

网络出版时间: 2020-12-25 11:23 网络出版地址: <https://kns.cnki.net/kcms/detail/34.1065.R.20201224.1107.024.html>

◇ 临床医学研究 ◇

# 结肠癌列线图预后风险模型构建: 基于 SEER 数据库的回顾性研究

黄 凯, 张长乐, 吴文涌, 刘省存, 方昌义, 余昌俊

**摘要** **目的** 分析结肠癌患者预后的重要风险因素, 构建列线图预测结肠癌患者的 1、3 及 5 年生存率, 并进行临床获益分析。**方法** 从 SEER 数据库中收集 2010—2015 年结肠癌患者年龄、病理类型等临床资料以及相关随访数据共 53 753 例。将年龄、病理分期、AJCC 第 7 版 T 分期、N 分期、M 分期、阳性淋巴结比值(LNR)、组织分级、性别和化疗等 9 个因素纳入结肠癌的预后分析中, 使用 Cox 比例风险模型分别进行单因素、多因素生存分析, 确定预后风险因素。使用逐步回归方法进一步筛选与预后关联最重要的临床因素, 用于构建列线图预后模型。校正曲线对列线图进行内部及外部验证, 检查其预测精度。决策曲线分析评估列线图模型对患者的临床获益预测的准确性。**结果** 年龄、病理分期、AJCC 第 7 版 T 分期、N 分期、M 分期、LNR、组织分级、性别和化疗等 9 个因素均为结肠癌患者的独立预后风险因素( $P < 0.01$ )。逐步回归分析筛选出年龄、M 分期、T 分期和 LNR 4 个因素与结肠癌的预后关联最大, 将 4 个因素纳入并构建了列线图。训练集和校验集的列线图一致性指数(C 指数)分别为 0.754(95% CI: 0.749 ~ 0.759) 和 0.748(95% CI: 0.741 ~ 0.755), ROC 曲线分析显示训练集和校验集样本的 1、3、5 年 AUC 均达到了 0.7 以上, 显示出 4 个因素列线图具有较好的性能。**结论** 基于年龄、M 分期、T 分期和 LNR 4 个因素构建的结肠癌预后总生存列线图具有良好的预测精度, 且好于传统的 TNM 分期系统。预后总生存列线图的建立有助于临床肿瘤医师对患者预后作出较为准确的预后评估, 可为结肠癌患者实施个体化诊疗提供益处。

**关键词** 结肠癌; 特征选择; 列线图; 预后; 生存分析**中图分类号** R 735.3 + 5**文献标志码** A **文章编号** 1000-1492(2021)02-0299-07

doi: 10.19405/j.cnki.issn1000-1492.2021.02.024

结肠癌是世界最常见的肿瘤之一, 每年导致约 90 万人死亡<sup>[1]</sup>。在中国, 随着生活方式的改变和营

养水平的提升, 结肠癌的发生率也呈现逐年上升的趋势<sup>[2-3]</sup>。基于长期的随访数据显示, 中国结肠癌患者的 5 年总生存率要低于美国和欧洲<sup>[4-5]</sup>, 这表明建立结肠癌准确而可靠的预后风险模型对改善结肠癌预后具有重要意义。特征选择是构建预后模型重要的一步, 选取有效的特征即可以去除大量冗余信息和不相关特征的干扰, 又能够降低模型的复杂度, 便于研究者更好的应用于临床中<sup>[6]</sup>。常见的特征选择方法包括 LASSO<sup>[7]</sup>、随机森林<sup>[8-9]</sup>和逐步回归<sup>[10]</sup>等。尽管 Marisa et al<sup>[11]</sup>建立了基于转录组的结肠癌分类法, 从而改善了当前的预后分层, 然而 AJCC 的病理分期仍然未能准确预测结肠癌患者的预后状态。该研究应用逐步回归方法筛选最优的临床风险因素, 并利用这些临床因素构建了结肠癌预后多因素列线图, 为结肠癌患者实施个体化诊疗提供参考。

## 1 材料与方法

**1.1 研究对象与处理** 通过 SEER\* Stat 8.3.6 软件, 根据肿瘤原发部位标记: C18.2(升结肠)、C18.3(结肠肝区)、C18.4(横结肠)、C18.5(结肠脾区)、C18.6(降结肠)、C18.7(乙状结肠), 搜集 SEER 数据库中 2010—2015 年的结肠癌患者的相关资料。每个患者的资料包括: 年龄、种族、性别、组织分级、AJCC 第 7 版 T 分期、N 分期、M 分期、手术方式、肿瘤位置、阳性淋巴结数目、手术活检淋巴结数目、化疗和生存信息。具体纳入标准如下: ① 病理确诊的原发性结肠癌; ② 采用 AJCC 第 7 版分期系统对病例进行分期; ③ 随访信息完整。排除标准如下: ① 多源性肿瘤; ② 尸检获取的和只有死亡报告上的病例; ③ 肿瘤分化程度及分期不完整; ④ 30 d 内死亡的病例。根据以上标准, 共 53 753 例患者纳入本研究。本研究纳入的生存时间为总生存时间(overall survival, OS), 阳性淋巴结比值(lymph node ratio, LNR)根据阳性淋巴结数目除以手术活检淋巴

2020-09-20 接收

基金项目: 国家自然科学基金(编号: 81902451)

作者单位: 安徽医科大学第一附属医院胃肠外科, 合肥 230022

作者简介: 黄 凯, 男, 硕士研究生;

余昌俊, 男, 主任医师, 博士生导师, 责任作者, E-mail:

yuchangjun321@163.com

结数目计算得到。本研究基于 SEER 数据库发布的公开数据,所有患者已去除个人标识,因此无需伦理委员会的批准,也无需患者知情同意。

**1.2 基于逐步回归的特征选择** 逐步回归是多元线性回归分析中一种检查每个自变量的统计显著性的方法,通过在模型中逐步引入或者去除一个变量实现对变量的选择<sup>[12]</sup>。本研究中我们使用 R package ‘MASS’<sup>[13]</sup> 进行逐步回归分析,选择 Bidirectional elimination 方法。简单来说,初始状态下逐个加入临床因素,评估新加入的因素是否使得模型发生显著性变化( $F$  检验),若原有的某个因素由于新加入的因素而不再显著时,则去除该因素,重复此过程直到没有显著的临床因素可以加入回归模型,也没有不显著的临床因素从回归模型中被去除,最终得到一个最优的变量集合。逐步回归过程中计算每一个临床因素的信息准则值(akaike information criterion, AIC), AIC 越大表明该因素对模型越重要。

**1.3 统计学处理** 应用 R version 3.6.1<sup>[14]</sup> 软件对数据进行分析。设定随机数种子 = 54 321,对全部样本按照 2 : 1 比例进行随机分组,分别得到训练集和校验集。类别型变量的分组比较使用卡方检验,数值型变量的二分组和多分组差异显著性分析分别使用秩和检验和 ANOVA 进行,类别型变量的二分组比较使用卡方检验。预后生存分析使用 Cox 比例风险模型,Log-Rank 检验比较生存率的差异。使用 R package ‘rms’ 进行 nomogram 分析,并画出预后预测列线图。模型的净获益分析使用 R package ‘dca’,分别选取 1、3、5 年的时间点进行比较。

## 2 结果

**2.1 患者基本临床病理资料** 本研究共包含了 53 753 例结肠癌个体样本,在随访时间内生存状态为存活和死亡的样本分别有 39 586 和 14 167 例。分析所用的患者特征以老年(71.27%)、白色人种(78.04%)、病理分期 I ~ III(88.69%) 为主。全部样本集按照 2 : 1 随机分成训练集和校验集(随机数种子 = 54 321),两组样本的患者生存状态和 OS 差异无统计学意义,见表 1。

**2.2 结肠癌患者危险因素分析** 为了确定临床因素与结肠癌的预后关系,我们分别构建了单因素和和多因素预后风险模型。我们看到年龄、病理分期、AJCC 第 7 版 T 分期、N 分期、M 分期、LNR、组织分

表 1 SEER 数据库结肠癌患者训练集和

校验集样本的临床病理特征

	全部数据集 ( <i>n</i> = 53 753)	训练集 ( <i>n</i> = 35 835)	校验集 ( <i>n</i> = 17 918)	<i>P</i> 值
生存状态				0.419
存活	39 586	26 351	13 235	
死亡	14 167	9 484	4 683	
年龄(岁)				0.592
<60	15 445	10 270	5 175	
≥60	38 308	25 565	12 743	
群体				0.088
黑种人	6 668	4 525	2 143	
其他肤色	5 139	3 349	1 790	
白种人	41 946	27 961	13 985	
性别				0.448
女性	26 235	17 448	8 787	
男性	27 518	18 387	9 131	
肠道来源				0.824
升结肠	15 838	10 591	5 247	
结肠肝区	3 493	2 296	1 197	
横结肠	7 697	5 116	2 581	
结肠脾区	2 469	1 642	827	
降结肠	4 751	3 156	1 595	
乙状结肠	19 505	13 034	6 471	
肿瘤分级				0.149
G1	4 401	2 887	1 514	
G2	39 838	26 544	13 294	
G3	7 855	5 307	2 548	
G4	1 659	1 097	562	
AJCC 分期				0.684
I 期	13 368	8 863	4 505	
II 期	17 584	11 761	5 823	
III 期	16 721	11 138	5 583	
IV 期	6 080	4 073	2 007	
T 分期				0.390
T1	7 915	5 258	2 657	
T2	7 982	5 265	2 717	
T3	29 830	19 968	9 862	
T4	8 026	5 344	2 682	
N 分期				0.392
N0	32 076	21 382	10 694	
N1	13 988	9 281	4 707	
N2	7 689	5 172	2 517	
M 分期				0.579
M0	47 673	31 762	15 911	
M1	6 080	4 073	2 007	

级、性别、化疗与否均与患者总体生存率存在相关性( $P < 0.01$ ),并且 LNR 具有最大的风险比(hazard ratio, HR),表明 LNR 对结肠癌预后的影响较其他因素更大(表 2)。多因素生存分析也表明这些因素与预后存在关联( $P < 0.01$ )。相对于其他因素而

言,化疗的  $HR < 1$ ,说明接受化疗有利于患者的预后,这也与真实情况是一致的(表2)。

**2.3 结肠癌预后危险因素的特征选择** 为了进一步对结肠癌预后风险因素进行筛选,利用训练集数据我们使用逐步回归方法评估每一个临床因素与预后的重要性,根据 AIC 对临床因素进行排序,发现对结肠癌的预后影响排名前 4 的因素包括年龄、M 分期、T 分期和 LNR(图 1A),提示这 4 个临床因素与结肠癌预后关联最紧密。多因素 Cox 风险比例模型计算的年龄、M 分期、T 分期和 LNR 4 个因素预后模型在训练集和校验集上均具有最大的一致性指数 (concordance index, C 指数),分别为 0.754 和 0.748 (图 1B,表 3),表明年龄、M 分期、T 分期和 LNR 4 个临床因素组成的多因素模型对结肠癌预后风险评估具有较好的表现。此外,ROC 曲线分析发现 4 个因素预后风险模型 1、3、5 年 AUC 均达到了 0.7 以上(图 2),进一步确认模型的可靠性。

## 2.4 构建基于 nomogram 的肠癌预后模型 逐步

回归分析筛选了 4 个与预后关联最紧密的临床因素,我们进一步构建了基于年龄/pT/pM/LNR 4 个因素的结肠癌患者 1、3 和 5 年总生存率的列线图(图 3)。通过将每个变量对应的点向“Points”轴投射获得的分数相加,所得总分数可对应相应的预测结果。4 因素的列线图 C 指数为 0.748, C 指数越接近于 1 表示列线图可信度越高。通过绘制 4 个因素列线图模型的预测值与实际值校准图,显示列线图预后模型的 1、3 以及 5 年 OS 与真实结果具有较高的一致性,平均错误率为 1.63%(图 4)。

**2.5 列线图的临床获益分析** 在评估了该模型的预测准确性之后,我们进一步评估了 4 个临床因素的加入是否可以使结肠癌患者在临床实践中受益。列线图的决策曲线分析(decision curve analysis, DCA)评估了患者可以净获益,净获益率越大,预后风险模型的预测性能越好。结果表明基于 4 个临床因素的预后模型较 AJCC 第 7 版 TNM 系统的预测模型将使得患者有更多获益(图 5)。

表 2 结肠癌患者总体生存相关预后因素的单因素及多因素分析

临床因素	单因素		多因素	
	HR(95% CI)	P 值	HR(95% CI)	P 值
年龄	1.178 0(1.169 9 ~ 1.186 3)	<0.01	1.196 5(1.185 7 ~ 1.207 3)	<0.01
AJCC 分期	1.961 3(1.925 6 ~ 1.997 6)	<0.01	1.176 6(1.113 0 ~ 1.243 9)	<0.01
T 分期	1.868 7(1.827 4 ~ 1.911 0)	<0.01	1.417 2(1.368 5 ~ 1.467 6)	<0.01
N 分期	1.821 9(1.785 3 ~ 1.859 2)	<0.01	1.180 5(1.120 5 ~ 1.243 7)	<0.01
M 分期	4.655 2(4.485 5 ~ 4.831 4)	<0.01	3.176 3(2.920 5 ~ 3.454 5)	<0.01
LNR	10.587 8(9.937 7 ~ 11.280 5)	<0.01	3.133 5(2.737 3 ~ 3.587 0)	<0.01
肿瘤分级	1.572 1(1.532 2 ~ 1.613 0)	<0.01	1.205 5(1.165 7 ~ 1.246 8)	<0.01
性别				
男 vs 女	1.070 3(1.035 6 ~ 1.106 2)	<0.01	1.160 9(1.114 6 ~ 1.209 0)	<0.01
化疗				
否 vs 是	1.071 2(1.034 7 ~ 1.109 0)	<0.01	0.527 3(0.500 1 ~ 0.556 0)	<0.01

表 3 临床因素在训练集和校验集上一致性指数

临床因素	训练集			校验集		
	P 值	C 指数	95% CI	P 值	C 指数	95% CI
年龄	0.001	0.625	0.619 ~ 0.631	0.001	0.623	0.615 ~ 0.632
AJCC 分期	0.001	0.668	0.662 ~ 0.674	0.001	0.665	0.657 ~ 0.673
T 分期	0.001	0.635	0.629 ~ 0.640	0.001	0.633	0.626 ~ 0.641
N 分期	0.001	0.623	0.618 ~ 0.629	0.001	0.626	0.618 ~ 0.634
M 分期	0.001	0.610	0.606 ~ 0.615	0.001	0.606	0.599 ~ 0.612
LNR	0.001	0.631	0.625 ~ 0.637	0.001	0.632	0.623 ~ 0.640
年龄/M 分期	0.001	0.724	0.718 ~ 0.729	0.001	0.718	0.710 ~ 0.725
年龄/T 分期/M 分期	0.001	0.742	0.737 ~ 0.747	0.001	0.734	0.727 ~ 0.741
年龄/T 分期/M 分期/LNR	0.001	0.754	0.749 ~ 0.759	0.001	0.748	0.741 ~ 0.755
肿瘤分级	0.001	0.572	0.567 ~ 0.578	0.001	0.574	0.567 ~ 0.581
群体	0.001	0.512	0.508 ~ 0.516	0.001	0.518	0.512 ~ 0.524
性别	0.001	0.509	0.504 ~ 0.514	0.200	0.504	0.497 ~ 0.512
化疗	0.001	0.497	0.492 ~ 0.502	0.135	0.496	0.489 ~ 0.503

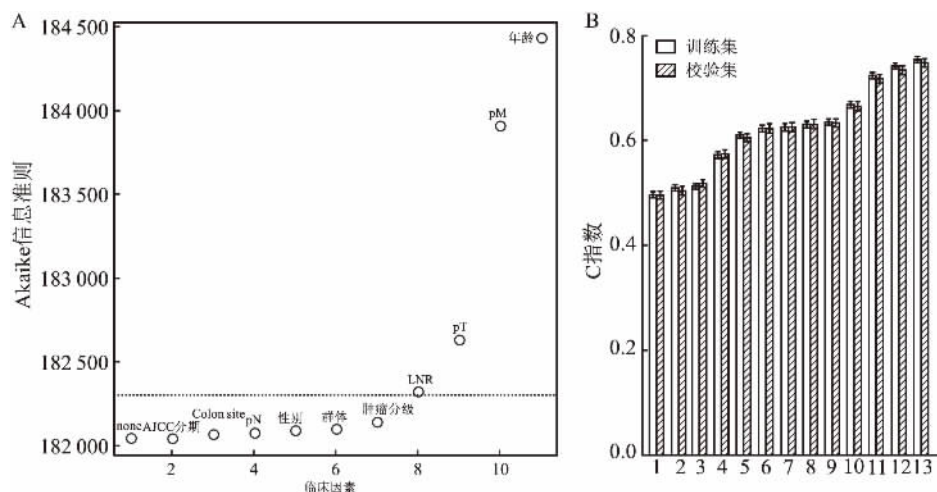


图1 基于逐步回归的结肠癌预后风险因素筛选

A: 临床因素的 AIC 分布; B: 不同临床因素的构建预后模型在训练集和校验集上的一致性指数; 1: 化疗; 2: 性别; 3: 群体; 4: 肿瘤分级; 5: pM; 6: pN; 7: 年龄; 8: LNR; 9: pT; 10: AJCC 分期; 11: 年龄/pM; 12: 年龄/pT/pM; 13: 年龄/pT/pM/LNR

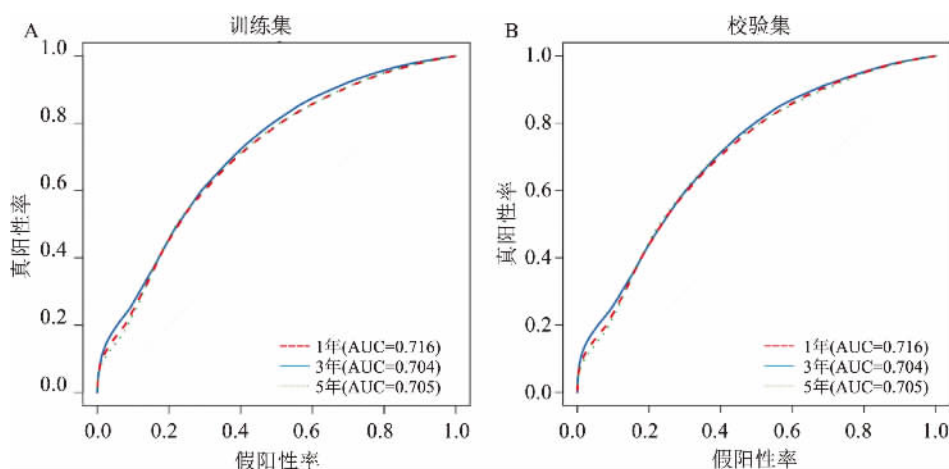


图2 年龄/pT/pM/LNR 4个因素预后模型1、3和5年 ROC 曲线

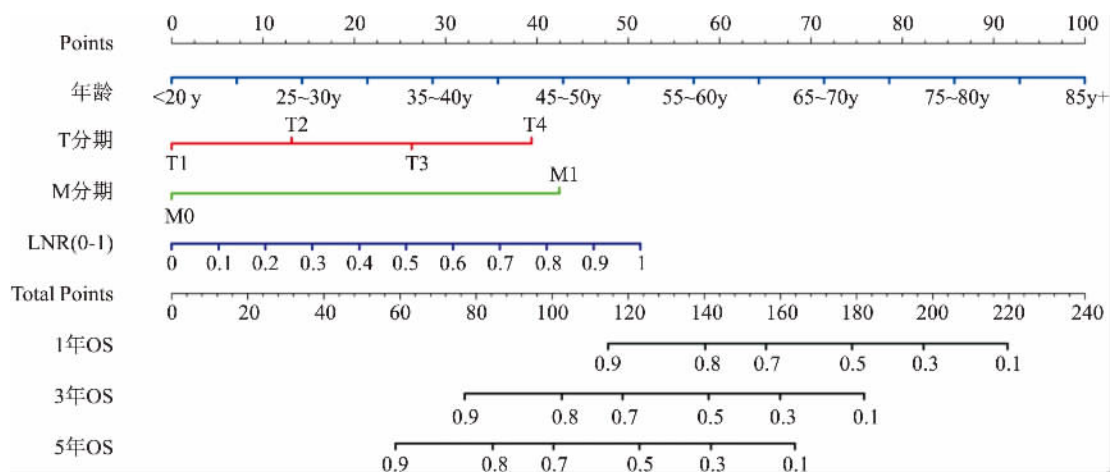


图3 全部数据集的结肠癌患者的1、3、5年 OS 预测的列线图

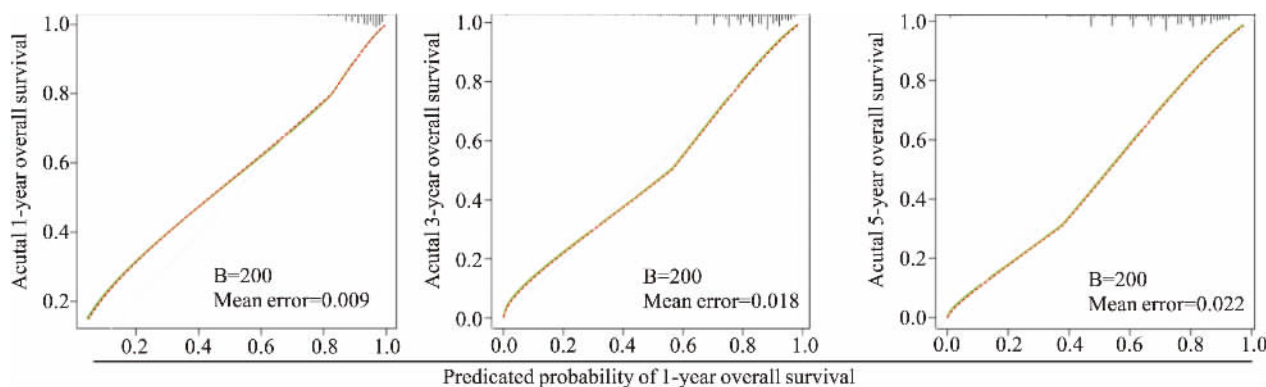


图4 1、3、5年OS概率的校准图

灰色实线为理想情况下的概率,红色虚线为校正后最优的概率,绿色实线为观察到的概率,bootstrap = 200

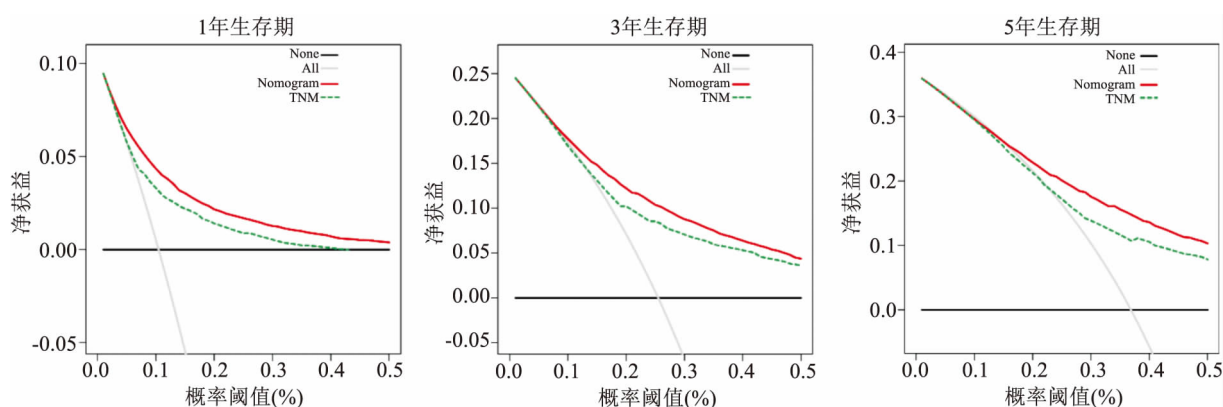


图5 结肠癌全部数据集列线图预后风险模型的决策曲线分析

横坐标为概率阈值,纵坐标为不同模型预测患者OS的净获益率;黑色实线和灰色实线用于作为参考,分别假设没有患者为高风险和所有患者均为高风险时净获益率;绿色虚线:TNM分期的预后风险模型评估的患者净获益率;红色实线:列线图预后风险模型评估的患者净获益率

### 3 讨论

目前AJCC-TNM分期为结肠癌患者生存预测和临床决策选择最广泛的系统,然而在精确区分不同阶段结肠癌患者生存差异方面仍然具有局限性,事实上一些研究已经表明列线图比传统的tumor-node-metastasis分期系统具有更好的预后风险预测的性能<sup>[15-19]</sup>,然而大量的临床因素被整合到预后风险模型中增加了预后判别的复杂性。本研究中我们首先对结肠癌预后风险因素进行了筛选,保留了较少的临床因素同时不损失预后风险模型的精确性,进一步利用这些因素构建了结肠癌的列线图,通过整合多种临床因素建立一个更准确而实用的预测模型。

在本研究,我们看到年龄是结肠癌非常重要的

预后风险因素(图1A, HR = 1.178),说明高龄的结肠癌患者预后风险较年轻患者更高,更应该注意预后监控和随访。另一方面也提示高龄患者的治疗,比如手术切除等创伤影响较大的方案可能需要更谨慎选择。此外,我们也观察到SEER数据库结肠癌患者的年龄以 $\geq 60$ 岁为主,而中国结肠癌患者则更加年轻,60岁以下的患者占大多数,这也预示着年龄因素可能对中国结肠癌患者的影响异于西方的结肠癌患者<sup>[20]</sup>。

淋巴结转移是结肠癌发生转移的主要形式之一,同时也是影响患者预后的重要因素<sup>[21]</sup>。虽然由AJCC定义的淋巴结分期已被广泛接受,但这种以阳性淋巴结数量为基础的分期方法明显受到切除的淋巴结总数影响,导致N分期存在一定偏差<sup>[22]</sup>。本研究中,我们使用LNR作为结肠癌预后因素,单因素

和多因素预后生存分析均表明 LNR 是结肠癌患者预后的重要风险因子 ( $HR > 1, P < 0.01$ )。在训练集和校验集中, LNR 的一致性指数比 N 分期的更大 ( $0.630\ 9/0.631\ 6$  vs  $0.623\ 1/0.625\ 7$ ), 也说明使用 LNR 来评估结肠癌的预后风险较 N 分期更具有优势。

单因素和多因素生存分析显示年龄、病理分期、AJCC 第 7 版 T 分期、N 分期、M 分期、LNR、组织分级、性别、化疗等因素均与结肠癌的总生存有关系, 逐步回归方法进一步对这些因素进行筛选, 最终确定年龄、T 分期、M 分期和 LNR 4 个因素作为最重要的预后因子, 由这 4 个因素构建的预后模型在训练集和校验集上一致性指数均最大, 显示出 4 个因素的模型具有较好的准确性。我们注意到逐步回归分析的结果没有纳入 N 分期而保留了 LNR, 这从另一方面说明结肠癌的预后风险模型中 LNR 似乎可以替代 N 分期。尽管逐步回归方法存在一些不足, 包括建模过程中基于同一数据所带来的偏好性、过拟合等, 但其优点在于可以通过对模型进行微调以选择最佳变量组合, 同时能够提供变量对模型的重要性信息, 此外应用独立的验证集数据可以解决该方法的一些不足。

### 参考文献

- [1] Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries [J]. CA: A Cancer Journal for Clinicians, 2018, 68(6): 394–424.
- [2] Fu J, Huang Y, Yang J, et al. Clinical characteristics and prognosis of young patients with colorectal cancer in eastern China [J]. World Journal of Gastroenterology, 2013, 19(44): 8078–84.
- [3] Li M, Gu J. Changing patterns of colorectal cancer in China over a period of 20 years [J]. World Journal of Gastroenterology, 2005, 11(30): 4685–8.
- [4] 万德森. 结直肠癌流行病学与预防 [J]. 中国中西医结合外科杂志, 2011, 17(1): 3–7.
- [5] 李明, 顾晋. 中国结直肠癌 20 年来发病模式的变化趋势 [J]. 中华胃肠外科杂志, 2004, 7(3): 214–7.
- [6] 李郅琴, 杜建强, 聂斌, 等. 特征选择方法综述 [J]. 计算机工程与应用, 2019, 55(24): 10–9.
- [7] 闫慈, 田翔华, 阿拉依·阿汗, 等. 基于 Lasso 特征选择的代谢综合征征数据分类 [J]. 公共卫生与预防医学, 2017, 28(6): 31–3.
- [8] 罗治文, 陈晓, 张业繁, 等. 机器学习算法和 COX 列线图在肝细胞癌术后生存预测中的应用价值 [J]. 中华消化外科杂志, 2020, 19(2): 166–7.
- [9] 耿尧. 基于随机生存森林的改进肿瘤预后预测模型 [D]. 大连: 大连海事大学, 2019.
- [10] 黄庆愿, 高钰琪, 刘福玉, 等. 急性高原反应相关因素的多元线性回归分析 [J]. 第三军医大学学报, 2006, 28(12): 1269–72.
- [11] Marisa L, De Reyniès A, Duval A, et al. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value [J]. PLoS Medicine, 2013, 10(5): e1001453.
- [12] Hocking R R. A Biometrics invited paper: The analysis and selection of variables in linear regression [J]. Biometrics, 1976, 32(1): 1–49.
- [13] Ripley B, Venables B, Bates D M, et al. Package ‘mass’ [M]. Cran R, 2013: 538.
- [14] Team RDC. R: a language and environment for statistical computing. R Foundation for statistical computing: Vienna, Austria [J]. Computing, 2013, 1: 12–21.
- [15] 吴志平, 张一帆, 邓信朗, 等. 肝内胆管癌根治术后生存的列线图预测模型 [J/CD]. 中华普通外科学文献(电子版), 2020, 14(2): 141–6.
- [16] 刘奥, 黄海, 张传杰, 等. 基于 2014 ISUP 分组预测根治性前列腺切除术后病理升级的列线图的建立与内部验证 [J]. 中华泌尿外科杂志, 2020, 41(4): 297–302.
- [17] 刘维, 高辉, 李华, 等. 基于 DSBs 修复蛋白表达构建预测食管鳞癌化疗预后的列线图模型 [J]. 实用医学杂志, 2020, 36(8): 1041–7.
- [18] 辛道, 刘偲, 刘航睿, 等. 基于 SEER 数据库的食管癌患者预后列线图模型构建与验证 [J]. 食管疾病, 2020, 2(1): 63–8.
- [19] 辛道, 刘偲, 刘航睿, 等. 结肠癌切除术患者预后列线图模型及危险分层系统构建与验证 [J]. 中国肿瘤临床, 2019, 46(22): 1145–54.
- [20] 詹天成, 张大奎, 彭亦凡, 等. 国内单中心 1321 例结直肠癌术后的临床分析 [J]. 实用肿瘤杂志, 2016, 31(4): 353–6.
- [21] 纪云. 结直肠癌淋巴结切除数目的变化与预后 [J]. 中国药物与临床, 2017, 17(6): 903–4.
- [22] 裴俊鹏, 张春东, 戴冬秋. 三种淋巴结分期方法对结直肠癌患者预后准确性的研究进展 [J]. 现代肿瘤医学, 2020, 28(4): 683–6.

## Stablish prognostic nomogram of colon cancer nomograms: a retrospective study based on the SEER database

Huang Kai, Zhang Changle, Wu Wenyong, et al

(*Dept of Gastrointestinal Surgery, The First Affiliated Hospital  
of Anhui Medical University, Hefei 230022*)

**Abstract Objective** To analyze the prognosis of patients with colon cancer risk factors, constructs the nomogram prediction of colon cancer patients 1 year, 3 years and 5 years survival rate and to analyze clinical benefit. **Methods**

A total of 53 753 colon cancer patients with pathologically confirmed between 2010 and 2015 were obtained from the Surveillance, Epidemiology, and End Results (SEER) program. All samples were randomly divided into training ( $n = 35\,835$ ) and validation ( $n = 17\,918$ ) set based on the prognosis status and overall survival time. Univariate and multivariate analysis were performed using Cox proportional hazards regression model to evaluate the prognostic value of 9 clinical parameters including age, pathological stage, AJCC 7th edition T stage, N stage, M stage, positive lymph node ratio (LNR), tissue grade, gender, and chemotherapy. The stepwise regression method was used to identify the potential prognostic risk parameters and a nomogram prognostic model was then established to predict the overall survival (OS). The internal and external calibration curve and decision curve analysis were conducted to evaluate the predictive accuracy of the nomogram. **Results** Age at diagnosis, T, N, M, LNR, tissue grade, gender and chemotherapy were identified as independent prognostic risk parameters of OS based on the training set in both univariate and multivariate analysis ( $P < 0.01$ ). Four clinical parameters including age, T stage, M stage and LNR were screened by stepwise regression analysis and a nomogram prognostic model was established based on the four factors. The average area under the curve (AUC) of ROC curves for 1-year, 3-year and 5-year OS was 0.708 (95% CI: 0.669 ~ 0.753) in the training set and 0.704 (95% CI: 0.681 ~ 0.715) in the validation set. The calibration curve for probability of survival showed excellent consistency between nomogram prediction and actual observation in the entire set, with a bootstrap-corrected concordance index of 0.748 (95% CI: 0.741 ~ 0.755). **Conclusion** We successfully developed a nomogram incorporating four clinical parameters to predict overall survival of colon cancer patients after surgery resection, which showed better performance than the traditional TNM staging system. The established nomogram might help oncologist make appropriate treatment decisions for colon cancer patients.

**Key words** colon cancer; feature selection; nomogram; prognosis; survival analysis

(上接第 298 页)

intestinal inflammation after infection of BALB/c suckling mice. **Methods** 7-day-old SPF grade BALB/c suckling mice were administrated with RV by gavage. The experiment was divided into five groups: normal control group, PBS control group, RV infection group, RV + STAT3 inhibitor group and STAT3 inhibitor group. Five suckling mice were killed at 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10 days after infection, respectively. Enzyme linked immunosorbent assay (ELISA) was used to detect the change of IL-22 content in intestinal homogenate, Western blot was used to detect the expression of STAT3, pSTAT3 and Reg III- $\gamma$  protein at different time points in each group. **Results** After RV infection, the intestinal tissue showed obvious inflammatory reaction, and the level of IL-22 increased gradually with the infection time which reached the maximum value on the fourth day of infection. The protein levels of pSTAT3 and Reg III- $\gamma$  showed the similar change trend with IL-22, while the protein level of STAT3 showed the opposite trend, and the expression level was the lowest on the fourth day after infection. **Conclusion** RV can activate IL-22/pSTAT3/Reg III- $\gamma$  signaling pathway and regulate the development of intestinal inflammation.

**Key words** rotavirus; IL-22; pSTAT3; Reg III- $\gamma$ ; intestinal inflammation