk factors of cardiovascular complications of diabetes mellitus, and build a predictive model to provide a basis for early prevention and slow down the progression of the disease. **Methods** Based on the chronic disease cohort of the natural population in the Beijing-Tianjin-Hebei community, the baseline diabetic patients from 2017 to 2019 were selected as the study objects, and the occurrence of cardiovascular complication was determined as the outcome index based on the chronological order of the self-reported illness. The model was randomly divided into a training set and a test set at 7 : 3 to build and verify the model, respectively. Lasso-Logistic regression model was used to screen risk factors, which were incorporated into the multi-factor Logistic regression model to build a prediction model for the risk of cardiovascular complication of diabetes, and a nomogram was drawn for visualization. The receiver operating characteristic curve (ROC) was drawn and the area under the curve (AUC) was calculated. The Hosmer-Lemeshow tests and plotting calibration curve to assess the calibration degree, and the prediction model was evaluated and verified. **Results** A total of 813 patients with type 2 diabetes were included, with an average age of (62.6±10.4) years, including 569 patients in the training set and 244 patients in the test set. There were differences in the level of hypersensitive C-reactive protein between the two groups ($P=0.028$), and there were no statistically significant differences in other basic characteristics. Multivariate Logistic regression analysis showed that the risk factors for cardiovascular complications of diabetes mellitus included age [OR=1.040, 95%CI (1.010, 1.073), $P=0.010$], hypertension [OR=2.211, 95%CI (1.263, 3.975), $P=0.006$], duration of diabetes [OR=1.063, 95%CI (1.028, 1.099), $P<0.001$], level of fasting blood glucose [OR=1.186, 95%CI (1.075, 1.309), $P=0.001$], dyslipidemia [OR=2.051, 95%CI (1.167, 3.583), $P=0.012$], family history of cardiovascular disease [OR=2.794, 95%CI (1.650, 4.774), $P<0.001$] and smoking [OR=1.975, 95%CI (1.133, 3.462), $P=0.017$]; the protective factor was level of serum bilirubin [OR=0.940, 95%CI (0.889, 0.991), $P=0.027$]. The nomogram shows that patients with type 2 diabetes can calculate the probability of developing cardiovascular complications of diabetes based on the dynamic changes in the eight predictors in the model. The AUC of the training set and the test set were 0.803 and 0.820, the P-values of the Hosmer-Lemeshow test were 0.776 and 0.554, respectively, and the average absolute error between the calibration curve and the ideal curve was 0.013, which proved that the differentiation and calibration degree of the prediction model were good. **Conclusions** In this study, the risk prediction model of diabetic cardiovascular complication built on the basis of community natural population has good effect, and provides a convenient and feasible tool for the early prediction and early warning of cardiovascular complication in diabetic.

【Keywords】 Diabetes; Cardiovascular complication; Lasso regression; Predictive model

2 型糖尿病是由组织胰岛素抵抗和相对胰岛素缺乏引起的高血糖^[1]。2019 年全球男性年龄标准化患病率为 6.0%，女性为 5.0%^[2]，2021 年全球约有 5.37 亿人患有 2 型糖尿病，预计到 2045 年将达到 7.83 亿^[3]。长期罹患糖尿病可能发生多器官或系统长期受损的慢性并发症，其中心血管病

是最常见的大血管并发症之一，糖尿病患者发生心血管并发症的风险是非糖尿病患者的 2~4 倍^[4]，糖尿病死亡患者中约三分之二是因心血管并发症导致^[5]。因此，通过构建预测模型探索糖尿病患者发生心血管并发症的危险因素具有重要意义，可帮助医生更准确地评估患者心血管并发症发生

风险,有效识别出高危个体,对其尽早采取有针对性的干预措施,提供更紧密的监测管理,同时有利于患者了解自身状况,增强自身管理,提高患者生活质量。

既往三大经典临床试验ADVANCE^[6]、ACCORD^[7]和VADT^[8]探讨了2型糖尿病患者在强化/标准降糖治疗后对心血管疾病结局的影响,发现强化控制血糖并不能减少2型糖尿病患者心血管疾病的发生,提示血糖水平并非唯一影响心血管并发症的因素。英国前瞻性糖尿病研究(UKPADS)^[9-10]以及日本和中国香港人群的相关研究^[11-12]指出,年龄、吸烟、体力活动、糖尿病病程、心血管病史、体质量指数(body mass index, BMI)、收缩压、糖化血红蛋白、非高密度脂蛋白胆固醇和白蛋白与肌酐比值等是糖尿病心血管并发症的危险因素。目前,国内仅针对糖尿病心血管并发症构建预测模型的研究较少,如基于上海4335例糖尿病患者构建了糖尿病并发卒中的风险预测模型,发现年龄、糖尿病病程、血压、BMI、低密度脂蛋白胆固醇等是发生卒中的危险因素^[13],其主要在中国2型糖尿病人群中利用既往的心血管病风险预测模型进行验证,发现美国弗明汉(Framingham)模型和新西兰糖尿病队列研究(diabetes cohort study, DCS)模型不适用于我国2型糖尿病人群心血管病风险预测^[14]。同时,2型糖尿病的代谢环境,包括高血糖、胰岛素抵抗和过量的游离脂肪酸的释放,通过内皮功能障碍、氧化应激和低度炎症等机制影响血管壁,促进血栓形成,最终导致动脉粥样硬化,从而罹患心血管疾病和脑卒中^[15]。因此,本研究将冠心病和卒中共同作为糖尿病心血管并发症事件,基于京津冀社区自然人群慢性病队列研究的调查数据,利用Lasso回归筛选糖尿病心血管并发症的危险因素并构建预测模型,将为社区糖尿病患者早期预测预警提供有效工具,并为早期干预筛选高危人群提供科学依据。

1 资料与方法

1.1 研究对象

选取京津冀社区自然人群慢性病队列研究(CHCN-BTH,注册号:ChiCTR1900024725)基线(2017—2019年)调查数据进行横断面研究。纳入标准:①年龄18岁以上2型糖尿病患者;

②心血管疾病确诊时间在2型糖尿病确诊时间之后。排除标准:①脑出血及卒中类型不确定;②急性感染或炎症患者;③相关生化指标,如血脂四项(甘油三酯、胆固醇、低密度脂蛋白、极低密度脂蛋白)和脂蛋白a等数据缺失。CHCN-BTH研究主要覆盖北京、天津和河北三个省市18岁以上3.7万社区自然人群,2017—2019年通过问卷调查、体格检查和实验室检查等方法收集健康和慢性病相关的基线数据,建立社区自然人群健康数据库,旨在明确其病因和危险因素,从而建立疾病风险预测模型,用于社区慢性病的精准预防和干预。CHCN-BTH研究已获得首都医科大学伦理委员会的审核批准[批号:2018SY81],所有研究对象均签署了知情同意书。

1.2 资料收集

使用标准化问卷收集一般人口学特征(年龄、性别、婚姻状况、教育水平)、生活行为方式(吸烟、饮酒、体力劳动、体育锻炼频率)、个人健康情况(高血压、血脂异常、糖尿病病程、心理症状)和疾病家族史(高血压、糖尿病、冠心病和脑卒中)。通过体格检查获得血压、身高、体重等数据;收集的血液样本均在北京某三甲医院使用贝克曼库尔特化学分析仪AU5800进行统一临床实验室检测,空腹血糖使用己糖激酶法、糖化血红蛋白使用离子交换高压液相色谱法、尿酸使用脲酶比色法、尿素使用脲酶紫外速率法、超敏C反应蛋白使用免疫比浊法、血清胆红素使用重氮试剂法进行检测。

1.3 指标定义

一般人口学特征中婚姻状况分为未婚、已婚、离异和丧偶,教育水平分为小学及以下、初中、高中和大学及以上,将曾经和现在正在吸烟或饮酒的情况判定为有吸烟或饮酒情况。BMI为体重(kg)与身高(m)平方的比值。血压(使用欧姆龙HEM-907电子血压计)由医生在研究对象静坐休息后间隔测量三次取平均值,高血压诊断依据《中国老年高血压管理指南2019》,将收缩压 ≥ 140 mmHg(1 mmHg=0.133 kPa)和(或)舒张压 ≥ 90 mmHg,或自我报告曾在二级及以上医院明确诊断高血压且接受降压药物治疗的研究对象定义为患有高血压。血脂异常是由研究对象自我报告曾在二级及以上医院明确诊断,或根据《中国血脂管理指南(2023年)》中胆固醇和(或)

甘油三酯水平升高或高密度脂蛋白胆固醇水平降低进行定义。心理症状使用抑郁、焦虑和压力量表-21 (Depression, Anxiety and Stress Scale-21, DASS-21) 进行收集^[16-17], 该量表包括三个子量表, 每个子量表包含七个问题, 抑郁量表 > 9 判定为有抑郁症状, 焦虑量表 > 7 判定为有焦虑症状, 压力量表 > 14 判定为有压力症状, 存在任一状态则判定为有心理症状。

根据《中国 2 型糖尿病防治指南 (2020 年版)》^[18], 2 型糖尿病患者定义为糖化血红蛋白在 6.5% 及以上, 或自我报告曾在二级及以上医院诊断过, 或正在进行规律的降糖治疗。糖尿病心血管并发症定义为糖尿病患者确诊为糖尿病之后发生心血管病, 其确诊时间在糖尿病确诊时间之后。心血管病是由研究对象在 2017—2019 年基线数据收集时自我报告曾在二级及以上医院明确诊断, 按照《国际疾病分类第 10 版》(International Classification of Disease, 10th Edition, ICD-10) 标准, 本研究主要包括冠心病 (I21~I25) 和缺血性脑卒中 (I64)。

1.4 统计学分析

符合正态分布的连续性变量用 $\bar{x} \pm s$ 表示, 两组间比较采用 *t* 检验; 非正态分布的连续性变量以中位数 (*M*) 和四分位数 (P_{25}, P_{75}) 表示, 两组间比较采用 Mann-Whitney U 检验; 分类变量分布用例数 (*n*) 和百分比 (%) 表示, 两组间比较采用 χ^2 检验。将研究对象按 7:3 比例随机分为训练集和测试集, 分别进行模型的构建与验证。采用最小绝对值收敛和选择算子算法 (least absolute shrinkage and selection operator, Lasso) 回归方法, 根据既往报道的糖尿病心血管并发症模型的危险因素

以及本研究数据库包含的其他未探讨的因素, 分析筛选训练集的 30 个自变量, 采取十折交叉验证法确定参数 λ , 选取模型误差在一个标准差时的 λ , 将此时 β 不为 0 的变量纳入并拟合多因素 Logistic 回归模型。根据多因素 Logistic 回归的结果, 选取 $P \leq 0.05$ 的变量构建糖尿病心血管并发症的预测模型并绘制列线图进行可视化, 列线图中每个预测因素对应的线段上都标注了刻度, 代表该因素的可取值范围, 线段的长度反应该因素对结局事件的贡献大小。图中顶部的分数表示预测因素在不同取值下所对应的得分, 所有预测因素分数相加, 进而通过总分对应图中底部线得到糖尿病心血管并发症发生的概率预测值。基于测试集对构建的预测模型进行验证。绘制受试者工作特征曲线 (receiver operating characteristic curve, ROC) 并计算曲线下面积 (area under curve, AUC) 作为区分度评价指标, 校准度通过 Hosmer-Lemeshow 拟合优度检验和绘制校准曲线来评估, 以确保模型的预测能力。所有统计分析使用 R 语言软件 4.2.0 版本完成, 采用的软件包为 “rms” “glmnet” “pROC” “ResourceSelection”。双侧检验, $P \leq 0.05$ 认为差异具有统计学意义。

2 结果

2.1 一般情况

共纳入 813 例 2 型糖尿病患者, 其中训练集 569 例 (62.6 ± 10.3 岁), 测试集 244 例 (62.8 ± 10.6 岁), 分别有 85 (14.9%) 例和 27 (11.1%) 例 2 型糖尿病患者发生心血管并发症, 两组间除超敏 C 反应蛋白水平具有差异 ($P=0.028$) 外, 其余基本特征差异无统计学意义 (表 1)。

表1 患者基线资料 (n, %)

Table 1. Basic characteristics of patients (n, %)

| 项目 | 全部研究对象 (n=813) | 训练集 (n=569) | 测试集 (n=244) | <i>t</i> / <i>Z</i> / χ^2 值 | <i>P</i> 值 |
|--------------------------|----------------|-------------|-------------|----------------------------------|------------|
| 年龄 (岁, $\bar{x} \pm s$) | 62.6 ± 10.4 | 62.6 ± 10.3 | 62.8 ± 10.6 | -0.208 | 0.835 |
| 性别 | | | | 0.026 | 0.872 |
| 男 | 448 (55.1) | 312 (54.8) | 136 (55.7) | | |
| 女 | 365 (44.9) | 257 (45.2) | 108 (44.3) | | |
| 婚姻状况 | | | | 4.170 | 0.244 |
| 未婚 | 17 (2.1) | 11 (1.9) | 6 (2.5) | | |
| 已婚 | 739 (90.9) | 512 (90.0) | 227 (93.0) | | |
| 离异 | 17 (2.1) | 15 (2.6) | 2 (0.8) | | |
| 丧偶 | 40 (4.9) | 31 (5.4) | 9 (3.7) | | |

续表1

| 项目 | 全部研究对象 (n=813) | 训练集 (n=569) | 测试集 (n=244) | t/Z/ χ^2 值 | P值 |
|--|-------------------|-------------------|-------------------|-----------------|-------|
| 教育水平 | | | | 5.436 | 0.143 |
| 小学及以下 | 127 (15.6) | 89 (15.6) | 38 (15.6) | | |
| 初中 | 292 (35.9) | 200 (35.1) | 92 (37.7) | | |
| 高中 | 211 (26.0) | 160 (28.1) | 51 (20.9) | | |
| 大学及以上 | 183 (22.5) | 120 (21.1) | 63 (25.8) | | |
| 吸烟 | | | | 0.067 | 0.796 |
| 否 | 506 (62.2) | 352 (61.9) | 154 (63.1) | | |
| 是 | 307 (37.8) | 217 (38.1) | 90 (36.9) | | |
| 饮酒 | | | | 0.005 | 0.945 |
| 否 | 500 (61.5) | 349 (61.3) | 151 (61.9) | | |
| 是 | 313 (38.5) | 220 (38.7) | 93 (38.1) | | |
| 体力劳动 | | | | 0.851 | 0.654 |
| 轻度 | 661 (81.3) | 462 (81.2) | 199 (81.6) | | |
| 中度 | 131 (16.1) | 94 (16.5) | 37 (15.2) | | |
| 重度 | 21 (2.6) | 13 (2.3) | 8 (3.3) | | |
| 体育锻炼频率 | | | | 4.980 | 0.289 |
| 每周5~7天 | 435 (53.5) | 308 (54.1) | 127 (52.0) | | |
| 每周3~4天 | 99 (12.2) | 61 (10.7) | 38 (15.6) | | |
| 每周1~2天 | 89 (10.9) | 64 (11.2) | 25 (10.2) | | |
| 每月 \leq 3天 | 30 (3.7) | 19 (3.3) | 11 (4.5) | | |
| 从不锻炼 | 160 (19.7) | 117 (20.6) | 43 (17.6) | | |
| 高血压 | | | | 0.191 | 0.662 |
| 否 | 392 (48.2) | 271 (47.6) | 121 (49.6) | | |
| 是 | 421 (51.8) | 298 (52.4) | 123 (50.4) | | |
| 血脂异常 | | | | 0.073 | 0.787 |
| 否 | 613 (75.4) | 427 (75.0) | 186 (76.2) | | |
| 是 | 200 (24.6) | 142 (25.0) | 58 (23.8) | | |
| 糖尿病病程 (年, $\bar{x} \pm s$) | 12.5 \pm 7.4 | 12.3 \pm 7.4 | 12.9 \pm 7.3 | -0.992 | 0.321 |
| 心理症状 | | | | 0.467 | 0.494 |
| 无 | 724 (89.1) | 510 (89.6) | 214 (87.7) | | |
| 有 | 89 (10.9) | 59 (10.4) | 30 (12.3) | | |
| 高血压家族史 | | | | 2.535 | 0.111 |
| 否 | 367 (45.1) | 246 (43.2) | 121 (49.6) | | |
| 是 | 446 (54.9) | 323 (56.8) | 123 (50.4) | | |
| 糖尿病家族史 | | | | 0.021 | 0.885 |
| 否 | 415 (51.0) | 289 (50.8) | 126 (51.6) | | |
| 是 | 398 (49.0) | 280 (49.2) | 118 (48.4) | | |
| 心血管病家族史 | | | | 0.145 | 0.704 |
| 否 | 547 (67.3) | 380 (66.8) | 167 (68.4) | | |
| 是 | 266 (32.7) | 189 (33.2) | 77 (31.6) | | |
| BMI ($\text{kg} \cdot \text{m}^{-2}$) [*] | 26.2 (23.9, 28.4) | 26.4 (23.9, 28.6) | 25.9 (23.9, 28.1) | -1.161 | 0.246 |
| 腰臀比 [*] | 0.9 (0.9, 1.0) | 0.9 (0.9, 1.0) | 0.9 (0.9, 1.0) | -0.807 | 0.420 |

续表1

| 项目 | 全部研究对象 (n=813) | 训练集 (n=569) | 测试集 (n=244) | t/Z/χ ² 值 | P值 |
|--|----------------------|----------------------|----------------------|----------------------|-------|
| 空腹血糖 (mmol · L ⁻¹) * | 7.6 (6.6, 9.1) | 7.5 (6.5, 8.9) | 7.8 (6.6, 9.5) | -1.901 | 0.057 |
| 糖化血红蛋白 (%) * | 6.7 (6.0, 7.5) | 6.7 (6.0, 7.5) | 6.7 (6.0, 7.5) | -0.028 | 0.978 |
| 谷丙转氨酶 (mmol · L ⁻¹) * | 18.0 (14.0, 26.0) | 18.0 (14.0, 26.1) | 18.4 (13.9, 25.6) | -0.834 | 0.404 |
| 谷草转氨酶 (mmol · L ⁻¹) | 19.0 (15.2, 23.0) | 19.0 (15.5, 23.0) | 18.4 (15.0, 23.0) | -0.706 | 0.480 |
| 尿素 (mmol · L ⁻¹) * | 5.3 (4.3, 6.3) | 5.2 (4.3, 6.3) | 5.4 (4.5, 6.3) | -0.544 | 0.587 |
| 尿酸 (μmol · L ⁻¹) * | 322.0 (269.0, 382.0) | 325.0 (267.3, 383.1) | 317.6 (271.0, 373.7) | -0.748 | 0.455 |
| 血肌酐 (μmol · L ⁻¹) * | 62.0 (51.0, 75.0) | 63.0 (50.0, 75.5) | 61.9 (52.0, 73.0) | -0.053 | 0.958 |
| 超敏C反应蛋白 (mmol · L ⁻¹) * | 1.4 (0.7, 2.5) | 1.5 (0.8, 2.5) | 1.3 (0.6, 2.3) | -2.201 | 0.028 |
| 总蛋白 (g · L ⁻¹ , $\bar{x} \pm s$) | 74.7 ± 4.1 | 74.7 ± 4.1 | 74.8 ± 4.1 | 0.058 | 0.954 |
| 白蛋白 (g · L ⁻¹) * | 46.4 (44.2, 48.8) | 46.4 (44.1, 49.0) | 46.1 (44.2, 48.5) | -0.574 | 0.566 |
| 球蛋白 (g · L ⁻¹) * | 27.9 (25.5, 30.2) | 27.8 (25.5, 30.2) | 27.9 (25.7, 30.0) | -0.955 | 0.955 |
| 白球比* | 1.7 (1.5, 1.9) | 1.7 (1.5, 1.9) | 1.7 (1.5, 1.9) | -0.154 | 0.878 |
| 血清胆红素 (μmol · L ⁻¹) * | 13.1 (10.3, 17.1) | 13.0 (9.9, 17.1) | 13.4 (10.6, 17.5) | -1.225 | 0.220 |

注: *非正态分布的连续性变量以中位数和四分位数表示[M (P₂₅, P₇₅)]。

2.2 模型构建

根据训练集数据将问卷信息和临床生化指标的 30 个潜在危险因素进行 Lasso-Logistic 回归分析和十折交叉验证 (图 1), 根据距离模型均方误差最小值一个标准误的 λ 值 (lambda.1se=0.032) 时结果, 确定了 10 个非零特征变量, 包括年龄、吸烟、高血压、血脂异常、糖尿病病程、心血管病家族史、空腹血糖、尿素、肌酐和血清胆红素水平。

将 10 个非零特征变量纳入多因素 Logistic 回归模型, 结果显示年龄 [OR=1.040, 95%CI (1.010, 1.073), P=0.010]、高血压 [OR=2.211, 95%CI (1.263, 3.975), P=0.006]、糖尿病病程 [OR=1.063, 95%CI (1.028, 1.099), P < 0.001]、空腹血糖水平 [OR=1.186, 95%CI (1.075, 1.309), P=0.001]、血脂异常 [OR=2.051, 95%CI (1.167, 3.583), P=0.012]、心血管病家族史 [OR=2.794, 95%CI (1.650, 4.774), P < 0.001] 和吸烟 [OR=1.975, 95%CI (1.133, 3.462), P=0.017] 是糖尿病心血管并发症的危险因素, 而血清胆红素水平 [OR=0.940, 95%CI (0.889, 0.991), P=0.027] 是其保护因素 (表 2)。最终预测模型表达式为:

发生概率 $P=1/[1+e^{-(-7.758+0.040 \text{ 年龄} +0.793 \text{ 高血压} +0.061 \text{ 糖尿病病程} +0.171 \text{ 空腹$

$\text{血糖} +0.718 \text{ 血脂异常} +1.028 \text{ 心血管病家族史} +0.681 \text{ 吸烟} -0.062 \text{ 血清胆红素})]$ 。

根据 Lasso 回归联合多因素 Logistic 回归筛选出 8 个最佳预测因素绘制列线图 (图 2)。通过各个预测因素相加得到的总分对应糖尿病心血管疾病并发症的发生风险, 例如, 某罹患 2 型糖尿病 5.5 年的患者, 年龄 73 岁、不吸烟、患有高血压和血脂异常、有心脑血管疾病家族史、空腹血糖为 7.5 mmol · L⁻¹、血清胆红素为 18.08 μmol · L⁻¹, 由列线图计算该患者总分为 205.8, 发生糖尿病心血管疾病并发症的概率为 21.4%。

2.3 模型验证

对训练集构建的糖尿病心血管并发症的预测模型进行评估, 同时利用测试集对预测模型进行内部验证。训练集和测试集分别绘制 ROC 曲线, AUC 值分别为 0.803[95%CI (0.753, 0.852)] 和 0.820[95%CI (0.736, 0.903)], 提示模型区分度较好 (图 3); 对训练集和测试集进行 Hosmer-Lemeshow 拟合优度检验, P 值分别为 0.776 4 和 0.554 2, 提示模型预测值与实际观测值之间并无统计学差异, 再经 500 次重抽样校正后的校准曲线与理想曲线的平均绝对误差为 0.013, 提示曲线之间一致性较好, 模型校准度良好 (图 4)。

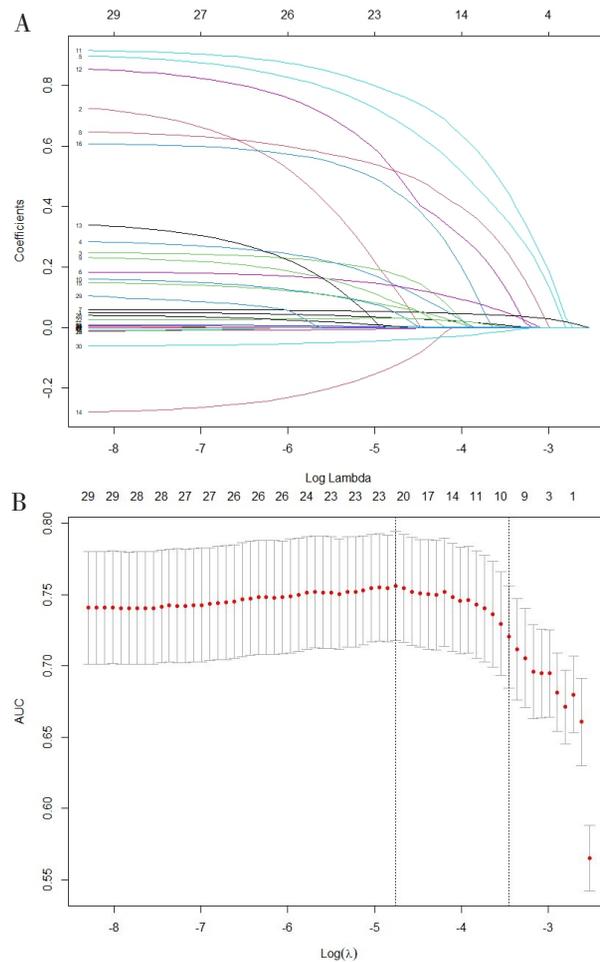


图1 基于十折交叉验证的Lasso回归特征变量选择

Figure 1. Lasso regression feature variable selection based on ten-fold cross-validation

注：A.预测因子系数平面图，图中横坐标为 λ 的对数，纵坐标为回归系数，图中的每一条彩色实线代表一个变量，随着 $\text{Log}(\lambda)$ 的增加变量系数不断减少，部分变量系数变为0；B.Lasso回归交叉验证模型交叉验证图，图中横坐标为 λ 的对数，纵坐标为均方误差，图中左边的虚线代表均方误差最小时的 λ 值，右边的虚线代表距离均方误差最小时的一个标准误差的 λ 值，在垂直虚线对应的最低点的惩罚值，即曲线最低点对应的上坐标，在图A的相应惩罚值的位置寻找对应位置的垂线，垂线与图A彩色实线相交的点数即为模型纳入的变量数目。

表2 Lasso-Logistic回归模型中糖尿病心血管并发症的预测因素

Table 2. Predictors of cardiovascular complications in diabetes mellitus in Lasso-Logistic regression models

| 变量 | Lasso回归 | 多因素Logistic回归 | |
|---|---------|----------------------|--------|
| | 系数 | OR值 (95%CI) | P值 |
| 年龄 (岁) | 0.007 | 1.040 (1.010, 1.073) | 0.010 |
| 高血压 | 0.345 | 2.211 (1.263, 3.975) | 0.006 |
| 糖尿病病程 (年) | 0.039 | 1.063 (1.028, 1.099) | <0.001 |
| 空腹血糖 ($\text{mmol} \cdot \text{L}^{-1}$) | 0.046 | 1.186 (1.075, 1.309) | 0.001 |
| 血脂异常 | 0.239 | 2.051 (1.167, 3.583) | 0.012 |
| 心血管病家族史 | 0.440 | 2.794 (1.650, 4.774) | <0.001 |
| 吸烟 | 0.111 | 1.975 (1.133, 3.462) | 0.017 |
| 尿素 ($\text{mmol} \cdot \text{L}^{-1}$) | 0.000 | 1.048 (0.914, 1.183) | 0.473 |
| 肌酐 ($\mu\text{mol} \cdot \text{L}^{-1}$) | 0.002 | 1.005 (0.997, 1.016) | 0.357 |
| 血清胆红素 ($\mu\text{mol} \cdot \text{L}^{-1}$) | -0.010 | 0.940 (0.889, 0.991) | 0.027 |

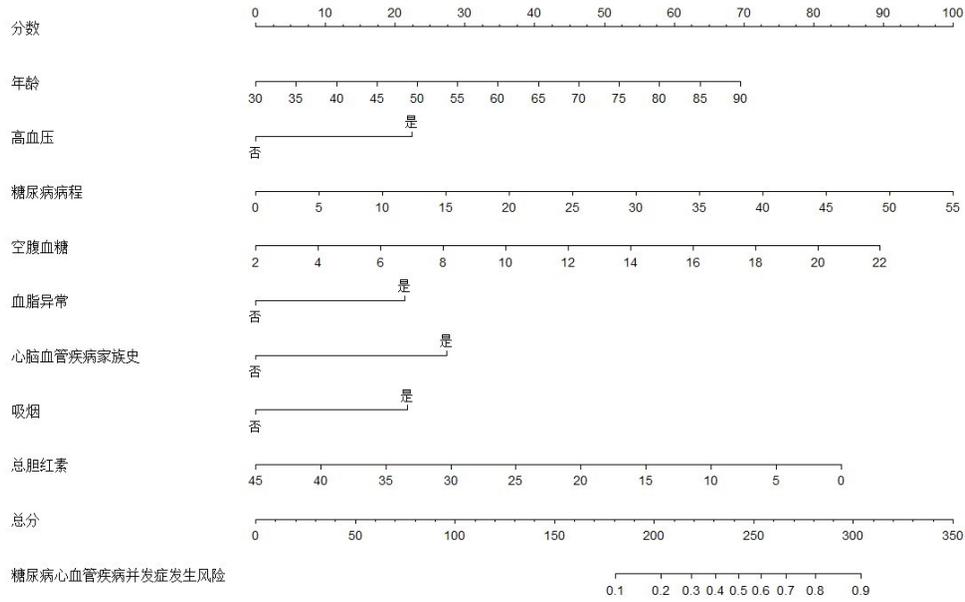


图2 2型糖尿病心血管疾病预测模型的列线图

Figure 2. A nomogram of a predictive model for cardiovascular complications of type 2 diabetes

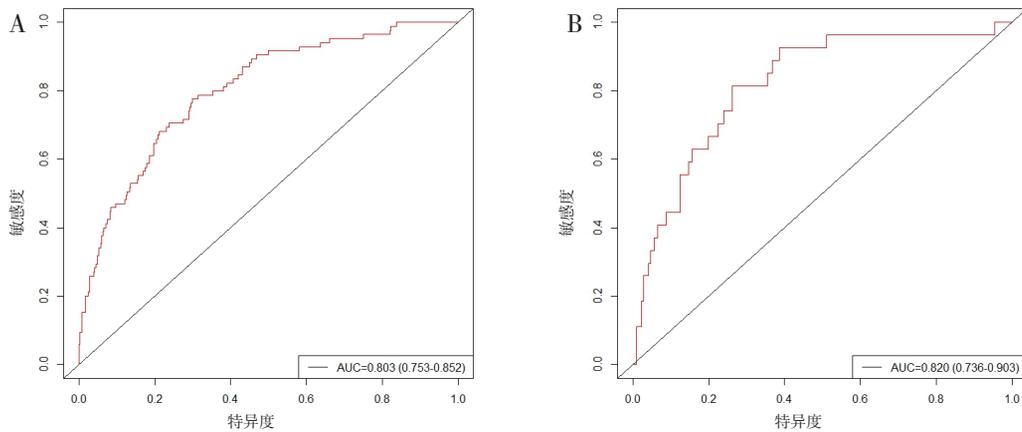


图3 2型糖尿病心血管疾病预测模型的受试者工作曲线

Figure 3. AUC of a predictive model for cardiovascular complications of type 2 diabetes

注: A.训练集, 曲线下面积为0.803; B.测试集, 曲线下面积为0.820。

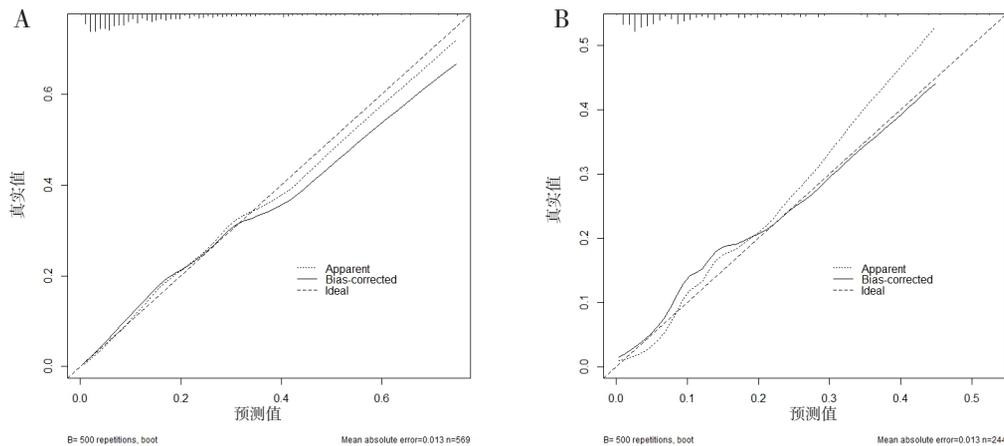


图4 糖尿病心血管疾病预测模型的校准曲线

Figure 4. Calibration curve of prediction model for cardiovascular complications of diabetes mellitus

注: 线段虚线为理想曲线, 点状虚线为未经bootstrap校正的校准曲线, 实线为500次bootstrap校正后的校准曲线; A.训练集; B.测试集。

3 讨论

本研究利用京津冀社区自然人群慢性病队列基线调查数据,采用 Lasso-Logistic 回归的分析方法构建糖尿病心血管并发症风险预测模型,纳入的预测因子包括年龄、吸烟、高血压、血脂异常、糖尿病病程、心血管病家族史、空腹血糖和血清胆红素,经内部验证表明该预测模型有较好的区分度和校准度。

本研究显示 2 型糖尿病患者的年龄和糖尿病病程是心血管并发症的危险因素,Wan 等^[19]、Morton 等^[20]和 De Jong 等^[21]分别针对中国、澳大利亚和英国的糖尿病患者心血管并发症发生风险的研究表明,随着年龄和病程时间的增加心血管病风险逐渐提高,与本研究的结果一致。Al-Rubeaan 等^[22]发现 2 型糖尿病患者中缺血性卒中的患病率在糖尿病病程时间长且老年人组最高。同时现在糖尿病确诊年龄越来越趋于年轻化也意味着病程时间的延长,确诊年龄每提前 10 年发生心血管并发症的风险逐渐增加^[23]。

本研究发现患有高血压和血脂异常以及空腹血糖升高是糖尿病患者发生心血管并发症的重要危险因素。Kaze 等^[24]发现与血脂代谢正常的 2 型糖尿病患者相比,血脂异常的患者发生心血管并发症的风险更高,其中冠心病发生风险增加 1.48 倍。血脂异常也是高血压和糖尿病两者共同的危险因素^[25]。既往研究证实,高血糖和高血压会共同加剧血管损伤和内皮功能障碍^[26-27],促进动脉粥样硬化的发生发展,糖尿病合并高血压患者可使大血管并发症的风险显著增加。Sibarani 等^[28]证明在血糖异常患者中合并高血压导致心血管并发症的发生风险增加 2.58 倍;Elley 等^[29]开展新西兰糖尿病队列研究发现收缩压每升高 10 mmHg,心血管并发症的风险增加 1.03 倍,与本研究结果一致。

本研究构建的糖尿病心血管并发症的预测模型与其他研究相比还包括危险因素心血管病家族史和保护因素血清胆红素。Tian 等^[30]利用中国慢性病前瞻性研究项目数据发现有卒中家族史的研究对象发生脑卒中的风险增加 1.5 倍,且患有家族史的人数与脑卒中发生风险有剂量反应关系。Lloyd-Jones 等^[31]基于 Framingham 原始队列的子代人群发现其父母有心血管病史的

男性子代发生心血管病风险提高 2.6 倍,女性子代提高 2.3 倍。血清胆红素是天然的内源性抗氧化剂,抑制氧化修饰低密度脂蛋白胆固醇生成,清除自由基,具有抗脂质过氧化、抗动脉粥样斑块等作用,因此对心血管病有保护作用^[32-33]。Kunutsor 等^[34]对 12 篇研究进行 Meta 分析,调整传统心血管病危险因素后,发现血清胆红素基线水平增加 2 倍,心血管病发生风险可降低 10%。然而血清胆红素与心血管病发病风险方面因果关系仍不明确,仍需进一步探讨。

本研究通过探究糖尿病心血管并发症的多种危险因素,初步构建了针对糖尿病心血管并发症的预测模型。既往研究构建或验证预测模型的研究人群大多来源于欧美国家^[35],由于种族和生活行为等差异,模型外推到亚洲地区时预测能力可能存在差异。与 2 型糖尿病患者 10 年心血管风险预测评分 SCORE2-Diabetes 相比^[36],两个研究所纳入的预测因素信息获取均较为便捷,同时本研究纳入的预测因素所反映信息与之基本相似,且进一步考虑了心血管疾病家族史的影响。另一方面,本研究构建的糖尿病心血管并发症预测模型的区分度与 SCORE2-Diabetes 相比更好,但 SCORE2-Diabetes 研究最终可计算出风险评分,可更为直观地为临床医生和糖尿病患者自身进行快速甄别心血管并发症高风险人群,进而提供预警信息和及时干预。

本研究仍存在一些局限性。首先,本研究利用基线调查时患者自报疾病史的信息判定先发生糖尿病后发生心血管病的时间先后顺序,可能存在回忆偏倚和信息偏倚,未来将对糖尿病患者进行长期随访获得心血管并发症发生情况,以提高预测模型的准确性;其次,因未能获取到其他糖尿病患者心血管并发症发生情况的研究数据,该预测模型没有开展外部验证,模型的外推效果尚有待进一步验证;此外,本研究为横断面调查,并限于研究设计,收集的有关实验室检查指标可能会受到基线糖尿病患病的影响,导致预测结果的精确性下降;且纳入样本量相对较少,结果可能不够稳定,可重复性有待验证,同时未包括患者的饮食和药物使用情况等信息,且研究对象仅集中在京津冀地区,未来对中国其他地区的糖尿病心血管并发症的危险因素预测需要设计更加严谨的前瞻性多中心队列研究来完善预测模型。

综上所述,本研究构建了一个简便的糖尿病心血管并发症风险预测模型,模型内部验证区分度和校准度较好,能够直观地反映不同糖尿病患者的心血管并发症发生风险,在一定程度上能够为临床开展早期预防和干预提供科学的依据。

参考文献

- 1 American Diabetes Association. 2. Classification and diagnosis of diabetes: standards of medical care in diabetes-2020[J]. *Diabetes Care*, 2020, 43(1): 14-31. DOI: [10.2337/dc20-S002](https://doi.org/10.2337/dc20-S002).
- 2 Tinajero MG, Malik VS. An update on the epidemiology of type 2 diabetes: a global perspective[J]. *Endocrinol Metab Clin North Am*, 2021, 50(3): 337-355. DOI: [10.1016/j.ecl.2021.05.013](https://doi.org/10.1016/j.ecl.2021.05.013).
- 3 International Diabetes Federation. IDF diabetes atlas 10th edition[J]. 2021. <https://diabetesatlas.org/atlas/tenth-edition/>.
- 4 Rawshani A, Rawshani A, Franzén S, et al. Mortality and cardiovascular disease in type 1 and type 2 diabetes[J]. *NEJM*, 2017, 376(15): 1407-1418. DOI: [10.1056/NEJMoa1608664](https://doi.org/10.1056/NEJMoa1608664).
- 5 Low Wang CC, Hess CN, Hiatt WR, et al. Clinical update: cardiovascular disease in diabetes mellitus: atherosclerotic cardiovascular disease and heart failure in type 2 diabetes mellitus-mechanisms, management, and clinical considerations[J]. *Circulation*, 2016, 133(24): 2459-2502. DOI: [10.1161/circulationaha.116.022194](https://doi.org/10.1161/circulationaha.116.022194).
- 6 Zoungas S, Chalmers J, Neal B, et al. Follow-up of blood-pressure lowering and glucose control in type 2 diabetes[J]. *NEJM*, 2014, 371(15): 1392-1406. DOI: [10.1056/NEJMoa1407963](https://doi.org/10.1056/NEJMoa1407963).
- 7 Gerstein HC, Miller ME, Byington RP, et al. Effects of intensive glucose lowering in type 2 diabetes[J]. *NEJM*, 2008, 358(24): 2545-2559. DOI: [10.1056/NEJMoa0802743](https://doi.org/10.1056/NEJMoa0802743).
- 8 Duckworth W, Abraira C, Moritz T, et al. Glucose control and vascular complications in veterans with type 2 diabetes [J]. *NEJM*, 2009, 360(2): 129-139. DOI: [10.1056/NEJMoa0808431](https://doi.org/10.1056/NEJMoa0808431).
- 9 Kothari V, Stevens RJ, Adler AI, et al. UKPDS 60: risk of stroke in type 2 diabetes estimated by the UK prospective diabetes study risk engine[J]. *Stroke*, 2002, 33(7): 1776-1781. DOI: [10.1161/01.str.0000020091.07144.c7](https://doi.org/10.1161/01.str.0000020091.07144.c7).
- 10 Stevens RJ, Kothari V, Adler AI, et al. The UKPDS risk engine: a model for the risk of coronary heart disease in Type II diabetes (UKPDS 56)[J]. *Clin Sci (Lond)*, 2001, 101(6): 671-679. DOI: [10.1042/cs20000335](https://doi.org/10.1042/cs20000335).
- 11 Tanaka S, Tanaka S, Iimuro S, et al. Predicting macro- and microvascular complications in type 2 diabetes: the Japan diabetes complications study/the Japanese elderly diabetes intervention trial risk engine[J]. *Diabetes Care*, 2013, 36(5): 1193-1199. DOI: [10.2337/dc12-0958](https://doi.org/10.2337/dc12-0958).
- 12 Yang X, So WY, Kong AP, et al. Development and validation of stroke risk equation for Hong Kong Chinese patients with type 2 diabetes: the Hong Kong Diabetes Registry[J]. *Diabetes Care*, 2007, 30(1): 65-70. DOI: [10.2337/dc06-1273](https://doi.org/10.2337/dc06-1273).
- 13 Shi R, Zhang T, Sun H, et al. Establishment of clinical prediction model based on the study of risk factors of stroke in patients with type 2 diabetes mellitus[J]. *Front Endocrinol (Lausanne)*, 2020, 11: 559. DOI: [10.3389/fendo.2020.00559](https://doi.org/10.3389/fendo.2020.00559).
- 14 梁靖媛, 李倩倩, 刘晓非, 等. 心血管病风险预测模型在 2 型糖尿病人群中的验证 [J]. 第三军医大学学报, 2021, 43(12): 1097-1102. [Liang JY, Li QQ, Liu XF, et al. Cardiovascular disease risk prediction models for type 2 diabetes: verification in Chinese patients[J]. *Journal of Third Military Medical University*, 2021, 43(12): 1097-1102.] DOI: [10.16016/j.1000-5404.202101028](https://doi.org/10.16016/j.1000-5404.202101028).
- 15 Kaur R, Kaur M, Singh J. Endothelial dysfunction and platelet hyperactivity in type 2 diabetes mellitus: molecular insights and therapeutic strategies[J]. *Cardiovasc Diabetol*, 2018, 17(1): 121. DOI: [10.1186/s12933-018-0763-3](https://doi.org/10.1186/s12933-018-0763-3).
- 16 Lovibond PF, Lovibond SH. The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck depression and anxiety inventories[J]. *Behav Res Ther*, 1995, 33(3): 335-343. DOI: [10.1016/0005-7967\(94\)00075-u](https://doi.org/10.1016/0005-7967(94)00075-u).
- 17 Antony Mm, Bieling Pj, Cox Bj, et al. Psychometric properties of the 42-item and 21-item versions of the depression anxiety stress scales in clinical groups and a community sample[J]. *Psychol Assess*, 1998, 10: 176-181. DOI: [10.1037/1040-3590.10.2.176](https://doi.org/10.1037/1040-3590.10.2.176).
- 18 中华医学会儿科学分会. 中国 2 型糖尿病防治指南 (2020 年版) [J]. *中华糖尿病杂志*, 2021, 13(4): 315-409.

- [Chinese Diabetes Society. Guideline for the prevention and treatment of type 2 diabetes mellitus in China (2020 edition)[J]. Chinese Journal of Diabetes Mellitus, 2021, 13(4): 315–409.] DOI: [10.3760/cma.j.cn115791-20210221-00095](https://doi.org/10.3760/cma.j.cn115791-20210221-00095).
- 19 Wan EYF, Fong DYT, Fung CSC, et al. Development of a cardiovascular diseases risk prediction model and tools for Chinese patients with type 2 diabetes mellitus: a population-based retrospective cohort study[J]. Diabetes Obes Metab, 2018, 20(2): 309–318. DOI: [10.1111/dom.13066](https://doi.org/10.1111/dom.13066).
- 20 Morton JI, Lazzarini PA, Polkinghorne KR, et al. The association of attained age, age at diagnosis, and duration of type 2 diabetes with the long-term risk for major diabetes-related complications[J]. Diabetes Res Clin Pract, 2022, 190: 110022. DOI: [10.1016/j.diabres.2022.110022](https://doi.org/10.1016/j.diabres.2022.110022).
- 21 De Jong M, Woodward M, Peters SAE. Duration of diabetes and the risk of major cardiovascular events in women and men: a prospective cohort study of UK biobank participants[J]. Diabetes Res Clin Pract, 2022, 188: 109899. DOI: [10.1016/j.diabres.2022.109899](https://doi.org/10.1016/j.diabres.2022.109899).
- 22 Al-Rubeaan K, Al-Hussain F, Youssef AM, et al. Ischemic stroke and its risk factors in a registry-based large cross-sectional diabetic cohort in a country facing a diabetes epidemic[J]. J Diabetes Res, 2016, 2016: 4132589. DOI: [10.1155/2016/4132589](https://doi.org/10.1155/2016/4132589).
- 23 Sattar N, Rawshani A, Franzén S, et al. Age at diagnosis of type 2 diabetes mellitus and associations with cardiovascular and mortality risks[J]. Circulation, 2019, 140(16): e724–e725. DOI: [10.1161/circulationaha.119.042457](https://doi.org/10.1161/circulationaha.119.042457).
- 24 Kaze AD, Santhanam P, Musani SK, et al. Metabolic dyslipidemia and cardiovascular outcomes in type 2 diabetes mellitus: findings from the look ahead study[J]. J Am Heart Assoc, 2021, 10(7): e016947. DOI: [10.1161/jaha.120.016947](https://doi.org/10.1161/jaha.120.016947).
- 25 Petrie JR, Guzik TJ, Touyz RM. Diabetes, Hypertension, and cardiovascular disease: clinical insights and vascular mechanisms[J]. Can J Cardiol, 2018, 34(5): 575–584. DOI: [10.1016/j.cjca.2017.12.005](https://doi.org/10.1016/j.cjca.2017.12.005).
- 26 Zhang YQ, Li Y, Dong YG, et al. A nationwide assessment of blood pressure control and the associated factors in Chinese type 2 diabetes mellitus patients[J]. J Clin Hypertens (Greenwich), 2019, 21(11): 1654–1663. DOI: [10.1111/jch.13675](https://doi.org/10.1111/jch.13675).
- 27 Gæde P, Oellgaard J, Carstensen B, et al. Years of life gained by multifactorial intervention in patients with type 2 diabetes mellitus and microalbuminuria: 21 years follow-up on the Steno-2 randomised trial[J]. Diabetologia, 2016, 59(11): 2298–2307. DOI: [10.1007/s00125-016-4065-6](https://doi.org/10.1007/s00125-016-4065-6).
- 28 Sibarani MHR, Wijaya IP, Rizka A, et al. Cardiovascular disease prediction model for Indonesian adult population with prediabetes and diabetes mellitus: the bogor cohort study of noncommunicable diseases risk factors[J]. Diabetes Metab Syndr, 2022, 16(1): 102330. DOI: [10.1016/j.dsx.2021.102330](https://doi.org/10.1016/j.dsx.2021.102330).
- 29 Elley CR, Robinson E, Kenealy T, et al. Derivation and validation of a new cardiovascular risk score for people with type 2 diabetes: the New Zealand diabetes cohort study[J]. Diabetes Care, 2010, 33(6): 1347–1352. DOI: [10.2337/dc09-1444](https://doi.org/10.2337/dc09-1444).
- 30 Tian T, Jin G, Yu C, et al. Family history and stroke risk in China: evidence from a large cohort study[J]. J Stroke, 2017, 19(2): 188–195. DOI: [10.5853/jos.2016.01270](https://doi.org/10.5853/jos.2016.01270).
- 31 Lloyd-Jones DM, Nam BH, D'agostino RB Sr, et al. Parental cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults: a prospective study of parents and offspring[J]. JAMA, 2004, 291(18): 2204–2211. DOI: [10.1001/jama.291.18.2204](https://doi.org/10.1001/jama.291.18.2204).
- 32 Seyed Khoei N, Wagner KH, Sedlmeier AM, et al. Bilirubin as an indicator of cardiometabolic health: a cross-sectional analysis in the UK biobank[J]. Cardiovasc Diabetol, 2022, 21(1): 54. DOI: [10.1186/s12933-022-01484-x](https://doi.org/10.1186/s12933-022-01484-x).
- 33 Mietus-Snyder M, Frier A, Glass CK, et al. Regulation of scavenger receptor expression in smooth muscle cells by protein kinase C: a role for oxidative stress[J]. Arterioscler Thromb Vasc Biol, 1997, 17(5): 969–978. DOI: [10.1161/01.atv.17.5.969](https://doi.org/10.1161/01.atv.17.5.969).
- 34 Kunutsor SK, Bakker SJ, Gansevoort RT, et al. Circulating total bilirubin and risk of incident cardiovascular disease in the general population[J]. Arterioscler Thromb Vasc Biol, 2015, 35(3): 716–724. DOI: [10.1161/atvbaha.114.304929](https://doi.org/10.1161/atvbaha.114.304929).
- 35 中华医学会神经病学分会. 中国缺血性脑卒中风险评估量表使用专家共识 [J]. 中华神经科杂志, 2016, 49(7): 519–525. [Neurology Branch of the Chinese Medical Association. Expert consensus on the use of the Chinese ischemic stroke risk assessment scale[J]. Chinese Journal

of Neurology, 2016, 49(7): 519–525.] DOI: [10.3760/cma.j.issn.1006-7876.2016.07.003](https://doi.org/10.3760/cma.j.issn.1006-7876.2016.07.003).

36 Score2–Diabetes Working Group and the Esc Cardiovascular Risk Collaboration. SCORE2–Diabetes: 10–year cardiovascular risk estimation in type 2 diabetes

in Europe[J]. Eur Heart J, 2023, 44(28): 2544–2556. DOI: [10.1093/eurheartj/ehad260](https://doi.org/10.1093/eurheartj/ehad260).

收稿日期: 2023 年 10 月 21 日 修回日期: 2023 年 12 月 06 日
本文编辑: 曹 越 桂裕亮

引用本文: 孙源, 刘括, 李冰潇, 等. 糖尿病心血管并发症预测模型的构建 [J]. 医学新知, 2024, 34(1): 2–13. DOI: [10.12173/j.issn.1004-5511.202310094](https://doi.org/10.12173/j.issn.1004-5511.202310094)
Sun Y, Liu K, Li BX, et al. Establishment of predictive model for cardiovascular complication of diabetes mellitus[J]. Yixue Xinzhi Zazhi, 2024, 34(1): 2–13. DOI: [10.12173/j.issn.1004-5511.202310094](https://doi.org/10.12173/j.issn.1004-5511.202310094)