

# 代谢风险评分在2型糖尿病人群心血管结局预测中的应用

申路佳<sup>1,2</sup>, 逯天威<sup>3</sup>, 巩伟明<sup>1,2</sup>, 赵岩松<sup>1,2</sup>, 王淑康<sup>1,2</sup>, 袁中尚<sup>1,2</sup>

(1.山东大学齐鲁医学院公共卫生学院生物统计学系, 山东 济南 250012;

2.国家健康医疗大数据研究院, 山东 济南 250003;

3.山东大学齐鲁医学院公共卫生学院儿少卫生与妇幼保健学系, 山东 济南 250012)

**摘要:**目的 识别2型糖尿病人群中分别与心肌梗死、心力衰竭和缺血性脑卒中相关的代谢物,进而构建代谢风险评分,并评价将代谢风险评分纳入传统临床模型是否能提高模型预测准确性。方法 利用英国生物银行队列数据,应用多变量Cox比例风险回归模型筛选出分别与心肌梗死、心力衰竭和缺血性脑卒中相关的代谢物,并进行时间截断的敏感性分析。随后,基于LightGBM算法构建代谢风险评分,最后将评分纳入传统临床模型,进行模型评价。结果 多变量Cox比例风险回归模型和时间截断的敏感性分析共发现了分别与心肌梗死、心力衰竭和缺血性脑卒中相关的119种、77种和12种代谢物,随后利用这些代谢物构建代谢风险评分。将构建的代谢风险评分纳入传统临床模型后,模型的预测性能显著提高,3种疾病的预测模型AUC提高到0.804、0.900和0.844,与仅利用传统临床模型相比,分别提高了0.145、0.198和0.188。此外,敏感性和特异性结果表明,这3个模型具有较高的预测准确性(敏感性:0.706、0.804、0.801;特异性:0.763、0.861、0.722),校准曲线和决策曲线同样显示模型具有较好的预测性能。结论 将代谢风险评分纳入传统临床模型可以显著提高2型糖尿病人群心血管疾病的预测准确性。

**关键词:**2型糖尿病;心肌梗死;心力衰竭;缺血性脑卒中;LightGBM;代谢风险评分;预测模型

中图分类号:R743.3

文献标志码:A

## Application of metabolomic risk score in predicting cardiovascular outcomes in patients with type 2 diabetes mellitus

SHEN Lujia<sup>1,2</sup>, LU Tianwei<sup>3</sup>, GONG Weiming<sup>1,2</sup>, ZHAO Yansong<sup>1,2</sup>,  
WANG Shukang<sup>1,2</sup>, YUAN Zhongshang<sup>1,2</sup>

(1. Department of Biostatistics, School of Public Health, Cheeloo College of Medicine, Shandong University, Jinan 250012, Shandong, China; 2. National Institute of Health and Medical Big Data, Jinan 250003, Shandong, China;

3. Department of Maternal and Child Health, School of Public Health, Cheeloo College of Medicine, Shandong University, Jinan 250012, Shandong, China)

**Abstract: Objective** To identify metabolites associated with myocardial infarction, heart failure, and ischemic stroke respectively in a population with type 2 diabetes mellitus, thereby constructing a metabolomic risk score and evaluating whether incorporation of the metabolomic risk score into a traditional clinical model improves the model's predictive accuracy. **Methods** Using data from the UK Biobank cohort, a multivariate Cox proportional risk regression model was applied to screen for metabolites associated with myocardial infarction, heart failure, and ischemic stroke, respectively, and a time-truncated sensitivity analysis was performed. Subsequently, metabolomic risk scores were constructed based on the LightGBM algorithm, and finally the scores were incorporated into traditional clinical models for model evaluation. **Results** Multivariate Cox proportional risk regression models and time-truncated sensitivity analyses identified a total of 119, 77, and 12 metabolites associated with myocardial infarction, heart failure, and ischemic stroke,

respectively, and these metabolites were then used to construct metabolomic risk scores. Incorporation of the constructed metabolomic risk scores into traditional clinical models significantly improved model prediction performance, with model AUCs improving to 0.804, 0.900, and 0.844 for the three diseases respectively, and the improvements were 0.145, 0.198, and 0.188, compared to utilizing only traditional clinical models. In addition, the sensitivity and specificity results showed that the three models had high prediction accuracy (sensitivity: 0.706, 0.804, 0.801; specificity: 0.763, 0.861, 0.722), and the calibration curves and decision curves likewise showed that the models had good prediction performance. **Conclusion** Incorporation of metabolomic risk scores into traditional clinical models can significantly improve the predictive accuracy of cardiovascular disease in the type 2 diabetes mellitus population.

**Key words:** Type 2 diabetes mellitus; Myocardial infarction; Heart failure; Ischemic stroke; LightGBM; Metabolomic risk score; Predictive modeling

心血管疾病仍然是世界疾病负担的主要原因,而中国是负担最重的国家之一,2019年中国约有450万人死于心血管疾病<sup>[1]</sup>。此外,2型糖尿病(type 2 diabetes mellitus, T2DM)人群患心血管疾病的风险是正常人的两倍<sup>[2]</sup>。队列研究表明,单纯糖尿病患者的后续病情发展中,首次并发症事件主要为慢性肾病、心力衰竭、卒中和心肌梗死<sup>[3]</sup>。因此,对此类疾病进行早期预测和尽早治疗具有重要的公共卫生意义。此外,缺血性脑卒中是T2DM患者最常见的中风亚型,其主要病理基础与糖尿病患者的血管病变特征高度相关<sup>[4]</sup>。

既往心血管疾病的预测模型仅基于经典的心血管危险因素,只取得了一般的预测准确性,并且之前开发的模型主要关注全局的心血管疾病负担,而忽略了不同亚型之间的复杂机制<sup>[5]</sup>。最近,美国心脏协会(American Heart Association, AHA)开发了结合心血管-肾脏-代谢健康的预测心血管疾病风险的新方程(predicting risk of CVD EVENTS, PREVENT),并用来预测罹患心力衰竭和动脉粥样硬化性心血管疾病(arteriosclerotic cardiovascular disease, ASCVD)的风险,还开发了附加模型用来预测基线患有慢性肾病或糖尿病人群的心血管疾病风险<sup>[6-7]</sup>。尽管已经取得很大进步,但仍需要提高这种敏感人群的心血管疾病风险评估的精度<sup>[8]</sup>。

代谢组学数据在分子水平上阐明了心血管疾病的早期病理生理变化<sup>[9]</sup>。核磁共振(nuclear magnetic resonance, NMR)代谢组学数据通过捕获与炎症、糖酵解、脂质代谢和氧化应激相关的各种代谢标记,提供了更全面的生物标志物来解释心血管疾病的复杂机制,比其他组学(如基因组学、转录组学和蛋白质组学)更接近疾病表型,并且它可以整合上游组学变异和环境因素暴露的影响。因此,代谢组学在T2DM患者的心血管疾病风险预测方面有很大的潜力。既往研究主要探索了代谢物在改善T2DM患者主要心血管不良事件预测中的作用<sup>[10]</sup>,

仅取得了相对较低的预测性能,且未针对具体的心血管疾病亚型进行预测。此外,只有少数研究通过构建代谢风险评分来进行预测<sup>[11]</sup>。本研究旨在利用英国生物银行(UK Biobank, UKB)的队列数据,初步筛选与3种心血管疾病(心肌梗死、心力衰竭和缺血性脑卒中)显著相关的代谢组生物标志物,进而构建代谢风险评分,并评估将代谢风险评分与PREVENT模型相结合是否可提高T2DM人群心血管疾病风险的预测能力。

## 1 资料与方法

### 1.1 资料

本研究数据来源于UKB。作为一项基于人群的前瞻性队列研究,UKB在2006年至2010年间共招募了50多万名年龄为40~69岁的参与者<sup>[12]</sup>,收集了包括全基因组基因分型、电子健康记录、磁共振成像、血液和尿液生物标志物检测以及人体测量等多维度数据,旨在加强对各种疾病的认识并完善其预防措施、诊断方式和治疗方法。该研究获得了英国国家医疗服务体系西北多中心研究伦理委员会的批准,所有参与者均提供了书面知情同意书,研究人员不需要单独的伦理批准,详细信息在线进行提供(<https://www.ukbiobank.ac.uk/media/0xsbmfmw/egf.pdf>)。本研究基于UKB申请号88159进行。

### 1.2 方法

#### 1.2.1 研究设计

在具有代谢组学数据的274 237名参与者中,根据国际疾病分类第10版(international statistical classification of diseases 10th revision, ICD-10)的编码定义,基线7 564例患有T2DM(E11),在排除基线患有心血管疾病[心力衰竭(I50)、心肌梗死(I21~I22)和缺血性脑卒中(I63)]的个体后剩余6 411例。进一步排除代谢物测量缺失超过10%(4例)的个体后,最终保留6 407例进行后续分析。

### 1.2.2 代谢物测量

UKB 通过高通量 NMR 代谢组学平台对随机抽取的 274 237 名参与者的基线血浆样本进行了代谢组学定量,量化了 249 个代谢组生物标记物(168 个绝对浓度和 81 个衍生比率),包括脂质、脂肪酸、氨基酸、酮体和其他必需的低分子量标记物,应用 R 包 ukbnmr<sup>[13]</sup>对代谢物进行重命名。

### 1.2.3 PREVENT 模型

PREVENT 方程是 AHA 从超过 600 万人样本中开发并验证的性别特异性无种族差异模型,用于预测 30~79 岁成年人罹患总体心血管疾病(以及动脉粥样硬化性心血管疾病和心力衰竭)的风险。该方程的附加模型通过添加糖化血红蛋白(glycated hemoglobin, HbA1c)作为预测因子,可用来预测 T2DM 人群的心血管疾病风险,包括:年龄、性别、吸烟(0 = 从未吸烟,1 = 曾经吸烟,2 = 正在吸烟)、HbA1c、估计肾小球滤过率(estimated glomerular filtration rate, eGFR)、体质量指数(body mass index, BMI)、收缩压(systolic blood pressure, SBP)、总胆固醇(total cholesterol, TC)、高密度脂蛋白胆固醇(high-density lipoprotein cholesterol, HDL\_C)、降压药物和降脂药物使用情况。eGFR 通过应用欧洲肾功能协会于 2021 年开发的基于肌酐的 CKD-Epi 方程<sup>[14]</sup>计算得出。基于最小缺失的目的,本研究利用高通量核磁共振平台测量的 TC、HDL\_C 和肌酐数值代替生化测量数值,既往研究表明这对于预测结果影响很小<sup>[10]</sup>。

### 1.2.4 电子健康数据

疾病结局是基于 UKB 电子健康记录中首次出现(Category 1712)的 3 位 ICD-10 编码确定的,该字段映射到死亡登记记录、初级保健数据、医院住院记录和自我报告的医疗状况中的 ICD-10 编码。基线日期是每个参与者首次参与 UKB 评估的日期,随访截至日期根据以下事件的最早日期确定:心血管疾病结局的发生、死亡、失访或随访结束(英格兰:2022 年 11 月;苏格兰:2021 年 8 月;威尔士:2021 年 5 月)。根据基线日期和随访截止日期计算每个参与者的随访时间(以年为单位)。

## 1.3 统计学处理

### 1.3.1 数据预处理

应用 R 4.3.2 软件,在最终保留的 6 407 例 T2DM 患者队列中,本研究使用 impute 包的 K-最邻近算法( $k=10$ )对缺失的代谢物数据进行插补,随后将所有代谢物浓度进行对数转换,并缩放到单位标准差。此外,使用 missForest 包对其他缺失变量进

行插补。

### 1.3.2 队列基本情况描述

正态分布的计量资料使用  $\bar{x}\pm s$  描述,组间比较采用成组  $t$  检验;非正态分布数据以  $M(P_{25}, P_{75})$  形式表示;计数资料以  $n(\%)$  表示,组间比较采用  $\chi^2$  检验。检验标准 = 0.05(双尾)。

### 1.3.3 多变量 Cox 比例风险回归分析

采用多变量 Cox 比例风险回归分析,分别筛选与心力衰竭、心肌梗死和缺血性脑卒中显著相关的代谢物,并将年龄、性别、BMI、吸烟和饮酒(0 = 从不,1 = 仅在特殊场合,2 = 1~3 次/月,3 = 1~2 次/周,4 = 3~4 次/周,5 = 每日或几乎每日)作为基础协变量进行调整。应用 Benjamini-Hochberg 错误发现率方法(false discovery rate, FDR)校正多重检验误差。作为敏感性分析,根据随访年限进行截断,将病例限制为随访 5 年和 10 年内发生,利用相同的方法筛选与 5 年和 10 年发病相关的代谢物。

### 1.3.4 代谢风险评分构建

对于多变量 Cox 比例风险回归分析和敏感性分析筛选出的与心血管结局具有显著关联的代谢物,取其并集,并利用 LightGBM 算法进行代谢风险评分的构建。作为传统梯度提升树改进的一种较为新颖的集成学习算法,LightGBM 可以用来处理本研究这一高维稀疏数据。相比于 XGBoost 算法,LightGBM 额外纳入了单边梯度采样(gradient-based one-side sampling, GOSS)和互斥特征捆绑(exclusive feature bundling, EFB)等策略,具有快速高效的优势;因随机森林(random forest, RF)算法随机选择特征,会有遗漏重要特征的风险,而 LightGBM 通过 EFB 策略,将全部特征压缩到更低维度,避免了信息损失,可以提升计算速度和评分的准确性。既往研究利用该算法构建的代谢风险评分,取得了较好的预测性能<sup>[15]</sup>。首先,对于每个心血管疾病,将数据随机分割为 70% 训练集和 30% 的验证集,使用 R 包 mlr3 采取 5 折交叉验证的方法在训练集中进行多轮超参数优化,使用最优参数集训练模型并以二元对数损失作为模型评估指标;然后,基于最终模型构建代谢风险评分。

### 1.3.5 模型评价

本研究使用受试者工作特征(receiver operating characteristic, ROC)曲线下面积(area under the curve, AUC)来评估将代谢风险评分添加到仅利用 PREVENT 模型变量的 Logistic 回归模型后对于 T2DM 人群心血管疾病的预测性能变化,采用

DeLong检验比较是否添加代谢风险评分的两个模型的AUC值。此外,本研究还评估了模型的敏感性和特异性,并利用校准曲线和决策曲线等来评估模型的预测性能。

## 2 结果

### 2.1 队列基本情况

研究队列共包括6407例基线患有T2DM且无

3种心血管疾病的UKB参与者,他们均具有可用的代谢组学数据。参与者(60.1±6.9)岁,包括3867例(60.4%)男性和2540例(39.6%)女性,T2DM病程为4.0(1.8, 7.2)年,糖化血红蛋白浓度为(53.2±13.9)mmol/mol。随访13.4(12.5, 14.3)年,有479例(7.5%)出现心肌梗死;随访13.3(12.4, 14.3)年,有827例(12.9%)出现心力衰竭;随访13.4(12.5, 14.3)年,有316例(4.9%)出现缺血性脑卒中。见表1。

表1 基线特征描述

Table 1 Description of baseline characteristics

变量	总例数 (n=6407)	心肌梗死		t/χ <sup>2</sup>	P
		非病例(n=5928)	病例(n=479)		
年龄/岁	60.1±6.9	59.9±6.9	61.6±6.5	-5.436	<0.001
性别				21.632	<0.001
男	3867(60.4)	3530(59.5)	337(70.4)		
女	2540(39.6)	2398(40.5)	142(29.6)		
吸烟				6.504	0.039
0	2957(46.2)	2753(46.4)	204(42.6)		
1	2750(42.9)	2543(42.9)	207(43.2)		
2	700(10.9)	632(10.7)	68(14.2)		
饮酒				12.384	0.030
0	1097(17.1)	994(16.8)	103(21.5)		
1	1289(20.1)	1181(19.9)	108(22.5)		
2	827(12.9)	769(13.0)	58(12.1)		
3	1417(22.1)	1324(22.3)	93(19.4)		
4	916(14.3)	850(14.3)	66(13.8)		
5	861(13.4)	810(13.7)	51(10.6)		
BMI	31.8±5.9	31.8±5.9	32.1±5.5	-1.203	0.229
eGFR/(mL·min·1.73 m <sup>2</sup> )	94.8±15.3	95.2±14.8	90.0±19.2	5.686	<0.001
HbA1c/(mmol/mol)	53.2±13.9	52.8±13.6	57.2±16.3	-5.716	<0.001
SBP/mmHg	141.1±17.1	140.9±16.9	144.0±19.2	-3.428	<0.001
TC/(mmol/L)	3.6±0.9	3.6±0.9	3.6±0.9	-0.197	0.844
HDL_C/(mmol/L)	1.1±0.2	1.1±0.3	1.0±0.3	4.076	<0.001
降压药	2558(39.9)	2302(38.8)	256(53.4)	39.604	<0.001
降脂药	2940(45.9)	2667(45.0)	273(57.0)	25.852	<0.001

变量	总例数 (n=6407)	心力衰竭		t/χ <sup>2</sup>	P
		非病例(n=5580)	病例(n=827)		
年龄/岁	60.1±6.9	59.7±6.9	62.6±5.9	-12.903	<0.001
性别				7.882	0.005
男	3867(60.4)	3331(59.7)	536(64.8)		
女	2540(39.6)	2249(40.3)	291(35.2)		
吸烟				10.739	0.005
0	2957(46.2)	2619(46.9)	338(40.9)		
1	2750(42.9)	2358(42.3)	392(47.4)		
2	700(10.9)	603(10.8)	97(11.7)		
饮酒				30.834	<0.001
0	1097(17.1)	905(16.2)	192(23.2)		
1	1289(20.1)	1127(20.2)	162(19.6)		
2	827(12.9)	725(13.0)	102(12.3)		
3	1417(22.1)	1263(22.6)	154(18.6)		

续表

变量	总例数 (n=6 407)	心力衰竭		t/ $\chi^2$	P
		非病例(n=5 580)	病例(n=827)		
4	916(14.3)	818(14.7)	98(11.9)		
5	861(13.4)	742(13.3)	119(14.4)		
BMI	31.8±5.9	31.6±5.8	33.4±6.2	-7.795	<0.001
eGFR/(mL·min·1.73m <sup>2</sup> )	94.8±15.3	95.7±14.5	88.5±18.7	10.553	<0.001
HbA1c/(mmol/mol)	53.2±13.9	52.7±13.5	56.4±15.9	-6.298	<0.001
SBP/mmHg	141.1±17.1	140.6±16.8	144.2±18.4	-5.289	<0.001
TC/(mmol/L)	3.6±0.9	3.7±0.9	3.6±0.8	3.332	0.001
HDL_C/(mmol/L)	1.1±0.2	1.1±0.3	1.0±0.3	3.706	<0.001
降压药	2 558(39.9)	2 149(38.5)	409(49.5)	35.754	<0.001
降脂药	2 940(45.9)	2 503(44.9)	437(52.8)	18.333	<0.001

变量	总例数 (n=6 407)	缺血性脑卒中		t/ $\chi^2$	P
		非病例(n=6 091)	病例(n=316)		
年龄/岁	60.1±6.9	60.0±6.9	61.9±6.2	-5.356	<0.001
性别				6.297	0.012
男	3 867(60.4)	3 655(60.0)	212(67.1)		
女	2 540(39.6)	2 436(40.0)	104(32.9)		
吸烟				4.576	0.101
0	2 957(46.2)	2 829(46.4)	128(40.5)		
1	2 750(42.9)	2 603(42.7)	147(46.5)		
2	700(10.9)	659(10.8)	41(13.0)		
饮酒				9.866	0.079
0	1 097(17.1)	1 045(17.2)	52(16.5)		
1	1 289(20.1)	1 227(20.1)	62(19.6)		
2	827(12.9)	792(13.0)	35(11.1)		
3	1 417(22.1)	1 345(22.1)	72(22.8)		
4	916(14.3)	854(14.0)	62(19.6)		
5	861(13.4)	828(13.6)	33(10.4)		
BMI	31.8±5.9	31.8±5.9	31.83±5.7	0.005	0.996
eGFR/(mL·min·1.73 m <sup>2</sup> )	94.8±15.3	95.0±14.9	89.6±19.9	4.812	<0.001
HbA1c/(mmol/mol)	53.2±13.9	53.0±13.8	55.7±15.3	-3.056	0.002
SBP/mmHg	141.1±17.1	140.8±16.9	146.3±20.4	-4.708	<0.001
TC/(mmol/L)	3.6±0.9	3.6±0.9	3.6±0.9	0.724	0.470
HDL_C/(mmol/L)	1.1±0.2	1.1±0.3	1.0±0.3	3.494	<0.001
降压药	2 558(39.9)	2 396(39.3)	162(51.3)	17.905	<0.001
降脂药	2 940(45.9)	2 785(45.7)	155(49.1)	1.364	0.243

注:1 mmHg=0.133 kPa。

## 2.2 代谢物 Cox 比例风险回归分析

在 249 种代谢物中,115 种与心肌梗死相关联,49 种与心力衰竭相关联,12 种与缺血性脑卒中相关联。其中,Albumin、Creatinine、S\_HDL\_FC\_pct 与 3 种心血管疾病均具有显著关联(图 1),107/115 个心肌梗死相关代谢物、44/49 个心力衰竭相关代谢

物、1/12 个缺血性脑卒中相关代谢物在敏感性分析中得到验证,并且新发现了 4 个和 28 个分别与心肌梗死和心力衰竭相关联的代谢物。

## 2.3 代谢风险评分构建

通过多变量 Cox 比例风险回归分析和敏感性分析,共发现了 119 种、77 种和 12 种分别与心肌梗

死、心力衰竭和缺血性脑卒中中具有显著关联的代谢物,将这些代谢物分别纳入3种心血管疾病代谢风险评分的构建中。经过超参数优化和模型训练,选择了在验证集上预测效果最佳的3个代谢物预测模型,分别保留了83种、77种和12种代谢物来构建基于特征重要性加权的代谢风险评分,其中,Creati-

nine、Albumin、ApoB\_by\_ApoA1这3种代谢物在3个模型中的重要性均排名前十,每种疾病前10个代谢物的重要性见图2。最终,对基线6407例T2DM患者分别构建了心肌梗死、心力衰竭和缺血性脑卒中的代谢风险评分,分别为 $(0.074 \pm 0.034)$ 、 $(0.129 \pm 0.079)$ 、 $(0.051 \pm 0.020)$ 分。

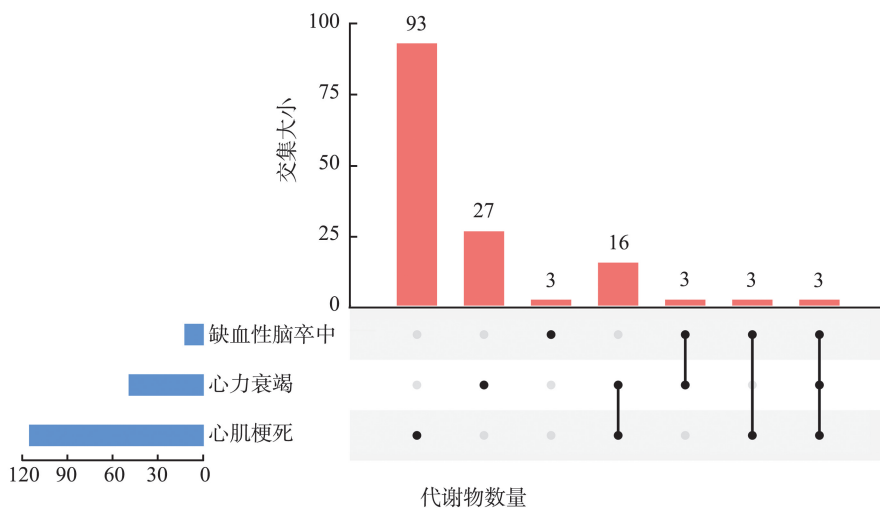


图1 与3种心血管疾病显著关联的代谢物集合图

Figure 1 Upset plots of metabolites significantly associated with three cardiovascular diseases

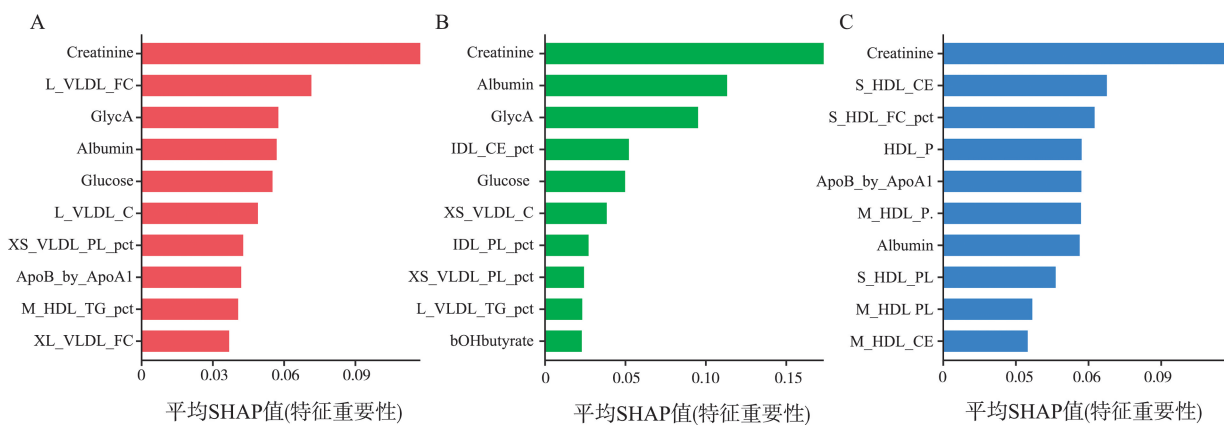


图2 3种心血管疾病预测模型中前10个代谢物的重要性排名

A: 心肌梗死; B: 心力衰竭; C: 缺血性脑卒中。

Figure 2 Importance ranking of the top 10 metabolites in three cardiovascular disease prediction models

A: Myocardial infarction; B: Heart failure; C: Ischemic stroke.

## 2.4 模型评价

将心肌梗死、心力衰竭和缺血性脑卒中的代谢风险评分分别纳入仅考虑 PREVENT 模型变量的预测模型后, AUC 值分别为 0.804、0.900 和 0.844, 与仅考虑 PREVENT 模型变量的预测模型 (AUC 分别为 0.659、0.702、0.656) 相比, 分别增加了 0.145、0.198 和 0.188 (DeLong 检验  $P$  值均  $< 0.001$ ), 见图 3。此外, 纳入代谢风险评分的 3 个

预测模型的敏感性和特异性均较好, 校准曲线的斜率、截距和 Brier 评分以及 Hosmer-Lemeshow 检验表明, 这 3 个模型的预测概率与实际发生概率一致性均较好, 能够准确估计 3 种心血管疾病的风险 (表 2, 图 4); 决策曲线显示, 模型的临床净收益在大多数阈值下都高于两种极端情况 (干预所有病例和不干预任何病例), 并且添加代谢风险评分的模型显示更高的临床净收益 (图 5)。

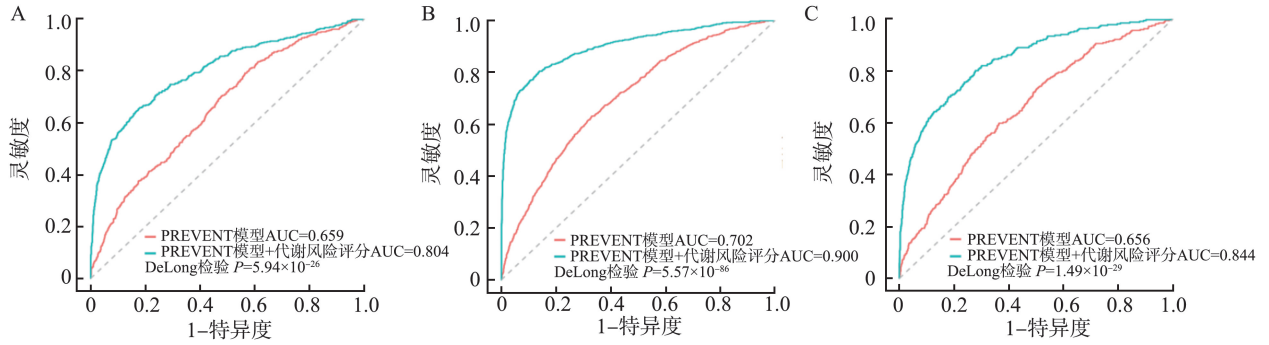


图3 3种心血管疾病预测模型的ROC曲线  
A:心肌梗死;B:心力衰竭;C:缺血性卒中。

Figure 3 ROC curves of predictive models for three cardiovascular diseases  
A: Myocardial infarction; B: Heart failure; C: Ischemic stroke.

表2 模型预测性能评估  
Table 2 Predictive performance evaluation of the models

疾病	模型	AUC (95%CI)	$P_{DeLong}$	敏感性	特异性	校准曲线			$P_{HL}$
						斜率	截距	Brier评分	
心肌梗死	PREVENT	0.659 (0.635~0.684)	$5.94 \times 10^{-26}$	0.791	0.437	0.968	0.002	0.067	0.1010
	PREVENT+代谢风险评分	0.804 (0.780~0.827)		0.706	0.763	1.052	-0.004	0.005	0.172
心力衰竭	PREVENT	0.702 (0.684~0.721)	$5.57 \times 10^{-86}$	0.630	0.670	0.997	0.000	0.105	0.661
	PREVENT+代谢风险评分	0.900 (0.887~0.913)		0.804	0.861	1.020	-0.003	0.058	0.807
缺血性脑卒中	PREVENT	0.656 (0.626~0.686)	$1.49 \times 10^{-29}$	0.598	0.636	0.907	0.005	0.046	0.613
	PREVENT+代谢风险评分	0.844 (0.820~0.867)		0.801	0.722	1.024	-0.001	0.038	0.710

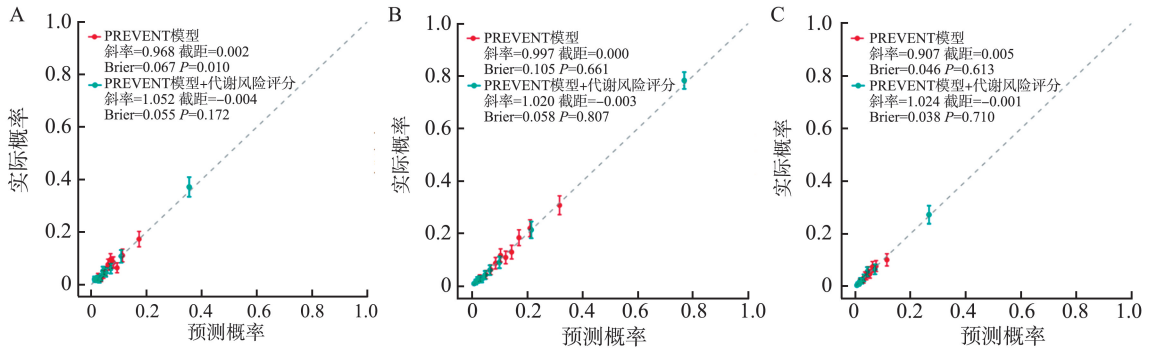


图4 3种心血管疾病预测模型的校准曲线  
A:心肌梗死;B:心力衰竭;C:缺血性卒中。

Figure 4 Calibration curves of predictive models for three cardiovascular diseases  
A: Myocardial infarction; B: Heart failure; C: Ischemic stroke.

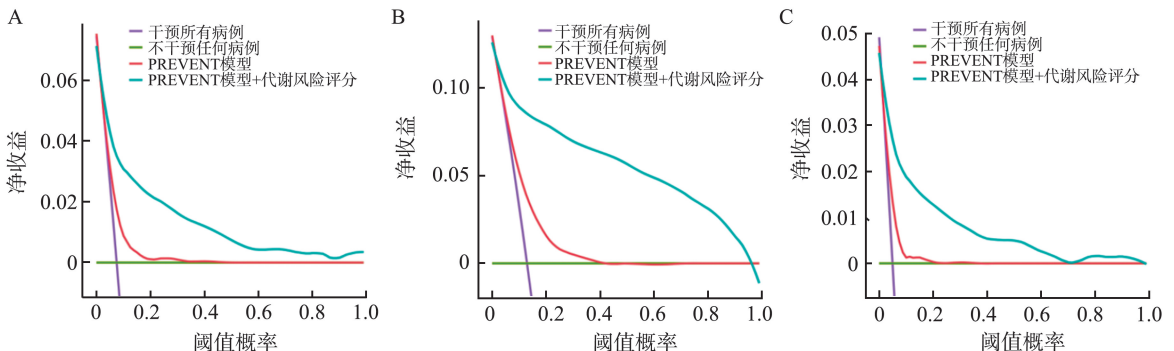


图5 3种心血管疾病预测模型的校准曲线  
A:心肌梗死;B:心力衰竭;C:缺血性卒中。

Figure 5 Decision curves of predictive models for three cardiovascular diseases  
A: Myocardial infarction; B: Heart failure; C: Ischemic stroke.

### 3 讨论

本研究基于 UKB 这一大型前瞻性队列,关注 T2DM 这一特殊群体的 3 种主要心血管疾病(心肌梗死、心力衰竭和缺血性脑卒中)风险预测,并结合代谢组学数据进行分析。首先通过多变量 Cox 比例风险回归分别筛选出与这 3 种疾病具有显著关联的代谢物,随后进行时间截断的敏感性分析来验证这些结果,并且又发现了一些新的代谢物。利用筛选出来的代谢物进一步通过 LightGBM 算法构建代谢风险评分,最后将代谢风险评分纳入仅考虑 PREVENT 模型变量的预测模型中。结果表明,纳入代谢风险评分后,对于每一种疾病的预测性能都得到了显著提升,心肌梗死、心力衰竭和缺血性脑卒中的 AUC 分别提高了 0.145、0.198 和 0.188,均具有较高的敏感性(0.706、0.804、0.801)和特异性(0.763、0.861、0.722),并且校准曲线和决策曲线也显示了模型的良好预测性能。但缺少来自其他队列的数据进行外部验证,导致研究结果在其他人群的预测性能未知,未来还需利用其他队列数据进行外部验证,以增强模型的稳健性。

先前基于代谢物的研究通常是在一般人群中进行的<sup>[15-17]</sup>,或是通过仅将单个或多个代谢产物直接添加到参考模型中来评估预测性能<sup>[10,18]</sup>。本研究结果表明,在特殊人群中构建代谢风险评分来预测心血管疾病或许具有更高的预测性能和临床应用价值。目前,在 T2DM 患者中进行心血管疾病风险预测的代谢组学研究相对较少,开发了一些仅考虑传统危险因素的临床预测模型,如:美国糖尿病协会建议使用美国心脏协会和美国心脏病学院于 2013 年开发的汇集队列方程(pooled cohort equations, PCEs)来评估糖尿病患者的动脉粥样硬化风险<sup>[19]</sup>;欧洲心脏病学会于 2023 年在 SCORE 模型的基础上开发了适用于糖尿病患者 10 年心血管疾病风险评估的新模型(SCORE2-Diabetes)<sup>[20]</sup>。此外,通过在小样本的 T2DM 队列中应用代谢风险评分预测心血管疾病也取得了较好的效果<sup>[11]</sup>,这表明添加代谢风险评分具有很大的潜在价值,进一步的心血管疾病亚型预测对于疾病特异性代谢物的发现和研究具有重要价值。

在多变量 Cox 比例风险回归分析中,3 种代谢物(Albumin、Creatinine、S\_HDL\_FC\_pct)与 3 种心血管疾病均具有显著关联,此外,在 3 种疾病的代谢风险评分中,Albumin、Creatinine、ApoB\_by\_ApoA1

的重要性均位列前十。血清白蛋白具有包括抗炎、抗氧化、抗凝和抗血小板聚集活性、胶体渗透作用等生理特性,而内皮功能障碍、炎症和氧化应激是动脉粥样硬化和许多心血管疾病发生、发展的重要因素<sup>[21]</sup>。多项流行病学证据表明,即使在调整了风险因素、体质量指数和炎症后,低血清白蛋白水平仍与缺血性心脏病<sup>[22]</sup>、心力衰竭<sup>[23-24]</sup>、和缺血性中风<sup>[25]</sup>的发生有关。血清肌酐是肌酸的酐形式,是肾功能的标志物,主要来源于肌肉代谢。此外,血清白蛋白和肌酐与肾功能密切相关,中度至重度慢性肾脏疾病患者的心血管疾病风险显著增加<sup>[26-27]</sup>,并且由血清肌酐估计的肾小球滤过率是 T2DM 人群心血管疾病的危险因素<sup>[6,20]</sup>。在整个高密度脂蛋白亚群中,小而致密、富含蛋白质的高密度脂蛋白表现出强大的抗动脉粥样硬化特性,这可归因于特定的蛋白质和脂质簇<sup>[28]</sup>。也有研究表明,游离胆固醇占小型高密度脂蛋白胆固醇总血脂的百分比(S\_HDL\_FC\_pct)与 T2DM 具有显著的遗传相关性,S\_HDL\_FC\_pct 的显著相关变体是 RS6073958,这是 PLTP 基因的表达数量性状位点,它参与磷脂从甘油三酸酯转移到高密度脂蛋白的过程,并且可能参与胆固醇代谢<sup>[29]</sup>,这表明 S\_HDL\_FC\_pct 对心血管疾病可能具有潜在影响。此外,几项队列研究表明,高密度脂蛋白颗粒大小及其胆固醇含量与心血管死亡和主要心血管结局相关<sup>[30]</sup>。多项队列研究表明,ApoB\_by\_ApoA1 可以反映动脉粥样硬化和抗动脉粥样硬化脂蛋白颗粒之间的胆固醇平衡,可以很好地预测心血管风险<sup>[31-33]</sup>。孟德尔随机化分析在遗传学方面的研究也表明,ApoB/ApoA1 比值升高与缺血性心脏病、主要心血管不良事件、主动脉瘤、脑缺血性疾病等多种心脏代谢性疾病均存在显著相关性<sup>[34]</sup>。

综上所述,本研究基于 UKB 队列,将代谢风险评分纳入基于传统危险因素的 PREVENT 模型中,显著提高了模型预测性能,具有一定的潜在临床价值,有助于进一步加强对 T2DM 这一心血管疾病高危人群疾病风险的预测精度。心血管-肾脏-代谢综合征的早期发现和干预往往与更大的临床获益相关,为了促进全生命周期心血管-肾脏-代谢健康,针对 CKM 2 期中糖尿病患者这一心血管疾病高危群体,对其进行早期筛查,尽早干预,可以在很大程度上预防其亚临床或临床心血管疾病的发生。

### 参考文献:

[1] Roth GA, Mensah GA, Johnson CO, et al. Global burden

- of cardiovascular diseases and risk factors, 1990—2019; update from the GBD 2019 study[J]. *J Am Coll Cardiol*, 2020, 76(25): 2982-3021.
- [2] Collaboration ERF, Sarwar N, Gao P, et al. Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies[J]. *Lancet*, 2010, 375(9733): 2215-2222.
- [3] Birkeland KI, Bodegard J, Eriksson JW, et al. Heart failure and chronic kidney disease manifestation and mortality risk associations in type 2 diabetes: a large multinational cohort study[J]. *Diabetes Obes Metab*, 2020, 22(9): 1607-1618.
- [4] Maida CD, Daidone M, Pacinella G, et al. Diabetes and ischemic stroke: an old and new relationship an overview of the close interaction between these diseases[J]. *Int J Mol Sci*, 2022, 23(4): 2397. doi: 10.3390/ijms23042397
- [5] Jia RR, Wang Q, Huang HY, et al. Cardiovascular disease risk models and dementia or cognitive decline: a systematic review[J]. *Front Aging Neurosci*, 2023, 15: 1257367. doi:10.3389/fnagi.2023.1257367
- [6] Khan SS, Matsushita K, Sang YY, et al. Development and validation of the American heart association's PREVENT equations[J]. *Circulation*, 2024, 149(6): 430-449.
- [7] Khan SS, Coresh J, Pencina MJ, et al. Novel prediction equations for absolute risk assessment of total cardiovascular disease incorporating cardiovascular-kidney-metabolic health: a scientific statement from the American heart association[J]. *Circulation*, 2023, 148(24): 1982-2004.
- [8] Dziopa K, Asselbergs FW, Gratton J, et al. Cardiovascular risk prediction in type 2 diabetes: a comparison of 22 risk scores in primary care settings[J]. *Diabetologia*, 2022, 65(4): 644-656.
- [9] McGarrah RW, Crown SB, Zhang GF, et al. Cardiovascular metabolomics[J]. *Circ Res*, 2018, 122(9): 1238-1258.
- [10] Xie RJ, Seum T, Sha S, et al. Improving 10-year cardiovascular risk prediction in patients with type 2 diabetes with metabolomics[J]. *Cardiovasc Diabetol*, 2025, 24(1): 18. doi:10.1186/s12933-025-02581-3
- [11] Huang Z, Klaric L, Krasauskaite J, et al. Combining serum metabolomic profiles with traditional risk factors improves 10-year cardiovascular risk prediction in people with type 2 diabetes[J]. *Eur J Prev Cardiol*, 2023, 30(12): 1255-1262.
- [12] Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age[J]. *PLoS Med*, 2015, 12(3): e1001779. doi:10.1371/journal.pmed.1001779
- [13] Ritchie SC, Surendran P, Karthikeyan S, et al. Quality control and removal of technical variation of NMR metabolic biomarker data in ~120, 000 UK Biobank participants[J]. *Sci Data*, 2023, 10(1): 64. doi:10.1038/s41597-023-01949-y
- [14] Inker LA, Eneanya ND, Coresh J, et al. New creatinine- and cystatin C-based equations to estimate GFR without race[J]. *N Engl J Med*, 2021, 385(19): 1737-1749.
- [15] Qiang YX, You J, He XY, et al. Plasma metabolic profiles predict future dementia and dementia subtypes: a prospective analysis of 274, 160 participants[J]. *Alzheimers Res Ther*, 2024, 16(1): 16. doi:10.1186/s13195-023-01379-3
- [16] Zhang SY, Wang Z, Wang YJ, et al. A metabolomic profile of biological aging in 250, 341 individuals from the UK Biobank[J]. *Nat Commun*, 2024, 15(1): 8081. doi:10.1038/s41467-024-52310-9
- [17] Julkunen H, Cichońska A, Tiainen M, et al. Atlas of plasma NMR biomarkers for health and disease in 118, 461 individuals from the UK Biobank[J]. *Nat Commun*, 2023, 14(1): 604. doi:10.1038/s41467-023-36231-7
- [18] McGranaghan P, Saxena A, Rubens M, et al. Predictive value of metabolomic biomarkers for cardiovascular disease risk: a systematic review and meta-analysis[J]. *Biomarkers*, 2020, 25(2): 101-111.
- [19] American Diabetes Association Professional Practice Committee. 10. Cardiovascular disease and risk management: standards of care in diabetes—2024[J]. *Diabetes Care*, 2024, 47(Suppl 1): S179-S218.
- [20] 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.
- [21] Arques S. Human serum albumin in cardiovascular diseases[J]. *Eur J Intern Med*, 2018, 52: 8-12. doi:10.1016/j.ejim.2018.04.014
- [22] Nelson JJ, Liao D, Sharrett AR, et al. Serum albumin level as a predictor of incident coronary heart disease: the Atherosclerosis Risk in Communities (ARIC) study[J]. *Am J Epidemiol*, 2000, 151(5): 468-477.
- [23] Gopal DM, Kalogeropoulos AP, Georgiopoulou VV, et al. Serum albumin concentration and heart failure risk The Health, Aging, and Body Composition Study[J]. *Am Heart J*, 2010, 160(2): 279-285.
- [24] Filippatos GS, Desai RV, Ahmed MI, et al. Hypoalbuminaemia and incident heart failure in older adults[J]. *Eur J Heart Fail*, 2011, 13(10): 1078-1086.
- [25] Xu WH, Dong CH, Rundek T, et al. Serum albumin

- levels are associated with cardioembolic and cryptogenic ischemic strokes: Northern Manhattan Study[J]. *Stroke*, 2014, 45(4): 973-978.
- [26] Visseren FLJ, Mach F, Smulders YM, et al. 2021 ESC guidelines on cardiovascular disease prevention in clinical practice: developed by the task force for cardiovascular disease prevention in clinical practice with representatives of the European Society of Cardiology and 12 medical societies with the special contribution of the European Association of Preventive Cardiology (EAPC) [J]. *Rev Esp Cardiol (Engl Ed)*, 2022, 75(5): 429. doi:10.1016/j.rec.2022.04.003
- [27] Matsushita K, Kaptoge S, Hageman SHJ, et al. Including measures of chronic kidney disease to improve cardiovascular risk prediction by SCORE2 and SCORE2-OP[J]. *Eur J Prev Cardiol*, 2023, 30(1): 8-16.
- [28] Camont L, John Chapman M, Kontush A. Biological activities of HDL subpopulations and their relevance to cardiovascular disease[J]. *Trends Mol Med*, 2011, 17(10): 594-603.
- [29] Bocher O, Singh A, Huang Y, et al. Disentangling the consequences of type 2 diabetes on targeted metabolite profiles using causal inference and interaction QTL analyses[J]. *PLoS Genet*, 2024, 20(12): e1011346. doi:10.1371/journal.pgen.1011346
- [30] Teis A, Cediél G, Amigó N, et al. Particle size and cholesterol content of circulating HDL correlate with cardiovascular death in chronic heart failure[J]. *Sci Rep*, 2021, 11(1): 3141. doi:10.1038/s41598-021-82861-6
- [31] Liu Y, Jia SD, Yuan DS, et al. Apolipoprotein B/A-I ratio predicts lesion severity and clinical outcomes in diabetic patients with acute coronary syndrome[J]. *Circ J*, 2020, 84(7): 1132-1139.
- [32] McQueen MJ, Hawken S, Wang XY, et al. Lipids, lipoproteins, and apolipoproteins as risk markers of myocardial infarction in 52 countries (the INTERHEART study): a case-control study [J]. *Lancet*, 2008, 372(9634): 224-233.
- [33] Walldius G, Jungner I, Holme I, et al. High apolipoprotein B, low apolipoprotein A-I, and improvement in the prediction of fatal myocardial infarction (AMORIS study): a prospective study [J]. *Lancet*, 2001, 358(9298): 2026-2033.
- [34] Fu C, Liu DB, Liu Q, et al. Revisiting an old relationship: the causal associations of the ApoB/ApoA1 ratio with cardiometabolic diseases and relative risk factors-a mendelian randomization analysis[J]. *Cardiovasc Diabetol*, 2024, 23(1): 51. doi:10.1186/s12933-024-02140-2

(编辑:相峰)