

基于血清 miR-504-3p、miR-151a-5p 及 VEGF 的多指标模型 预测晚期非小细胞肺癌患者长期生存与转移风险

黄 凯¹⁾, 丁海兵²⁾

(1)扬州大学附属泰州第二人民医院检验科; 2)胸外科, 江苏 泰州 225500)

[摘要] 目的 建立并内部验证结合血清中 miR-504-3p、miR-151a-5p 与血管内皮生长因子 (vascular endothelial growth factor, VEGF) 的多指标预测模型, 评估其对预测晚期非小细胞肺癌 (non-small cell lung cancer, NSCLC) 患者 5 年生存率及转移风险的预测价值。方法 采用回顾性队列研究, 选取 2016 年 1 月至 2019 年 12 月本院确诊的 264 例晚期 NSCLC 患者, 均完成 5 年追踪随访, 按生存状况分为存活组 ($n = 58$) 与死亡组 ($n = 206$)。采用多因素 Cox 回归筛选影响晚期 NSCLC 患者 5 年生存的关键因素并构建线性预后指数 (prognostic index, PI) 模型; 绘制时间依赖受试者工作特征曲线 (receiver operating characteristic curve, ROC) 曲线, 计算曲线下面积 (area under the curve, AUC) 评估模型区分度; 通过 Bootstrap 法 (重复 500 次) 进行内部验证, 以校准曲线及校准截距、斜率评估模型校准度; 采用决策曲线分析 (decision curve analysis, DCA) 评估模型的临床净收益。结果 miR-151a-5p、VEGF、临床分期 IV 期是晚期 NSCLC 患者 5 年生存的独立危险因素 ($P < 0.05$), miR-504-3p 是保护因素 ($P < 0.05$)。时间依赖 ROC 曲线显示, 模型预测死亡风险的 AUC 由 1 年的 0.703 (95% CI: 0.637 ~ 0.769) 提升至 5 年的 0.859 (95% CI: 0.814 ~ 0.904), 预测转移风险的 AUC 由 1 年的 0.694 (95% CI: 0.624 ~ 0.764) 提升至 5 年的 0.847 (95% CI: 0.802 ~ 0.892), 区分能力随时间推移逐渐增强。校准曲线及量化指标显示, 模型对 3、4、5 年死亡及转移风险的预测校准度良好 (校准截距接近 0, 校准斜率接近 1), 但 1、2 年校准度欠佳。DCA 结果显示, 在预测 3 年、4 年和 5 年死亡和转移风险时, PI 模型在广泛的阈值概率范围内均能提供高于“全部干预”和“全部不干预”策略的净获益。根据 PI 中位数将 264 例患者分为高风险组 ($n = 175$) 和低风险组 ($n = 89$)。Kaplan-Meier 生存分析显示, 低风险组患者的 5 年生存率与无转移率均显著优于高风险组 ($P < 0.05$)。结论 血清 miR-504-3p、miR-151a-5p、VEGF 与晚期 NSCLC 患者长期生存及转移风险密切相关。整合三者及临床分期的多指标预测模型, 对晚期 NSCLC 患者 3 年及以上的长期生存与转移风险具有良好的预测准确性, 有助于临床识别高复发转移风险的患者。

[关键词] 非小细胞肺癌; 微小 RNA; 血管内皮生长因子; 预测模型

[中图分类号] R734.2 **[文献标志码]** A **[文章编号]** 2095 - 610X(2026)06 - 0141 - 11

A Multi-index Model Based on Serum miR-504-3p, miR-151a-5p, and VEGF for Predicting Long-term Survival and Metastatic Risk in Patients with Advanced Non-small Cell Lung Cancer

HUANG Kai¹⁾, DING Haibing²⁾

(1) Department of Laboratory Medicine; 2) Department of Thoracic Surgery, Taizhou Second People's Hospital Affiliated to Yangzhou University, Taizhou Jiangsu 225500, China)

[Abstract] **Objective** To establish and internally validate a multi-indicator predictive model incorporating serum miR-504-3p, miR-151a-5p, and vascular endothelial growth factor (VEGF) to evaluate its predictive

[收稿日期] 2026 - 01 - 07

[基金项目] 江苏大学临床医学科技发展基金 (JLY2021187)

[作者简介] 黄凯 (1977 ~), 男, 江苏泰州人, 医学学士, 副主任技师, 主要从事生化分子方面的研究工作。

value for the 5-year survival rate and metastasis risk in patients with advanced non-small cell lung cancer (NSCLC).

Methods A retrospective cohort study was conducted on 264 advanced NSCLC patients diagnosed between January 2016 and December 2019 at our institution, all of whom completed a 5-year follow-up. Patients were stratified into survival ($n = 58$) and mortality ($n = 206$) groups based on survival status. Multivariate Cox regression analysis was used to identify key factors affecting 5-year survival in advanced NSCLC patients, and a linear prognostic index (PI) model was constructed. Time-dependent receiver operating characteristic (ROC) curves were generated, and the area under the curve (AUC) was calculated to evaluate model discrimination. Internal validation was performed using the Bootstrap method (500 iterations), with calibration curves, intercepts, and slopes used to assess model calibration. Decision curve analysis (DCA) was employed to evaluate the clinical net benefit of the model. **Results** miR-151a-5p, VEGF, and clinical stage IV were independent risk factors for 5-year survival in patients with advanced NSCLC ($P < 0.05$), while miR-504-3p was a protective factor ($P < 0.05$). Time-dependent ROC curves demonstrated that the model's AUC for predicting mortality risk increased from 0.703 (95%CI: 0.637 ~ 0.769) at 1 year to 0.859 (95%CI: 0.814 ~ 0.904) at 5 years, and for predicting metastasis risk increased from 0.694 (95%CI: 0.624 ~ 0.764) at 1 year to 0.847 (95%CI: 0.802 ~ 0.892) at 5 years, with discriminatory ability progressively strengthening over time. Calibration curves and quantitative metrics demonstrated good calibration of the model for predicting 3-, 4-, and 5-year mortality and metastasis risk (calibration intercept approaching 0, calibration slope approaching 1), but suboptimal calibration at 1 and 2 years. DCA results demonstrated that for predicting 3-, 4-, and 5-year mortality and metastasis risk, the PI model provided net benefit superior to both "intervene all" and "intervene none" strategies across a wide range of threshold probabilities. Using the PI median, 264 patients were stratified into high-risk ($n = 175$) and low-risk ($n = 89$) groups. Kaplan-Meier survival analysis showed that 5-year survival and metastasis-free rates in the low-risk group were significantly superior to the high-risk group ($P < 0.05$). **Conclusion** Serum miR-504-3p, miR-151a-5p, and VEGF are closely associated with long-term survival and metastasis in advanced NSCLC patients. A multi-indicator model integrating these three markers and clinical staging demonstrates good predictive accuracy for long-term survival and metastasis risk at 3 years and beyond in advanced NSCLC patients and may facilitate clinical identification of patients at high risk for recurrence and metastasis.

[**Key words**] Non-small cell lung cancer; microRNA; Vascular endothelial growth factor; Predictive model

非小细胞肺癌(non-small cell lung cancer, NSCLC)是肺癌的主要亚型^[1-2]。尽管近年来靶向治疗、免疫治疗及多学科综合诊疗模式取得了显著进展,但晚期 NSCLC 患者因肿瘤侵袭转移能力强,长期预后仍不理想^[3-4]。研究显示,患者预后存在明显的个体异质性,不仅源于肿瘤的临床病理特征,更与肿瘤的分子生物学行为密切相关^[5]。因此,探寻能够精准反映肿瘤侵袭、转移潜能及预后的分子标志物,并构建有效的风险预测模型,对制定精准个体化治疗方案、提升患者临床结局意义重大。

近年来,多种微小 RNA(microRNA, miRNA)被证实 NSCLC 患者血清中异常表达,并与预后相关^[6-9]。miR-504-3p 在多种肿瘤中发挥抑癌作用,通过调控下游靶基因抑制肿瘤发展。miR-151a-5p 多表现为促癌功能,与肿瘤侵袭转移密切相关。除 miRNA 外,肿瘤相关细胞因子表达失

调,在 NSCLC 进展与转移过程中起关键作用。血管内皮生长因子(vascular endothelial growth factor, VEGF)可通过促进新生血管形成加速肿瘤转移,其高表达与肿瘤富血管化、侵袭转移及不良预后明确相关^[10]。肿瘤的发生演进是多基因、多通路共同参与的繁复进程,单一生物标志物预测效能受限,整合多个具有互补生物学功能的分子标志物,可能构建出预测效能更佳的综合模型。本研究通过前期预实验验证了血清 miR-504-3p、miR-151a-5p 在晚期 NSCLC 患者与健康人群中的表达差异,二者在晚期 NSCLC 患者中存在显著的异常表达。为此,本研究采用回顾性队列研究设计,通过整合血清 miR-504-3p、miR-151a-5p 及 VEGF 构建多指标预测模型,探讨其对晚期 NSCLC 患者 5 年长期生存及转移风险的预测价值,旨在为临床精准评估预后、制定个体化干预策略提供理论依据。

1 资料与方法

1.1 研究对象

本研究为回顾性队列研究, 采用连续入组的方法, 选取 2016 年 1 月至 2019 年 12 月于扬州大学附属泰州第二人民医院就诊的 NSCLC 患者 264 例, 通过医院电子病历系统按时间顺序逐例筛选。纳入标准^[11]: (1) 经病理组织学检查确诊为 NSCLC^[12]; (2) 临床分期为ⅢB~Ⅳ期; (3) 年龄 18~80 岁; (4) 确诊时未接受过任何抗肿瘤治疗; (5) 基线临床病理资料和随访资料完整。排除标准^[13]: (1) 合并其他恶性肿瘤; (2) 严重脏器功能障碍; (3) 严重感染、自身免疫性疾病等; (4) 随访过程中失访或因非肿瘤相关原因死亡; (5) 血清样本不合格或检测数据缺失。所有患者确诊后均按规范接受抗肿瘤治疗^[13], 治疗方案包括: (1) 化疗; (2) 靶向治疗; (3) 免疫治疗; (4) 放疗; (5) 上述两种及以上方案的联合治疗, 治疗方案由多学科诊疗团队评估后确定。由于 2016 年至 2019 年晚期 NSCLC 治疗策略不断更新, 本研究将治疗年份作为校正变量纳入多因素 Cox 回归模型, 以控制潜在的时间偏倚。本研究经扬州大学附属泰州第二人民医院伦理委员会批准通过(202503196)。

样本量估算: 本研究依据多因素 Cox 回归分析的事件数要求确定样本量, 以主要结局(死亡)为终点。根据文献回顾^[14]及临床经验, 初步筛选可能影响晚期 NSCLC 患者长期预后的候选变量, 包括年龄、性别、吸烟史、临床分期、靶向治疗、ECOG 评分、miR-504-3p、miR-151a-5p、VEGF 等共 10 项潜在预测变量。按 EPV 的经验标准, 为保障模型稳定, 每个预测变量至少需 10~20 个事件。基于前期预实验数据, 预估死亡率约为 75%。为满足多因素分析要求, 设定至少需要 100 个事件数, 则所需总样本量 $n = \text{事件数}/\text{预期}$

发生率 $=100/0.75 \approx 133$ 例。考虑到随访过程中可能存在的失访率(预设 $\leq 5\%$), 最终本研究纳入 264 例患者, 完成 5 年随访后发生死亡事件 206 例, 实际结局事件数(206)与最终纳入模型的变量数(4)之比为 51.5:1, 远高于经验标准的下限(10:1), 确保了统计分析的检验效能。

1.2 研究方法

1.2.1 临床资料收集 采集患者基线临床病理信息, 包括年龄、性别、BMI、吸烟及饮酒情况、肿瘤大小、临床分期、ECOG 评分等。

1.2.2 血清样本处理与标志物检测 (1) 样本采集与处理: 所有患者在初次诊断且未接受任何抗肿瘤治疗前, 于清晨空腹状态下采集外周静脉血 5 mL。在室温环境静置 30 min 后, 以 3000 r/min 的转速离心 15 min, 分离出血清, 分装好置于 -80°C 超低温冰箱长期留存备用。

(2) 血清 miR-504-3p、miR-151a-5p 表达水平检测: 使用美国 Invitrogen 公司的 Trizol 试剂提取血清总 RNA, 用美国 Thermo Fisher Scientific 公司的 Nanodrop 2000 超微量分光光度计检测 RNA 纯度(A260/A280 比值处于 1.8~2.0 范围)与浓度。按日本 TaKaRa 公司 PrimeScript RT Reagent Kit 说明书进行逆转录, 合成 cDNA。用日本 TaKaRa 公司的试剂盒, 以 U6 snRNA 为内参基因, 通过 qRT-PCR 检测 miR-504-3p、miR-151a-5p 表达水平。引物由上海生工生物工程股份有限公司合成, 所有引物序列均经 NCBI BLAST 工具比对验证, 证实具有高度特异性, 无非特异性结合位点。qRT-PCR 反应体系: $2 \times \text{SYBR Premix Ex Taq II}$ 10 μL , 上游引物 0.4 μL , 下游引物 0.4 μL , cDNA 模板 2 μL , RNase-Free H_2O 7.2 μL , 总反应体积 20 μL 。反应条件: 95°C 预变性 30 s; 95°C 变性 5 s, 60°C 退火 30 s, 共 40 个循环; 最后 95°C 15 s, 60°C 1 min, 95°C 15 s 绘制溶解曲线。采用 $2^{-\Delta\Delta\text{Ct}}$ 法计算血清 miR-504-3p、miR-151a-5p 相对表达水平, 具体序列及反应条件见表 1。

表 1 qRT-PCR 引物序列及反应条件

Tab. 1 qRT-PCR primer sequence and reaction conditions

基因名称	引物类型	引物序列(5'→3')	产物长度(bp)	退火温度($^{\circ}\text{C}$)
miR-504-3p	上游引物	CGG CGG TGA GAG AGT AGT GAT	72	60
	下游引物	AGT GCA GGG TCC GAG GTA TT		
miR-151a-5p	上游引物	CGG CGG ACT GAG GCT ACA ACT	70	60
	下游引物	AGT GCA GGG TCC GAG GTA TT		
U6	上游引物	CTC GCT TCG GCA GCA CA	94	60
	下游引物	AAC GCT TCA CGA ATT TGC GT		

(3)血清VEGF水平检测:运用酶联免疫吸附试验检测血清中 VEGF 浓度,所用试剂盒来自上海酶联生物科技有限公司(货号:ml064294,规格:48T/96T),操作严格遵循试剂盒说明。依次将标准品和样本加入酶标孔,于 37℃ 孵育 2 h;弃去孔内液体,用洗涤液洗 3 次,每次 30s;加入酶标试剂,37℃ 孵育 1 h;再次重复洗涤步骤;加入显色剂,37℃ 避光孵育 15 min;最后加入终止液终止反应,随即用美国 Bio-Rad 公司的 iMark 型酶标仪测定 450 nm 波长处的吸光度。根据标准曲线计算血清 VEGF 浓度。

1.2.3 预后评估及分组 本研究所有纳入研究对象均已通过门诊复查、电话等方式完成为期 5 年的随访,随访截止日期为 2024 年 12 月 31 日。随访时间点定义为确诊后满 1 年、2 年、3 年、4 年、5 年(± 1 个月)。主要结局指标为总生存期,定义为从确诊日期至因任何原因死亡或末次随访日期。根据随访结束时的生存状态,将患者分为存活组($n = 58$)和死亡组($n = 206$)。次要结局指标为无转移生存期,定义为从确诊日期至首次出现影像学证实的远处转移(任何部位)的时间,若患者在出现转移前死亡则视为竞争风险事件。本研究中共 190 例患者在随访期间出现新发远处转移。

各时间点结局数据的获取:对于每个预设的时间点(1、2、3、4、5 年),记录患者在此时刻的生存状态(存活/死亡)及是否已发生转移。若患者在某一时间点之前已死亡或失访,则该时间点之后的随访数据视为缺失;若患者在某一时间点之后才出现事件,则该时间点记录为未发生事件。时间依赖 ROC 曲线分析中,各时间点的预测概率基于该时间点之前仍处于风险集中的患者计算得出。本研究所有入组患者均完整随访或记录死亡/转移事件,无缺失数据。

1.3 统计学分析

采用 SPSS 26.0 统计软件和 R 4.2.1 软件进行数据分析。若计量资料呈正态分布,用均数 $\pm$ 标准差($\bar{x} \pm s$)表示,组间对比用 t 检验;非正态分布的计量资料,以中位数(四分位数) $M(P_{25}, P_{75})$ 表示,组间比较选 Mann-Whitney U 检验。计数资料以例数(百分比) $[n(\%)]$ 表述,组间对比采用 χ^2 检验。模型构建:首先采用单因素 Cox 回归分析初步筛选与 5 年生存相关的潜在预后因素(纳入标准 $P < 0.05$)。在进行多因素 Cox 回归分析前,首先对连续型自变量与因变量之间的线性关系进行检验。采用受限立方样条(restricted cubic spline, RCS)方法评估各连续变量的非线性趋势,设定 3 个节

点(位于第 10、50、90 百分位数)。采用方差膨胀因子(variance inflation factor, VIF)对纳入多因素模型的变量进行共线性诊断, $VIF < 10$ 提示变量间不存在显著的多重共线性问题,可同时纳入多因素 Cox 回归模型进行分析。将单因素分析中 $P < 0.05$ 的变量纳入多因素 Cox 回归模型,采用基于似然比检验的逐步回归法进行变量筛选,纳入标准为 $P < 0.05$,剔除标准为 $P > 0.10$ 。为控制治疗策略随时间变化可能引入的偏倚,将“治疗年份”(按 2016—2019 年分类)作为校正变量纳入多因素模型。基于多因素 Cox 回归分析得到的各独立影响因素及其对应的回归系数(B),构建线性预后指数(prognostic index, PI)模型: $PI = \sum(\beta_i \times X_i)$ 。模型评价:运用时间依赖性受试者工作特征曲线(receiver operating characteristic curve, ROC)曲线评估模型于不同时间点对死亡和转移风险的辨别能力,计算各时间点 AUC 及其 95%CI。采用 Bootstrap 重抽样法(重复 500 次)开展内部验证,绘制不同时间点死亡与转移的校准曲线,并计算校准截距(calibration intercept)和校准斜率(calibration slope),量化评估模型预测概率与实际观察概率的吻合度。校准截距反映整体校准度(理想值为 0),校准斜率反映预测风险的准确性(理想值为 1)。采用决策曲线分析(decision curve analysis, DCA)评估模型的临床净收益。模型应用:根据 PI 中位数将 264 例患者分为高风险组和低风险组。采用 Kaplan-Meier 法绘制生存曲线,用 Log-rank 检验比较两组患者总生存差异。针对转移风险分析,死亡是转移的竞争风险事件,采用 Fine-Gray 竞争风险模型校正,基于校正后结果绘制无远处转移生存曲线,并采用 Gray 检验比较两组患者的无远处转移生存差异。以 $P < 0.05$ 为差异具有统计学意义。

2 结果

2.1 存活组和死亡组临床资料比较

死亡组年龄、吸烟史、IV 期占比高于存活组($P < 0.05$),靶向治疗占比低于存活组($P < 0.05$),见表 2。

2.2 血清 miR-504-3p、miR-151a-5p、VEGF 比较

存活组患者 miR-504-3p 表达水平显著高于死亡组($P < 0.05$),miR-151a-5p 表达水平及 VEGF 水平显著低于死亡组($P < 0.05$),见表 3。

2.3 生存的多因素 Cox 回归分析及模型构建

在进行多因素 Cox 回归分析前,对连续型自

表 2 存活组和死亡组临床资料比较 $[(\bar{x} \pm s)/n(\%)/M(P_{25}, P_{75})]$

Tab. 2 Comparison of clinical characteristics between two groups $[(\bar{x} \pm s)/n(\%)/M(P_{25}, P_{75})]$

项目	存活组(n=58)	死亡组(n=206)	$\chi^2/t/Z$	P
年龄(岁)	63.42 ± 8.17	67.83 ± 9.24	3.290	<0.001*
性别			0.067	0.796
男	36(62.07)	124(60.19)		
女	22(37.93)	82(39.81)		
BMI(kg/m ²)	22.86 ± 2.94	22.41 ± 3.17	0.970	0.333
肿瘤直径(cm)	4.3(3.2, 5.8)	4.7(3.6, 6.1)	-1.320	0.186
吸烟史			5.469	0.019*
有	27(46.55)	131(63.59)		
无	31(53.45)	75(36.41)		
饮酒史			0.044	0.834
有	16(27.59)	54(26.21)		
无	42(72.41)	152(73.79)		
病理类型			0.606	0.436
鳞癌	19(32.76)	79(38.35)		
腺癌	39(67.24)	127(61.65)		
临床分期			4.364	0.037*
IIIB-IIIC期	25(43.10)	59(28.64)		
IV期	33(56.90)	147(71.36)		
靶向治疗			4.025	0.045*
有	39(67.24)	108(52.43)		
无	19(32.76)	98(47.57)		
ECOG评分			2.201	0.333
0	18(31.03)	45(21.84)		
1	32(55.17)	125(60.68)		
2	8(13.79)	36(17.48)		
治疗年份			1.284	0.733
2016年	13(22.41)	54(26.21)		
2017年	15(25.86)	52(25.24)		
2018年	16(27.59)	48(23.30)		
2019年	14(24.14)	52(25.24)		

* $P < 0.05$ 。

表 3 存活组和死亡组血清 miR-504-3p、miR-151a-5p、VEGF 比较 $(\bar{x} \pm s)$

Tab. 3 Comparison of serum miR-504-3p, miR-151a-5p and VEGF levels between two groups $(\bar{x} \pm s)$

项目	存活组(n=58)	死亡组(n=206)	t	P
miR-504-3p	1.24 ± 0.35	0.93 ± 0.21	8.433	<0.001*
miR-151a-5p	1.02 ± 0.26	1.56 ± 0.42	9.296	<0.001*
VEGF(pg/mL)	375.68 ± 86.32	526.45 ± 124.56	8.646	<0.001*

* $P < 0.05$ 。

变量(miR-504-3p、miR-151a-5p、VEGF、年龄)与因变量(log 累积风险函数)之间的线性关系进行检验。采用 RCS 方法评估各连续变量的非线性趋势,结果显示,miR-504-3p、miR-151a-5p、VEGF

及年龄与 log 累积风险函数均呈线性关系(非线性检验 P 值分别为 0.324、0.416、0.278、0.562),满足 Cox 回归模型的比例风险及线性假设,故以原始连续变量形式纳入模型。采用 VIF 对纳入多

因素模型的变量进行共线性诊断。结果显示, miR-504-3p、miR-151a-5p、VEGF、临床分期、治疗年份的 VIF 值分别为 1.342、1.418、1.527、1.216、1.093, 均远低于经验阈值 10, 提示变量间不存在显著的多重共线性问题, 可同时纳入多因素 Cox 回归模型进行分析。

以患者 5 年生存状态为因变量(死亡=1, 存活=0), 将单因素 Cox 回归分析中 $P < 0.05$ 的变量(年龄、吸烟史、临床分期、靶向治疗、miR-504-3p、miR-151a-5p、VEGF)纳入多因素 Cox 回归模型。为控制治疗策略随时间变化可能引入的偏倚,

将“治疗年份”(2016 年=1, 2017 年=2, 2018 年=3, 2019 年=4)作为校正变量纳入模型。采用基于似然比检验的逐步回归法进行变量筛选, 纳入标准为 $P < 0.05$, 剔除标准为 $P > 0.10$ 。Cox 回归分析显示, miR-151a-5p、VEGF、临床分期Ⅳ期是晚期 NSCLC 患者 5 年生存的独立危险因素($P < 0.05$), miR-504-3p 是保护因素($P < 0.05$), 治疗年份在多因素分析中差异无统计学意义($P > 0.05$), 未进入最终预测模型, 年龄、吸烟史、靶向治疗在多因素分析中差异无统计学意义(均 $P > 0.05$), 见表 4。

表 4 影响晚期 NSCLC 患者 5 年生存的多因素 Cox 回归分析(预测模型变量筛选)

Tab. 4 Multivariable Cox regression analysis of factors associated with 5-year survival in patients with advanced NSCLC (variable selection for predictive model)

变量	赋值	β	<i>S.E.</i>	Wald	<i>HR</i>	95% <i>CI</i>	<i>P</i>
年龄	连续变量	0.215	0.187	1.324	1.240	0.859 ~ 1.789	0.250
吸烟史	无=0, 有=1	0.486	0.232	4.388	1.625	0.753 ~ 2.532	0.145
临床分期	IIIB ~ IIIC 期=0, IV 期=1	0.682	0.301	5.134	1.978	1.097 ~ 3.566	0.024*
靶向治疗	有=0, 无=1	0.451	0.256	3.104	1.570	0.951 ~ 2.592	0.078
miR-504-3p	连续变量	-0.394	0.185	4.534	0.674	0.469 ~ 0.969	0.033*
miR-151a-5p	连续变量	0.527	0.238	4.904	1.694	1.063 ~ 2.701	0.027*
VEGF	连续变量(每 100 pg/mL)	0.718	0.322	4.970	2.050	1.090 ~ 3.857	0.026*
治疗年份	2016 年=1, 2017 年=2, 2018 年=3, 2019 年=4	0.096	0.112	0.735	1.101	0.884 ~ 1.371	0.391

注: * $P < 0.05$; 多因素 Cox 回归分析中, miR-504-3p、miR-151a-5p 以原始检测值纳入模型, VEGF 因检测值范围较大(200 ~ 800 pg/mL), 为便于临床解释, 以每 100 pg/mL 为单位纳入模型; Wald 值计算公式为 $(\beta/SE)^2$, 近似服从自由度为 1 的卡方分布。表中所有 P 值均基于该分布计算得出, 经复核无误; Cox 比例风险模型采用部分似然估计, 未估计常数项(截距), 基线风险函数由 Breslow 方法估计。

基于多因素 Cox 回归分析筛选出的 4 个影响因素(miR-504-3p、miR-151a-5p、VEGF、临床分期Ⅳ期), 构建多指标预测模型。模型预测公式为: $PI = 0.682 \times \text{临床分期(IIIB-IIIC 期=0, IV 期=1)} - 0.394 \times \text{miR-504-3p} + 0.527 \times \text{miR-151a-5p} + 0.718 \times (\text{VEGF}/100)$ 。该 PI 为线性组合形式, 用于量化每位患者的综合风险水平。后续风险分层分析基于该 PI 值进行。

2.4 模型预测效能评估

2.4.1 模型对死亡风险的区分度评估 时间依赖 ROC 曲线分析显示, 模型预测死亡风险的时间依赖 AUC 随随访时间延长呈上升趋势。1 年、2 年、3 年、4 年、5 年的 AUC(95%CI) 分别为 0.703(0.637 ~ 0.769)、0.724(0.661 ~ 0.787)、0.752(0.693 ~ 0.811)、0.826(0.775 ~ 0.877)、0.859(0.814 ~ 0.904), 提示模型对长期死亡风险的预

测效能优于短期, 见图 1。

2.4.2 模型对转移风险的区分度评估 模型预测转移风险的时间依赖 AUC 亦随随访时间延长而提高。1 年、2 年、3 年、4 年、5 年的 AUC(95%CI) 分别为 0.694(0.624 ~ 0.764)、0.718(0.652 ~ 0.784)、0.767(0.708 ~ 0.826)、0.835(0.786 ~ 0.884)、0.847(0.802 ~ 0.892), 模型对长期转移风险的预测效能优于短期, 见图 2。

2.5 模型的准确性评估

2.5.1 模型对死亡风险的校准度评估 1 年和 2 年期的校准曲线与 45° 对角线偏离较大, 校准不佳; 而 3、4、5 年期的校准曲线与 45° 对角线贴合良好, 模型预测 3、4、5 年死亡风险的准确性较高, 见图 3、表 4。

2.5.2 模型对转移风险的校准度评估 1 年和 2 年期的校准曲线与 45° 对角线偏离较大, 校准不

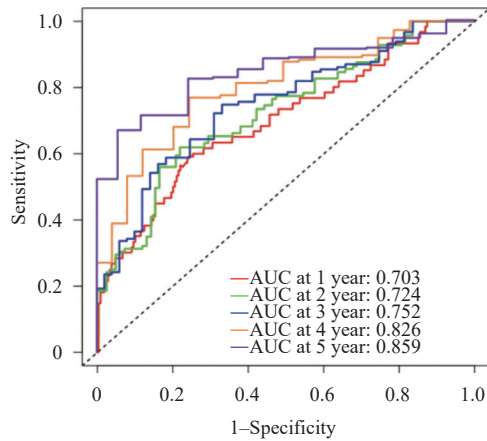


图 1 模型预测死亡风险的时间依赖的 ROC 曲线

Fig. 1 Time-dependent ROC curves for the model in predicting mortality risk

佳; 而 3、4、5 年期的校准曲线与 45° 对角线贴合良好, 模型预测 3、4、5 年转移风险的准确性较高, 见图 4、表 5。

2.5.3 模型临床实用性评估 DCA 结果显示, 在预测 3 年、4 年和 5 年死亡风险时, PI 模型在广泛的阈值概率范围内均能提供高于“全部干预”和“全部不干预”策略的净获益。在预测 3 年、4 年和 5 年转移风险时, 该模型在大部分临床合理的阈值概率范围内也展现出显著的临床净获益, 见图 5、图 6。

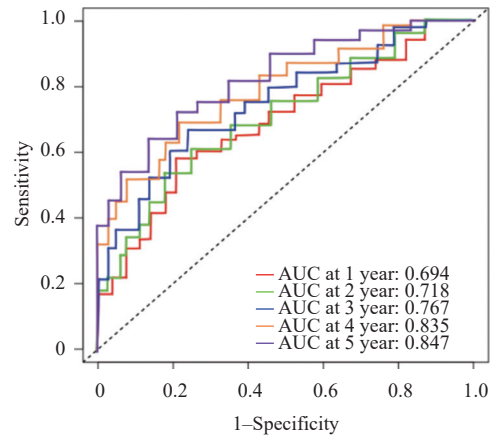


图 2 模型预测转移风险的时间依赖的 ROC 曲线

Fig. 2 Time-dependent ROC curves for the model in predicting metastasis risk

2.6 不同风险组患者生存及转移情况比较

根据 PI 中位数将 264 例患者分为高风险组 ($n = 175$) 和低风险组 ($n = 89$)。Kaplan-Meier 生存分析显示, 低风险组患者 5 年生存率显著高于高风险组 [47(52.81%) vs 11(6.29%), Log-rank $\chi^2 = 74.486$, $P < 0.001$]; 经 Fine-Gray 竞争风险模型校正死亡竞争风险后, 低风险组患者 5 年无远处转移率显著高于高风险组 [51(57.30%) vs 23(13.14%), Gray $\chi^2 = 57.032$, $P < 0.001$], 见图 7。

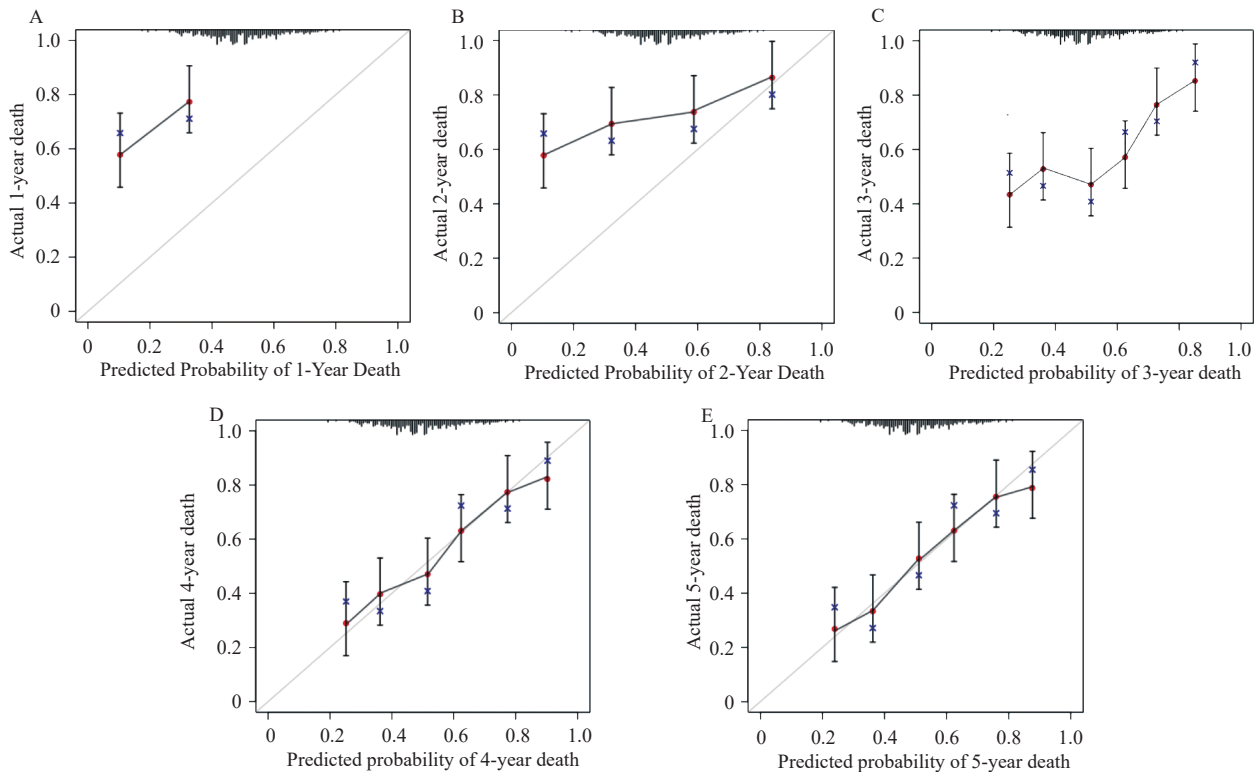


图 3 模型预测死亡风险的校准曲线

Fig. 3 Calibration curves for the model in predicting mortality risk

A: 1 年; B: 2 年; C: 3 年; D: 4 年; E: 5 年。

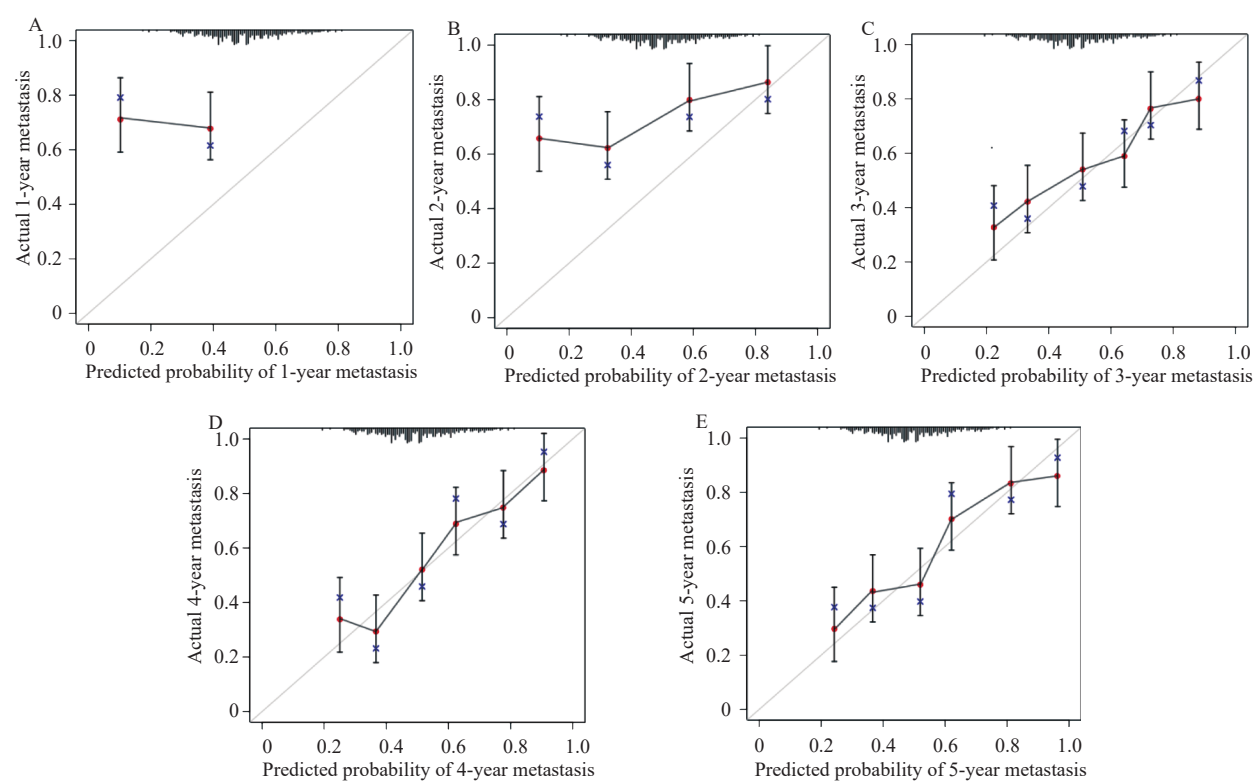


图 4 模型预测转移风险的校准曲线

Fig. 4 Calibration curves for the model in predicting metastasis risk

A: 1 年; B: 2 年; C: 3 年; D: 4 年; E: 5 年。

表 5 模型校准度的量化评估指标(Bootstrap 内部验证, 500 次重抽样)

时间点	死亡风险		转移风险	
	校准截距(95%CI)	校准斜率(95%CI)	校准截距(95%CI)	校准斜率(95%CI)
1年	-0.213(-0.412 ~ -0.014)	0.762(0.635 ~ 0.889)	-0.185(-0.373 ~ 0.003)	0.784(0.658 ~ 0.910)
2年	-0.156(-0.298 ~ -0.013)	0.819(0.703 ~ 0.935)	-0.133(-0.269 ~ 0.003)	0.837(0.723 ~ 0.951)
3年	-0.042(-0.176 ~ 0.092)	0.945(0.842 ~ 1.048)	-0.056(-0.188 ~ 0.076)	0.932(0.834 ~ 1.030)
4年	0.038(-0.074 ~ 0.150)	1.012(0.921 ~ 1.103)	0.027(-0.083 ~ 0.137)	1.008(0.918 ~ 1.098)
5年	0.021(-0.085 ~ 0.127)	0.985(0.894 ~ 1.076)	0.016(-0.089 ~ 0.121)	0.992(0.902 ~ 1.082)

3 讨论

精准预测晚期 NSCLC 患者的长期生存结局及转移风险,是实现个体化诊疗、改善预后的关键前提^[15-16]。尽管 TNM 分期是临床评估肺癌进展的经典标准,但受肿瘤异质性影响,其对个体预后的预测效能有限。所以,探寻新型预后生物标志物对提升晚期 NSCLC 患者生存率意义重大。

既往研究显示,miR-504-3p 可抑制肿瘤细胞侵袭迁移,其高表达者生存期更久^[17]。本研究发现,存活组患者血清 miR-504-3p 表达水平显著高于死亡组,miR-504-3p 是晚期 NSCLC 患者长

期生存的独立保护因素,推测其或靶向调控下游促癌基因,抑制 NSCLC 细胞增殖、侵袭与转移,进而改善患者预后。miR-151a-5p 为促癌 miRNA,与多种肿瘤的不良预后相关^[18]。有研究显示,miR-151a-5p 能靶向 PTEN 基因激活 PI3K/Akt 通路,促进肿瘤细胞增殖转移,其高表达者 5 年生存率明显下降^[19]。本研究表明,死亡组患者血清该指标水平显著升高,是晚期 NSCLC 患者 5 年生存的独立危险因素,提示血清中升高的 miR-151a-5p 可能反映了体内肿瘤更具侵袭性的生物学表型。VEGF 作为调控肿瘤血管生成的核心因子,与肺癌预后密切相关^[20]。VEGF 能促进新生血管生成,给肿瘤细胞输送丰富营养和氧气,还

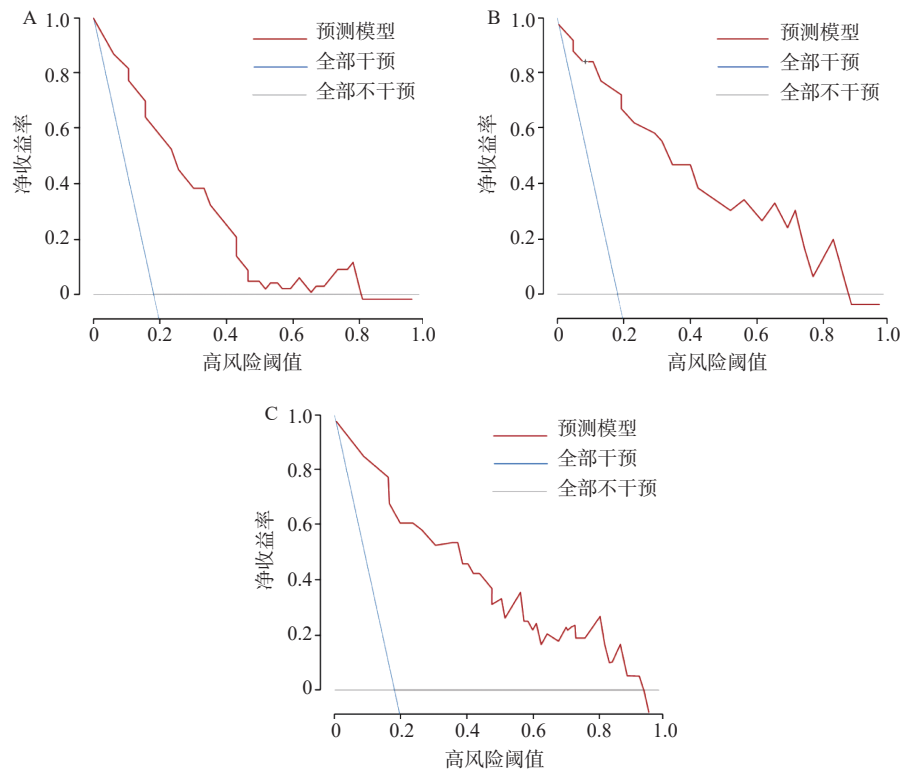


图 5 模型预测不同时间点死亡风险的决策曲线分析

Fig. 5 Decision curve analysis for the model in predicting mortality risk at different time points

A: 3 年; B: 4 年; C: 5 年。

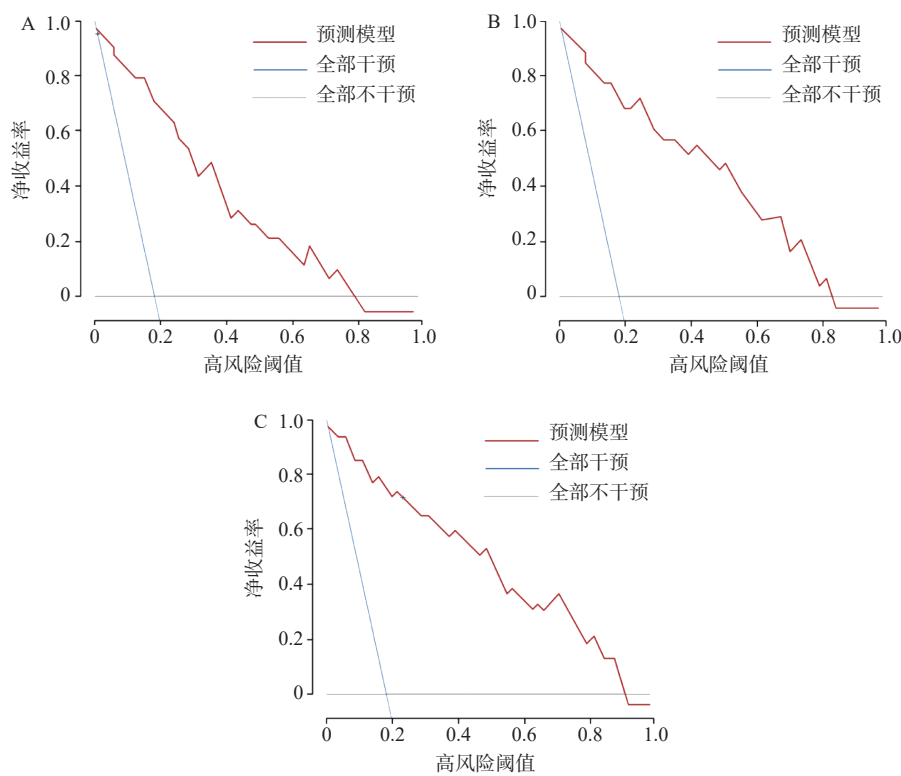


图 6 模型预测不同时间点转移风险的决策曲线分析

Fig. 6 Decision curve analysis for the model in predicting metastasis risk at different time points

A: 3 年; B: 4 年; C: 5 年。

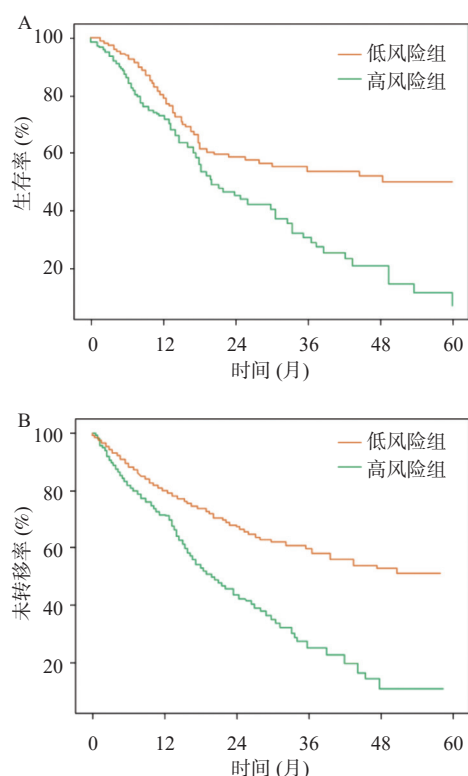


图 7 不同风险组患者生存及转移情况比较

Fig. 7 Comparison of survival and metastasis between different risk groups

A: 死亡风险; B: 转移风险, 经 Fine-Gray 竞争风险模型校正。

能加快肿瘤细胞远处转移, 其高表达常与肺癌晚期分期、不良预后有关^[21]。本研究结果和过往研究相符, 显示 VEGF 促进晚期 NSCLC 进展, 其高表达是此类患者长期生存的独立危险因素。

既往研究构建的多指标预测模型多集中于短期预后^[22], 针对 5 年长期生存的预测模型较少, 且预测效能有限。本研究整合 miR-504-3p、miR-151a-5p、VEG 与临床分期构建 PI 模型。随着时间推移, PI 模型对 5 年生存结局及转移风险的预测效能逐渐提高, 区分能力不断增强; 模型在 1 年、2 年期校准效果不理想, 但 3、4、5 年期校准曲线与 45° 对角线贴合度较高, 提示该模型在预测晚期 NSCLC 患者 3 年及以上的长期生存与转移风险时具有良好的准确性。

本研究所构建的多指标预测模型在短期与长期预测准确性上存在差异。晚期 NSCLC 患者早期死亡原因的复杂性可能削弱了模型在短期的预测效能。研究表明, 在晚期 NSCLC 确诊后的 1 ~ 2 年内, 患者的死亡原因除肿瘤进展外, 还常涉及初始治疗相关毒性、感染、心血管事件等并发症, 以及急性治疗反应不佳等因素^[23]。这些短期事件具有较强的随机性和个体差异性, 与肿瘤固

有的生物学行为关联度较低, 因此难以被基于基线分子标志物的模型准确预测。研究发现, 根治性治疗组患者的死亡风险率在 6 至 24 个月期间达到峰值, 且死因构成随时间呈现动态变化, 早期死亡中治疗相关因素占比较高^[24]。此外, 本模型所整合的标志物更倾向于反映肿瘤的长期侵袭潜能, 而非短期治疗反应。miRNA 作为表观遗传调控分子, 通过影响肿瘤细胞增殖、上皮间质转化及血管生成等过程, 塑造肿瘤的内在生物学特性, 其对预后的影响往往在较长随访期内逐渐显现。研究发现, 血浆外泌体 miRNA 对 NSCLC 患者总生存期的预测价值随时间推移而增强, 时间依赖 AUC 在长期随访中优于短期^[21]。VEGF 作为促血管生成因子, 其水平反映的是肿瘤微环境的血管生成状态, 这一状态对肿瘤转移潜能的塑造具有持续性, 但其对短期预后的影响可能被更直接的治疗因素所掩盖。此外, 其构建的 NLR-LMR 动态评分系统对免疫治疗晚期 NSCLC 患者的预测效能随时间推移而提升。这反映了炎症标志物对长期免疫监视重建的预测价值优于短期治疗应答。本研究结果与之类似, 提示基于基线分子标志物的预后模型可能在预测长期结局时更具优势, 而短期预测需纳入动态变化的治疗反应指标。此外, 本研究按 PI 中位数划分患者为高、低风险组, 低风险组 5 年生存率和未转移率明显高于高风险组, 凸显了 PI 模型的临床应用价值。进一步通过 DCA 评估模型的临床净收益, 结果显示, 在 3、4、5 年时间点, 模型在阈值合理范围内均表现出优于“全部干预”和“全部不干预”策略的净获益, 提示模型在指导临床决策方面具有良好的实用价值。临床医生可通过该模型快速识别出高复发转移风险的晚期 NSCLC 患者, 对其采取更积极的个体化治疗策略, 从而改善患者的长期生存获益; 而对于低风险患者, 则可适当简化治疗方案, 减少过度治疗带来的不良反应与医疗负担。

本研究存在以下局限性: 首先, 研究设计为单中心回顾性研究, 所有 264 例患者均来自同一医疗中心, 尽管样本量满足统计要求, 但单中心数据可能引入选择性偏倚, 限制了模型的通用性和外推性。因此, 本研究所构建的多指标预测模型需在独立的多中心前瞻性队列中进行外部验证, 以确认其稳定性和临床适用性; 其次, 未深入探讨 miR-504-3p、miR-151a-5p 与 VEGF 之间的相互作用机制, 后续可结合体外实验及多组学分析进一步阐明其分子机制。

综上所述, miR-504-3p、miR-151a-5p、VEGF 与晚期 NSCLC 患者长期生存及转移风险密切相关。整合三者及临床分期的多指标预测模型, 对晚期 NSCLC 患者 3 年及以上的长期生存与转移风险具有良好的预测准确性, 有助于临床识别高复发转移风险的患者。

[参考文献]

- [1] Haratake N, Ozawa H, Morimoto Y, et al. MUC1-C is a common driver of acquired osimertinib resistance in NSCLC[J]. *J Thorac Oncol*, 2024, 19(3): 434-450.
- [2] Molero A, Hernandez S, Alonso M, et al. Assessment of PD-L1 expression and tumour infiltrating lymphocytes in early-stage non-small cell lung carcinoma with artificial intelligence algorithms[J]. *J Clin Pathol*, 2025, 78(7): 456-464.
- [3] Meyer M L, Fitzgerald B G, Paz-Ares L, et al. New promises and challenges in the treatment of advanced non-small-cell lung cancer[J]. *Lancet*, 2024, 404(10454): 803-822.
- [4] Wu F, Fan J, He Y, et al. Single-cell profiling of tumor heterogeneity and the microenvironment in advanced non-small cell lung cancer[J]. *Nat Commun*, 2021, 12: 2540.
- [5] Alessi J V, Elkrief A, Ricciuti B, et al. Clinicopathologic and genomic factors impacting efficacy of first-line chemoimmunotherapy in advanced NSCLC[J]. *J Thorac Oncol*, 2023, 18(6): 731-743.
- [6] Charkiewicz R, Sulewska A, Charkiewicz A, et al. miRNA-seq tissue diagnostic signature: A novel model for NSCLC subtyping[J]. *Int J Mol Sci*, 2023, 24(17): 13318.
- [7] Abdipourbozorgbaghi M, Vancura A, Radpour R, et al. Circulating miRNA panels as a novel non-invasive diagnostic, prognostic, and potential predictive biomarkers in non-small cell lung cancer (NSCLC) [J]. *Br J Cancer*, 2024, 131(8): 1350-1362.
- [8] Zu L, He J, Zhou N, et al. Identification of multiple organ metastasis-associated hub mRNA/miRNA signatures in non-small cell lung cancer[J]. *Cell Death Dis*, 2023, 14: 798.
- [9] Wang M, Ding X, Fang X, et al. Circ6834 suppresses non-small cell lung cancer progression by destabilizing AN-HAK and regulating miR-873-5p/TXNIP axis[J]. *Mol Cancer*, 2024, 23(1): 128.
- [10] Liu P I, Jiang Y J, Chang A C, et al. ANGPTL2 promotes VEGF-A synthesis in human lung cancer and facilitates lymphangiogenesis[J]. *Aging*, 2023, 15(5): 1652-1667.
- [11] 余会敏, 辛鑫, 梁志伟. KIAA1199 基因和 Linc00673 基因对非小细胞肺癌化疗疗效的预测价值分析 [J]. *昆明医科大学学报*, 2025, 46(1): 111-116.
- [12] 张波, 钟润波, 钟华. 《中华医学会肺癌临床诊疗指南 (2025 版)》更新要点解读 [J]. *中华肿瘤杂志*, 2025, 47(10): 981-986.
- [13] 王洁, 赫捷, 国家肿瘤质控中心肺癌质控专家委员会, 等. 原发性肺癌化疗规范化应用中国指南 (2024 版) [J]. *中国肿瘤临床与康复*, 2024, 31(9): 525-548.
- [14] 热哈提·阿曼交尔, 顾国民, 刘春玲. CA19-9 水平在非小细胞肺癌预后评估中的价值 [J]. *热带医学杂志*, 2024, 24(12): 1670-1675.
- [15] 洪超, 向旭东, 李盈甫, 等. ERK1/2 信号通路基因 3'UTR 多态性与非小细胞肺癌的相关性 [J]. *昆明医科大学学报*, 2024, 45(3): 7-17.
- [16] Liu Z, Zhang W, Zhang B, et al. miR-504-3p has tumor-suppressing activity and decreases IFITM1 expression in non-small cell lung cancer cells[J]. *Genet Test Mol Biomark*, 2022, 26(7-8): 351-359.
- [17] Ebrahimi A, Derakhshan S M, Ghavi D, et al. The role of mir-151a-5p in tumorigenesis: A systematic review[J]. *Pathol Res Pract*, 2023, 249: 154576.
- [18] Hu H, Jiang L, Kang X, et al. Extracellular vesicles derived from lung cancer cells promote the progression of lung cancer by delivering miR-151a-5p[J]. *Exp Cell Res*, 2023, 425(1): 113526.
- [19] Zhao Y, Guo S, Deng J, et al. VEGF/VEGFR-targeted therapy and immunotherapy in non-small cell lung cancer: Targeting the tumor microenvironment[J]. *Int J Biol Sci*, 2022, 18(9): 3845-3858.
- [20] Zhang Y M, Miao Z M, Chen Y P, et al. Ononin promotes radiosensitivity in lung cancer by inhibiting HIF-1 α /VEGF pathway[J]. *Phytomedicine*, 2024, 125: 155290.
- [21] Guo Y, Li L, Zheng K, et al. Development and validation of a survival prediction model for patients with advanced non-small cell lung cancer based on LASSO regression[J]. *Front Immunol*, 2024, 15: 1431150.
- [22] Zaborowska-Szmit M, Krzakowski M, Kowalski D M, et al. Cardiovascular complications of systemic therapy in non-small-cell lung cancer[J]. *J Clin Med*, 2020, 9(5): 1268.
- [23] Robinson A G, Young K, Balchin K, et al. Causes of death and subsequent treatment after initial radical or palliative therapy of stage III non-small-cell lung cancer[J]. *Curr Oncol*, 2015, 22(5): 333-340.
- [24] Liu Q, Yu Z, Yuan S, et al. Circulating exosomal microRNAs as prognostic biomarkers for non-small-cell lung cancer[J]. *Oncotarget*, 2017, 8(8): 13048-13058.