·论　　著·
（临床研究）

# 基于机器学习算法的胰腺鳞状细胞癌预后模型的构建与验证

黄坤，章慧，赵攀，何运胜
绵阳市中医医院普外科，四川 绵阳 621000

［摘要］ **目的**　胰腺鳞状细胞癌（pancreatic squamous cell carcinoma，PSCC）预后差，缺乏个体化的预后预测工具。研究通过SEER数据库中的大样本真实世界数据，基于机器学习算法，构建PSCC病人预后列线图，旨在精准化、个体化评价PSCC病人的预后，为临床决策制定提供参考。**方法**　提取SEER数据库中2000～2019年期间经病理学确诊的PSCC病人的临床资料，以7∶3的比例随机划分为训练集和验证集，采用多因素Cox比例风险模型、LASSO回归模型和随机生存森林模型筛选变量，构建预测3、6、12个月肿瘤特异性生存期（cancer-specific survival，CSS）和总生存期（overall survival，OS）的Nomogram模型，利用一致性指数、受试者操作特征曲线操作、校准曲线、生存曲线、决策曲线分析对模型进行验证和评估。**结果**　本研究共纳入367例病人，其中训练集256例，验证集111例。训练集和验证集病人的中位随访时间分别为3（1，7）个月和2（1，8）个月。两组间的基线特征均衡（均$P>0.05$）。多因素Cox比例风险模型显示：肿瘤大小、T分期、M分期、手术、化疗是OS和CSS的独立影响因素（均$P<0.05$）。LASSO回归模型显示：M分期、手术、化疗和OS、CSS相关。随机生存森林模型显示，影响OS的重要性评分前4位变量分别为化疗、M分期、手术和年龄，而影响CSS的重要性评分前4位变量分别为化疗、M分期、手术和肿瘤大小。基于这些因素所构建的Nomogram用于预测病人3、6个月的OS和CSS。验证结果表明：对于OS，训练集和验证集中一致性指数分别为0.753（95%*CI*：0.720～0.790）和0.723（95%*CI*：0.660～0.780）；对于CSS，两者分别为0.749（95%*CI*：0.720～0.780）和0.721（95%*CI*：0.660～0.780）。受试者操作特征曲线操作显示：对于3个月OS，训练集和验证集的曲线下面积（AUC）分别为79.8%和75.9%；对于6个月OS，训练集和验证集的AUC为78.9%和76.8%；对于12个月OS，训练集和验证集的AUC为78.7%和77.5%；对于3个月CSS，训练集和验证集的AUC为79.3%和76.3%；对于6个月CSS，训练集和验证集的AUC为78.6%和76.9%；对于12个月CSS，训练集和验证集的AUC为77.8%和77.4%。校准曲线均靠近理想的45°参考线，表现出良好的一致性。**结论**　年龄、M分期、肿瘤部位、手术和化疗是病人预后的独立影响因素。研究构建的Nomogram预测模型具有良好的预测价值，有利于临床对PSCC病人选择个性化治疗。

［关键词］ 胰腺肿瘤；癌鳞状细胞；列线图；预后；SEER数据库
［中图分类号］R735.9　［文献标识码］A　**DOI**：10.3969/j.issn.1003-5591.2024.04.006

**Developing and validating nomograms for predicting survival in patients with pancreatic squamous cell carcinoma based upon machine learning**

*Huang Kun*, *Zhang Hui*, *Zhao Pan*, *He Yunshen*
*Department of General Surgery*, *Mianyang Hospital of Traditional Chinese Medicine*, *Sichuan*

基金项目：绵阳市卫健委课题（202309），绵阳市中医医院课题（MYSZYYYKT202317），成都中医药大学校院联合创新基金（LH202402010）

作者简介：黄坤，副主任医师，主要从事普外基础与临床研究，Email：mshuangkun@163.com
通信作者：何运胜，Email：hys2520856@163.com

*Mianyang* 621000, *China*
*Corresponding author*: *HE Yunshen*, *Email*: *hys*2520856@163. *com*

[**Abstract**] **Objective** Pancreatic Squamous Cell Carcinoma (PSCC) has a poor prognosis and it lacks individualized prognostic tools. This study aimed to construct prognostic nomograms for PSCC patients based upon machine learning and using large-scale real-world data from the database of SEER, provide precise and individualized prognostic assessments and offer valuable references for clinical decision-making. **Methods** From 2000 to 2019, the relevant clinical data of 367 pathologically diagnosed PSCC patients were extracted from the database of SEER. They were randomized by a ratio of 7∶3 into training ($n$=256) and verification ($n$=111) sets. Multivariate Cox proportional hazard model, LASSO regression and random survival forest model were utilized for identifying independent prognostic factors for patient survival. These factors were utilized for constructing nomograms for predicting cancer specific survival (CSS) and total survival (OS) at Month 3/6. Subsequently, the models were internally and externally validated in training and validation sets by concordance index (C-index), receiver operating characteristic (ROC) and calibration curves for assessing their accuracy and predictive capacity. **Results** The median follow-up period in training and verification sets were 3(1,7) and 2(1,8) month. Baseline profiles were comparable between two groups (all $P$>0.05). Multivariate Cox proportional hazard model analysis indicated that tumor size, M/N stage, surgery and chemotherapy were independent influencing factors for OS/CSS. LASSO regression analysis revealed that M stage, surgery and chemotherapy were associated with OS/CSS. For OS, top four scoring variables for via random survival forest model were chemotherapy, M stage, surgery and age; For CSS, chemotherapy, M stage, surgery and tumor size. Nomograms for predicting OS/CSS at Month 3/6 were developed based upon these independent prognostic factors. Validation results showed that C-index for OS in training and verification sets were 0.753 (95%*CI*: 0.720-0.790) and 0.723 (95%*CI*: 0.660-0.780) and for CSS 0.749 (95%*CI*: 0.720-0.780) and 0.721 (95%*CI*: 0.660-0.780). ROC curve analysis indicated that AUC values for OS in training and verification sets were 79.8% and 75.9% at Month 3, 78.9% and 76.8% at Month 6 and 78.7% and 77.5% at Month 12; for CSS, 79.3% and 76.3% at Month 3, 78.6% and 76.9% at Month 6 and 77.4% and 78.4% at Month 12 respectively. Calibration curve analysis demonstrated a decent agreement between predicted and actual OS/CSS. Both were closely situated near ideal 45° reference line, demonstrating a high consistency. **Conclusion** Age, M stage, tumor size, surgery and chemotherapy are independent prognostic factors for PSCC patients. And the above constructed nomogram prediction models exhibit favorable predictive value and facilitate personalized therapeutics for PSCC patients in clinical practices.

[**Key words**] Pancreatic neoplasms; Carcinoma, Squamous cell; Nomogram; Prognosis; SEER

胰腺癌是一种起病隐匿且高度恶性的肿瘤[1]，是欧美国家癌症致死的第四大原因[2-4]。据统计，我国1990年至2020年期间胰腺癌的发病和死亡率呈现持续上升趋势，2020年我国胰腺癌的发病率和死亡率分别为8.32/100 000和8.48/100 000[5]；预计到2030年胰腺癌的发病率和死亡率将会进一步升高，分别达到12.71/100 000和12.94/100 000[6]。胰腺鳞状细胞癌(pancreatic squamous cell carcinoma, PSCC)，是胰腺癌中极为罕见的一种病理学类型，约占胰腺癌的0.3%[7]。PSCC是一种预后极差的高侵袭性恶性肿瘤[8]，中位(总)生存时间为4个月，1年、2年和5年的生存率分别为11.5%、5.7%、3.5%[7]。PSCC发病罕见、临床预后极差，目前缺乏准确的预后评价工具。列线图可同时纳入多个独立预后影响因素，个体化地预测病人生存概率，因而被临床广泛应用[9-12]。机器学习是一种实现人工智能的基本方法，通过特定算法来解析数据、并从中不断学习，最后对真实世界中的事件做出决策和预测，其准确性高，在病情评估、医疗决策、预后分析、个性化诊断与治疗等方面应用广泛[13]。本研究基于美国SEER(the Surveillance, Epidemiology and End Results, SEER)数据库临床病例样本量大、资料丰富的特点[14-15]，用机器学习算法构建PSCC病人预后列线图，旨在精准化、个体化评价PSCC病人的预后，为临床决策制定提供参考。

## 资料与方法

一、资料来源

利用SEER*Stat v8.3.9软件提取SEER数据库中2000～2019年期间通过病理学确诊PSCC的病人临床资料，随访截止日期为2019年12月31日。

二、纳入和排除标准

纳入标准：(1)病理学确诊为原发性PSCC的病人；(2)确诊年份为2000～2019年；(3)ICD-O-3编码为“8070/3”，“8071/3”，“8072/3”，“8074/3”，“8075/

3”,“8083/3”,“8084/3”。排除标准:(1)多源性肿瘤;(2)经过尸检或死亡证明确诊的病例;(3)研究指标无法获取;(4)临床或随访信息记录不完整。

三、提取数据

参照先前的研究[16-19],提取病人的诊断年龄、种族、性别、婚姻状况、肿瘤大小、分期、治疗信息、生存时间及生存结局等资料。根据数据特征并结合文献报道[20-24],本研究中,年龄界定标准设置为<60 岁和≥60 岁,肿瘤大小分类标准设置为<5 cm、5～10 cm 和>10 cm。依据 PSCC 病人的年龄、性别、种族、婚姻状况、肿瘤大小、肿瘤分期、是否接受手术、和化疗对其临床资料进行分组后用于后续分析。

四、统计学方法

本研究的研究终点分别为:肿瘤特异性生存期(cancer-specific survival, CSS)和总生存期(overall survival, OS)。CSS 定义:随访期间因 PSCC 死亡的时长(月),非 PSCC 死亡、随访截止或失访数据作为删失数据。OS 定义:随访期间因任何病因死亡的时长(月),随访截止或失访数据作为删失数据[9,25]。应用 Stata/MP 16.0 软件和 R(version 4.2.3)软件进行统计分析。正态分布计量资料以 $\bar{x}\pm s$ 表示,非正态分布计量资料采用 $M$ ($IQR$)进行描述。计数资料采用率表示,组间对比采用 $\chi^2$ 检验。按照 7∶3 将数据集随机划分为训练集和验证集,训练集用于列线图模型的构建及其内部验证,验证集被用于外部验证。通过 bootstrap 1 000 次重抽样的方法,绘制校准曲线对模型准确度进行验证;计算一致性指数(C-index)对模型区分度进行验证。采用 Kaplan-Meier 法计算生存率,生存率比较采用 log-rank 法。建模变量筛选原则:先构建单因素 Cox 比例风险模型,根据单因素 Cox 分析结果并结合临床重要性,进入多因素 Cox 比例风险模型分析,最终选出独立预后因素。根据多因素 Cox 比例风险模型计算风险比($HR$)及对应 95% 置信区间($CI$)[26]。在多因素 Cox 风险比例模型分析中,用方差扩展因子评估模型多重共线性,当取值大于 10,表明变量间存在严重的多重共线性问题,需排除后方可进行后续分析。在本研究中,同步运用随机生存森林模型进行独立的预后分析。随机生存森林是一种机器学习算法,具有高鲁棒性,它不受比例风险假定、对数线性假定等条件的约束,并且能够通过两个随机采样的过程来防止算法的过度拟合问题[27]。所有检验均为双侧尾。检验水准 $\alpha$=0.05。

## 结　果

一、纳入病历的基线特征

本研究共纳入 367 例 PSCC 病人,其中训练集 256 例,验证集 111 例。中位年龄 70 岁(IQR:59～78 岁),老年(>60 岁)273 例(74.4%),胰头肿瘤 40.6%(149/367),$T_3$ 和 $T_4$ 期 70.3%(78/111),$N_1$ 期 45.8%(55/120),$M_1$ 期 63.2%(134/212)。两组病人的临床病理特征见表 1。由表 1 可见,两组间的基线特征均衡(均 $P$>0.05)。

**表 1**　训练集和验证集中胰腺鳞状细胞癌病人的临床基线特征

| 变量 | 全组($n$=367) | 训练集($n$=256) | 验证集($n$=111) | $P$值 |
|---|---|---|---|---|
| 年龄[$n$(%)] | | | | 0.066 |
| <60 岁 | 94(25.6) | 58(22.7) | 36(32.4) | |
| ≥60 岁 | 273(74.4) | 198(77.3) | 75(67.6) | |
| 性别[$n$(%)] | | | | 0.520 |
| 女 | 173(47.1) | 124(48.4) | 49(44.1) | |
| 男 | 194(52.9) | 132(51.6) | 62(55.9) | |
| 种族[$n$(%)] | | | | 0.668 |
| 黑种人 | 52(14.2) | 39(15.2) | 13(11.7) | |
| 白种人 | 272(74.1) | 187(73.0) | 85(76.6) | |
| 其他种族 | 43(11.7) | 30(11.7) | 13(11.7) | |
| 婚姻状况[$n$(%)] | | | | 0.745 |
| 已婚 | 192(52.3) | 132(51.6) | 60(54.1) | |
| 未婚 | 175(47.7) | 124(48.4) | 51(45.9) | |
| 肿瘤部位[$n$(%)] | | | | 0.945 |
| 胰头 | 149(40.6) | 103(40.2) | 46(41.4) | |
| 胰体 | 59(16.1) | 43(16.8) | 16(14.4) | |
| 胰尾 | 67(18.3) | 47(18.4) | 20(18.1) | |
| 其他 | 92(25.1) | 63(24.6) | 29(26.1) | |
| 肿瘤大小[$n$(%)] | | | | 0.645 |
| <5 cm | 102(27.8) | 73(28.5) | 29(26.1) | |
| 5～10 cm | 71(19.3) | 53(20.7) | 18(16.2) | |
| >10 cm | 151(41.1) | 101(39.5) | 50(45.1) | |
| 不清楚 | 43(11.7) | 29(11.3) | 14(12.6) | |
| T 分期[$n$(%)] | | | | 0.338 |
| $T_2$ | 33(9.0) | 27(10.5) | 6(5.4) | |
| $T_3$ | 45(12.3) | 33(12.9) | 12(10.8) | |
| $T_4$ | 33(9.0) | 21(8.2) | 12(10.8) | |
| 不清楚 | 256(69.8) | 175(68.4) | 81(73) | |
| N 分期[$n$(%)] | | | | 0.131 |
| $N_0$ | 65(17.7) | 52(20.3) | 13(11.7) | |
| $N_1$ | 55(15.0) | 36(14.1) | 19(17.1) | |
| 不清楚 | 247(67.3) | 168(65.6) | 79(71.2) | |
| M 分期[$n$(%)] | | | | 0.488 |
| $M_0$ | 78(21.3) | 57(22.3) | 21(18.9) | |
| $M_1$ | 134(36.5) | 96(37.5) | 38(34.2) | |
| 不清楚 | 155(42.2) | 103(40.2) | 52(46.8) | |
| 化疗[$n$(%)] | | | | 0.620 |
| 否 | 194(52.9) | 138(53.9) | 56(50.5) | |
| 是 | 173(47.1) | 118(46.1) | 55(49.5) | |
| 手术[$n$(%)] | | | | 0.514 |
| 否 | 337(91.8) | 233(91.0) | 104(93.7) | |
| 是 | 30(8.2) | 23(9.0) | 7(6.3) | |

二、PSCC病人的预后分析

（一）单因素Cox比例风险模型

在训练集256例病人中，中位随访3（1，7）个月。在验证集111例病人中，中位随访2（1，8）个月。单因素Cox比例风险模型分析显示：肿瘤部位、大小、M分期、化疗、手术与PSCC病人的OS和CSS相关（均$P<0.05$）（表2）。

**表2**　胰腺鳞状细胞癌病人的单因素Cox分析

| 变量 | 总生存期 | | 肿瘤特异性生存期 | |
|---|---|---|---|---|
| | *HR*[95%*CI*] | *P*值 | *HR*[95%*CI*] | *P*值 |
| 年龄 | | | | |
| ＜60岁 | 1 | | 1 | |
| ≥60岁 | 1.32（0.98，1.79） | 0.069 | 1.33（0.97，1.82） | 0.078 |
| 性别 | | | | |
| 男 | 1 | | 1 | |
| 女 | 1.07（0.83，1.38） | 0.585 | 1.11（0.86，1.45） | 0.423 |
| 种族 | | | | |
| 白种人 | 1 | | 1 | |
| 黑种人 | 1.30（0.91，1.85） | 0.143 | 1.37（0.95，1.97） | 0.093 |
| 其他种族 | 1.11（0.75，1.65） | 0.595 | 1.18（0.78，1.78） | 0.423 |
| 婚姻状况 | | | | |
| 未婚 | 1 | | 1 | |
| 已婚 | 0.85（0.66，1.10） | 0.218 | 0.91（0.70，1.18） | 0.481 |
| 肿瘤部位 | | | | |
| 胰头 | 1 | | 1 | |
| 胰体 | 0.85（0.59，1.23） | 0.389 | 0.92（0.63，1.34） | 0.651 |
| 胰尾 | 1.19（0.83，1.69） | 0.345 | 1.13（0.78，1.65） | 0.508 |
| 其他 | 1.53（1.11，2.11） | 0.009 | 1.43（1.02，2.01） | 0.036 |
| 肿瘤大小 | | | | |
| ＜5 cm | 1 | | 1 | |
| 5～10 cm | 1.59（1.11，2.29） | 0.012 | 1.70（1.18，2.45） | 0.005 |
| ＞10 cm | 1.21（0.88，1.65） | 0.240 | 1.10（0.80，1.53） | 0.552 |
| 不清楚 | 1.58（1.03，2.45） | 0.038 | 1.31（0.81，2.11） | 0.269 |
| T分期 | | | | |
| $T_2$ | 1 | | 1 | |
| $T_3$ | 1.02（0.61，1.72） | 0.934 | 0.97（0.57，1.66） | 0.910 |
| $T_4$ | 1.11（0.62，2.00） | 0.720 | 1.05（0.58，1.90） | 0.881 |
| 不清楚 | 1.04（0.68，1.57） | 0.870 | 0.93（0.61，1.42） | 0.741 |
| N分期 | | | | |
| $N_0$ | 1 | | 1 | |
| $N_1$ | 0.78（0.50，1.21） | 0.269 | 0.79（0.50，1.25） | 0.316 |
| 不清楚 | 0.86（0.63，1.18） | 0.360 | 0.84（0.60，1.16） | 0.291 |
| M分期 | | | | |
| $M_0$ | 1 | | 1 | |
| $M_1$ | 2.35（1.66，3.35） | ＜0.001 | 2.34（1.63，3.36） | ＜0.001 |
| 不清楚 | 1.64（1.16，2.32） | 0.005 | 1.49（1.04，2.14） | 0.029 |
| 化疗 | | | | |
| 否 | 1 | | 1 | |
| 是 | 0.39（0.30，0.51） | ＜0.001 | 0.40（0.30，0.52） | ＜0.001 |
| 手术 | | | | |
| 否 | 1 | | 1 | |
| 是 | 0.31（0.19，0.52） | ＜0.001 | 0.32（0.19，0.55） | ＜0.001 |

（二）多因素Cox比例风险模型

多因素Cox比例风险模型分析结果显示：肿瘤大小、T分期、M分期、手术、化疗是PSCC病人OS和CSS的独立影响因素（均$P<0.05$）（图1）。

（三）LASSO回归模型

LASSO方法的原理是在最小二乘法的基础上增加一个惩罚项来压缩估计参数，进而对预后影响较大的自变量进行自动筛选并计算出相应的回归系数。模型筛选的主要参数是lambda. min和lambda. 1se。前者是指获得最小目标参量均值的λ值，而后者是指在lambda. min一个方差se范围内得到最简单模型的λ值。由于λ值到达一定值之后，再继续增加模型自变量个数并不能显著地提高模型性能。因此，lambda. 1se可以给出一个性能优良，同时自变量数最少的模型。为了尽可能减少变量共线性对结果的影响，同时采用基于10折交叉验证方法LASSO回归（以lambda. 1se为模型参数）进一步分析。结果显示：M分期、手术、化疗与PSCC病人的OS和CSS相关（图2）。

（四）随机生存森林模型

随机生存森林具有高鲁棒性的特点，不受比例风险假定、对数线性假定的约束，在分析过程中，能够通过两个随机采样的过程来防止算法的过度拟合问题。在研究中进一步通过该方法来筛选变量。结果显示，影响OS的重要性评分前4位变量分别为：化疗、M分期、手术和年龄（图3A）；而影响CSS的重要性评分前4位变量分别为：化疗、M分期、手术和肿瘤大小（图3B）。

三、PSCC病人生存预测模型的构建

基于多因素Cox回归模型、LASSO回归模型和随机生存森林模型的分析结果并结合临床相关性，研究最终选择：年龄、肿瘤部位、M分期、手术和化疗总共5个变量参与构建PSCC病人预后预测的Nomogram图。预测病人3、6、12个月的OS和CSS。图中每个变量都将被赋予相应的得分（表3），所有变量的分值之和等于总分值，分值越低，预后越好。根据总分值即可对不同时间点PSCC病人的OS和CSS做出预测（图4）。

四、PSCC病人生存预测模型的验证

通过C-index和时间依赖受试者操作特征（ROC）曲线对模型的区分度进行评价。结果显示，对于OS，训练集和验证集中C-index分别为0.753（95%*CI*：0.720～0.790）和0.723（95%*CI*：0.660～0.780）；对于CSS，两者分别为0.749（95%*CI*：

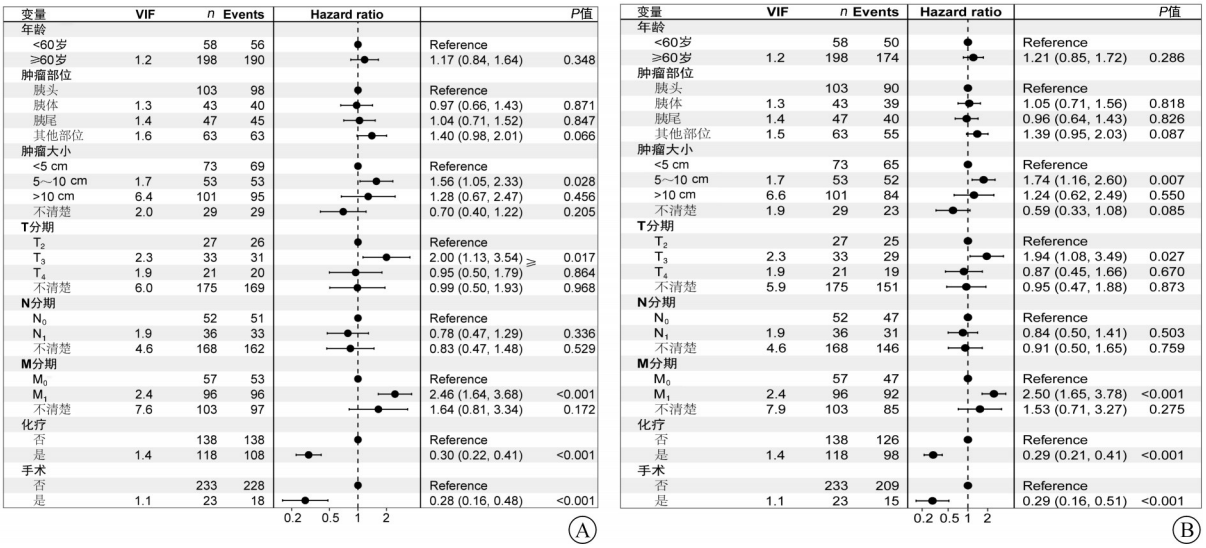



**图1**　胰腺鳞状细胞癌病人肿瘤特异性生存期和总生存期的森林图基于多因素Cox回归分析　A. 肿瘤特异性生存期；B. 总生存期

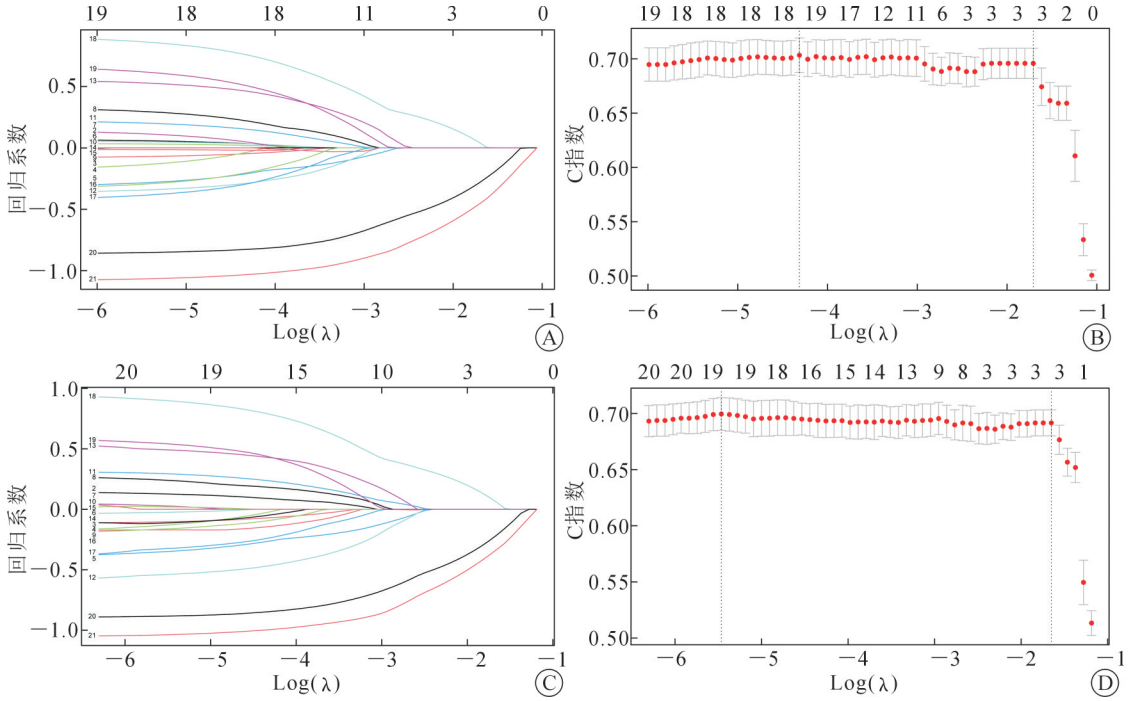



**图2**　基于LASSO回归的特征选择　A. LASSO回归系数随Log(λ)的变化曲线（总生存期）；B. 基于10折交叉验证C指数随Log(λ)的变化曲线（总生存期）；C. LASSO回归系数随Log(λ)的变化曲线（肿瘤特异性生存期）；D. 基于10折交叉验证C指数随Log(λ)的变化曲线（肿瘤特异性生存期）

0.720～0.780）和 0.721（95%*CI*：0.660～0.780）。ROC曲线显示：对于3个月OS，训练集和验证集的曲线下面积（AUC）分别为79.8%和75.9%；对于6个月OS，训练集和验证集的AUC分别为78.9%和76.8%；对于12个月OS，训练集和验证集的AUC分别为78.7%和77.5%；对于3个月CSS，训练集和验证集的AUC分别为79.3%和76.3%；对于6个月CSS，训练集和验证集的AUC分别为78.6%和76.9%；对于12个月CSS，训练集和验证集的AUC分别为77.8%和77.4%（图5），两者均表明模型具有良好的预测价值。为了评价模型的准确度，采用Bootstrap法对模型进行内部验证和外部验证，自抽样次数B=1 000，并绘制校准曲线。验证结果显示，在训练集（内部）与验证集（外部）中，PSCC病人的3、6、12个月的OS和CSS校正曲线均靠近理想的45°参考线，表明模型预测值与实际值之间具有

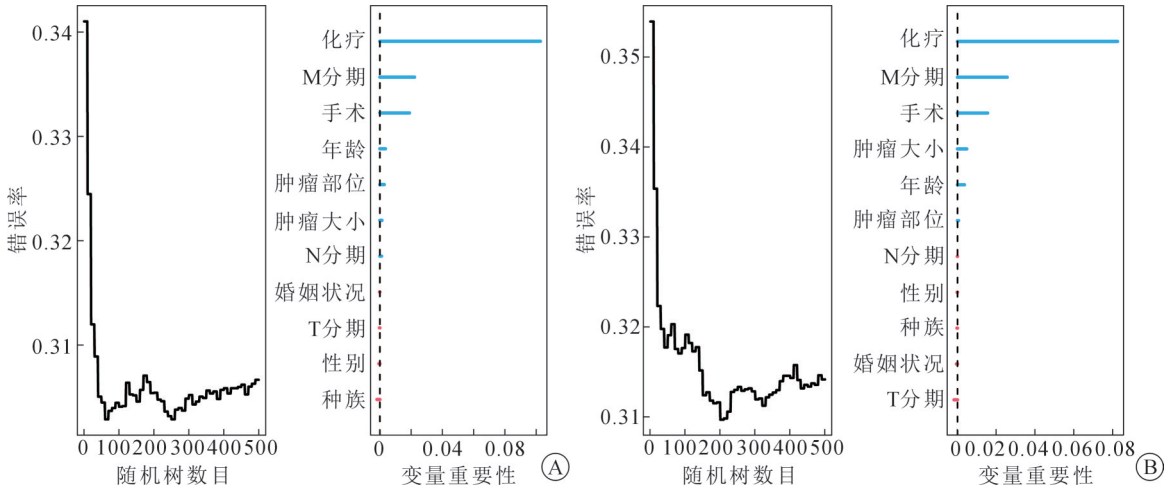
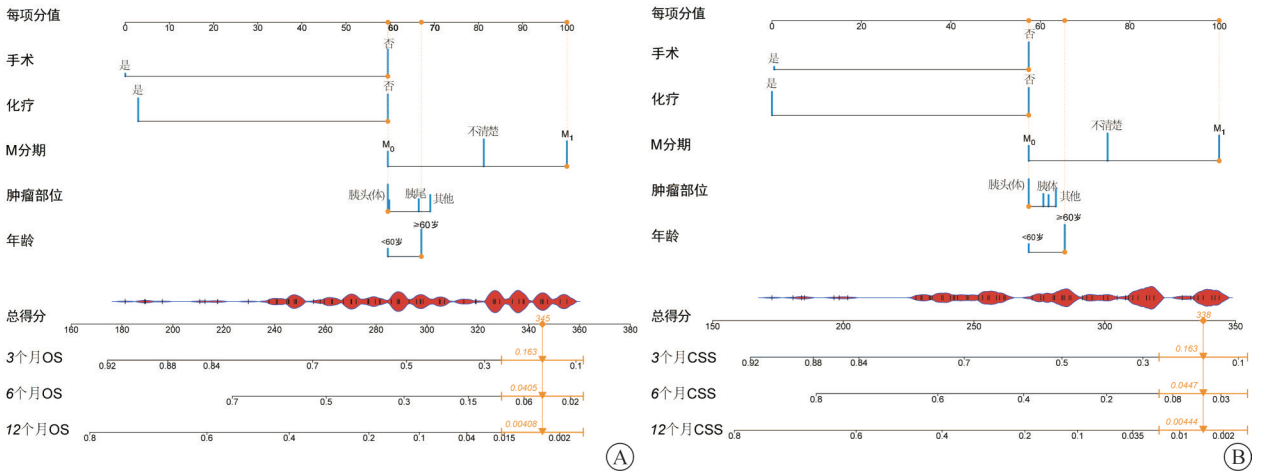



**图3**　基于随机生存森林模型评价变量对总生存期(A)和肿瘤特异性生存期(B)的重要性




注:OS,总生存期;CSS,肿瘤特异生存期

**图4**　预测胰腺鳞状细胞癌病人3、6、12个月的总生存期(A)和肿瘤特异性生存期(B)的列线图

良好的一致性(图6)。

五、模型的决策曲线分析

决策曲线分析(DCA)是一种新的预测模型评价方法,它更注重临床获益评价。研究中,通过DCA评估其在临床应用中该模型是否有获益。结果显示,与TNM分期相比,列线图模型在预测PSCC病人的OS和CSS均显示出更大的净收益(图7)。

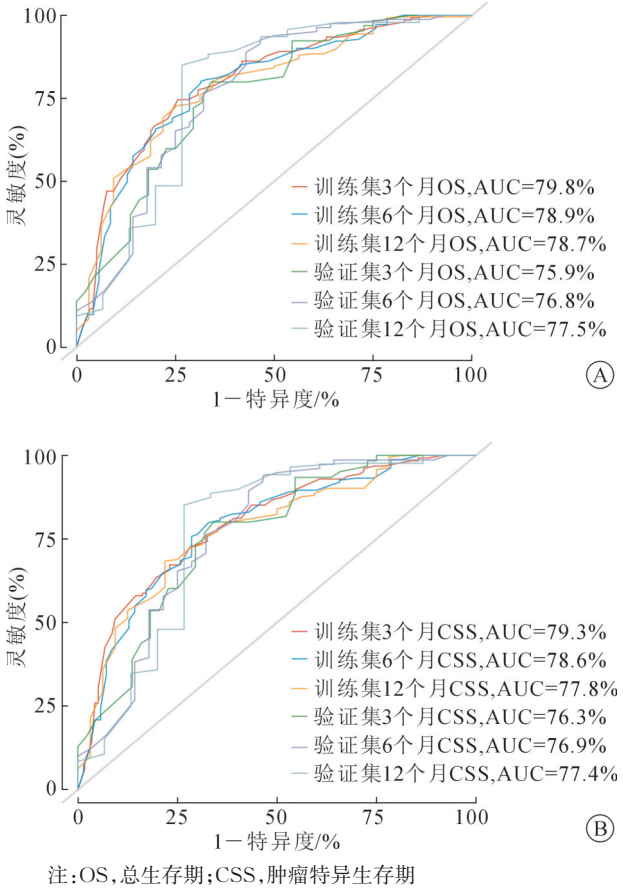
六、不同风险组PSCC病人的生存获益

为了进一步验证该模型在临床实践中的应用价值,根据构建的模型分别在训练集和验证集中计算出每个病人的总分值,将训练集中总分值的中位数设置为危险分层阈值。其中,对于OS和CSS,低风险组和高风险组的阈值分别为197分和198分。结果表明,在验证集和训练集中该模型均能对PSCC病人的生存预后做出良好的区分;生存曲线显示,在验证集和训练集中,相比于低风险组,高风险组病人预后更差(均$P<0.05$)(图8)。

## 讨　论

PSCC是一种极为罕见的恶性性肿瘤,具有高度的侵袭性[8]。由于发病罕见、预后差,缺乏针对性的临床预后评价工具。本研究利用SEER数据库基于机器学习算法构建PSCC病人预后列线图,旨在精准化、个体化评价PSCC病人的预后,为临床决策制定提供参考。

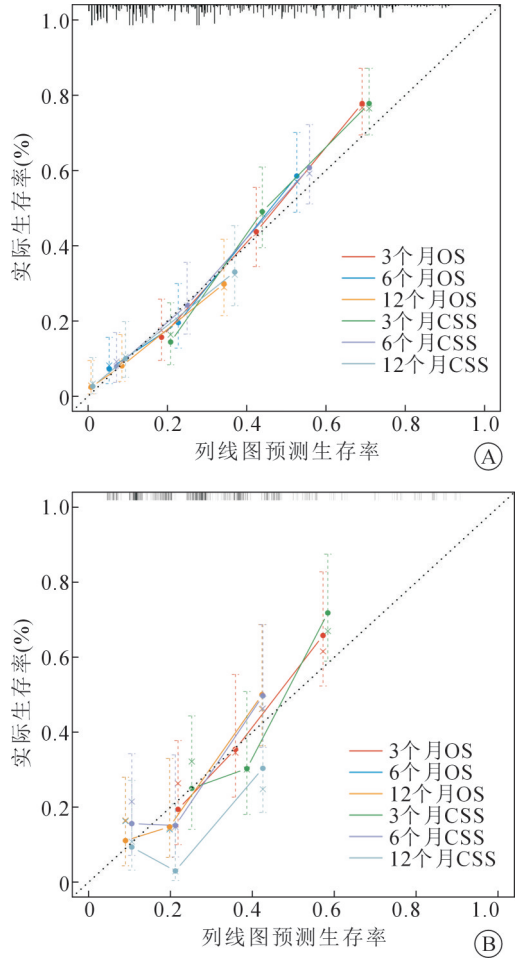
研究显示,PSCC以老年为主,好发于胰头,远处转移多见;手术、化疗是独立的预后影响因素,均能显著提高病人的OS和CSS。研究中,病人中位年龄70岁(*IQR*:59～78岁),老年(>60岁)273例(74.4%),胰头肿瘤40.6%(149/367),$T_3$和$T_4$期70.3%(78/111),$N_1$期45.8%(55/120),$M_1$期63.2%(134/212),低分化和未分化肿瘤80.3%(110/137)。Wang等[28]的一项包含124例PSCC的研究显示,老年(>60岁)、胰头肿瘤、$M_1$期病人的




注：OS，总生存期；CSS，肿瘤特异生存期

**图5**　不同时间点肿瘤特异生存期和总生存期模型验证的ROC曲线　A. 总生存期；B. 肿瘤特异生存期

构成比分别为74.4%、46.0%和52.4%。而在Makarova-Rusher等[29]的报道中，得出了相同的结论。上述研究结论同本研究一致的。在本研究中，多因素Cox回归分析显示，和未行手术治疗病人相比，手术病人的总死亡风险和肿瘤相关性死亡风险分别降低72.0%和71.0%。和未行化疗的病人相比，化疗病人的上述风险则分别下降70.0%和71.0%。并且，我们进一步通过LASSO和随机生存森林模型分析也结果得出了相同的结论。除此之外，江珍珍等[30]曾经报道1例胰体部的鳞状细胞癌67岁男性病人，在接受腹腔镜胰体尾切除联合脾切除术后，经过半年的随访，无复发和转移。何睿哲等[31]也报道一例胰头鳞状细胞癌43岁女性病人在接受胰十二指肠切除术，术后联合化疗和放疗，随访6个月未见复发和转移。金圣博等[32]报道一例原发性胰腺鳞状细胞癌伴肝转移病人经化疗取得良好的效果。上述报道，均支持此次研究结论。除此之外，在Qin等[33]和Wang等[28]的研究中也证实了手术和化疗能明显改善PSCC病人的预后，也支持这次的研究。




注：OS，总生存期；CSS，肿瘤特异生存期

**图6**　不同时间点总生存期和肿瘤特异生存期的移准曲线　A. 训练集；B. 验证集

列线图被广泛应用于肿瘤预后的评价，它可以个体化的预测病人特定时间点的累积生存概率，辅助临床医师做出针对性的临床决策[34-38]。在本研究中，我们基于上述PSCC预后的独立影响因素，成功构建了用于预测PSCC病人3、6个月OS和CSS的预测模型。通过进一步的预测模型验证显示，其具有较高的准确度和一致性。相比于TNM分期，该模型的临床净获益更大。不同于TNM分期，列线图可以直观的预测病人的生存概率[9,39]，帮助临床医师对病人的预后做出更准确的判断。在此基础上，还可进一步根据模型得分，对病人进行危险分层，从而区分病人是否属于高危人群，帮助临床医师做出最佳的临床决策。例如：1例病人，肿瘤位于胰头，未接受手术和化疗，$M_1$期，该病人的OS模型总得分为345分，其3、6、12个月的累积生存率分别为16.3%、4.1%、0.4%；该病人的CSS模型总得分为338分，其3、6、12个月的累积生存率分别为

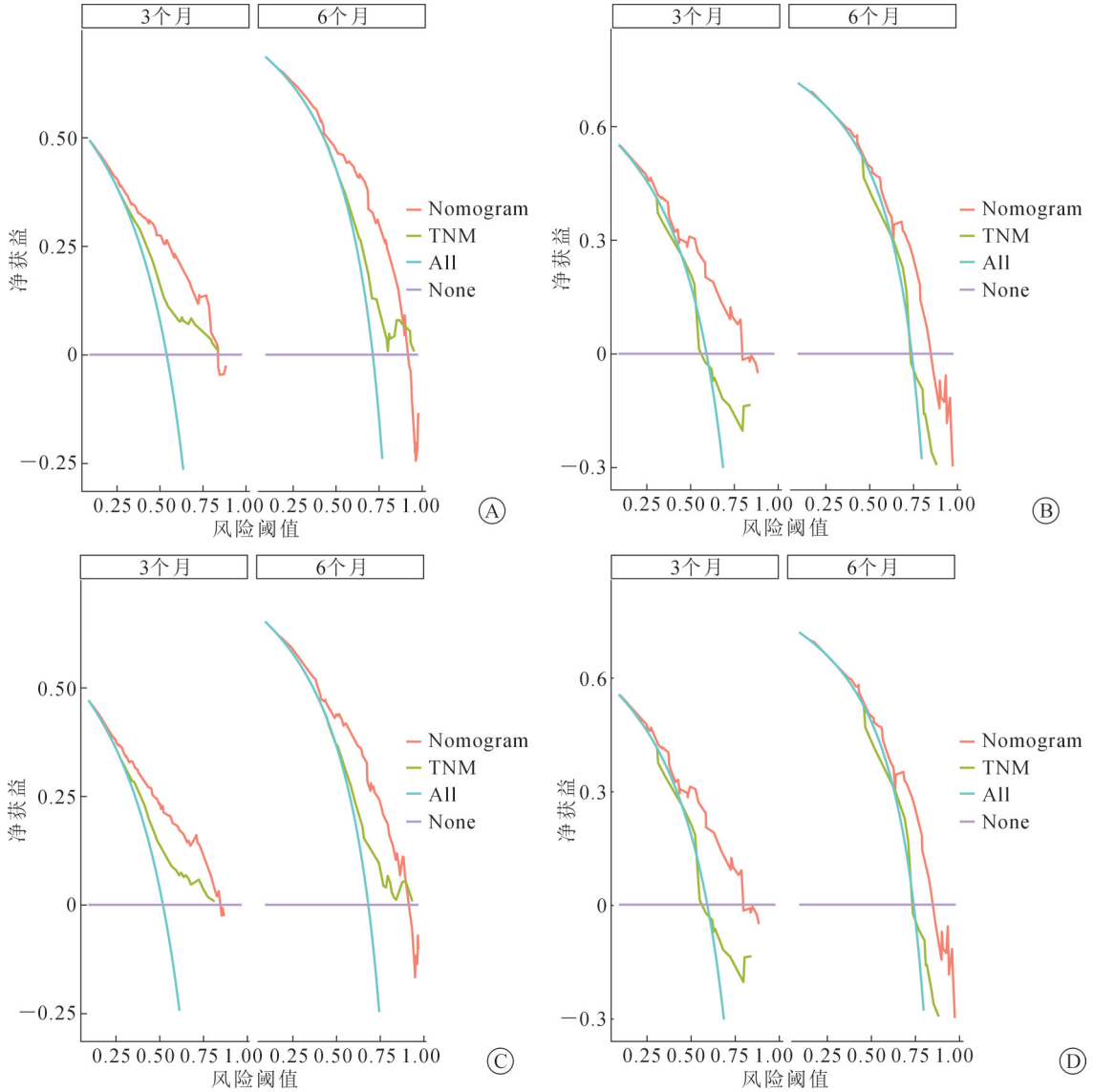



**图7**　列线图和TNM分期预测3个月、6个月总生存期和肿瘤特异生存期的决策曲线分析　A. 训练集总生存期；B. 验证集总生存期；C. 训练集肿瘤特异生存期；D. 验证集肿瘤特异生存期

16.3%、4.5%、0.4%（图4）。由于影响病人的OS的因素相对复杂，而相比之下CSS显得更加重要[34]。进一步根据CSS对该病人进行危险分层。由于其总分338分高于阈值198分，因此该病人被认定为高危人群，需要更加密切观察病人，制定更为严格的治疗和随访计划。

目前国内外尚无关于PSCC的预后模型的大样本真实世界研究的报道。本研究主要有以下优势：首先，在分析过程中，我们使用了随机生存森林模型。该模型作为一种机器学习算法，属于集成学习范畴，具有高鲁棒性，且不受比例风险假定、对数线性假定等条件的限制，而且还能够通过两个随机采样的过程来防止算法的过度拟合问题[27]。因而，该算法得出的结论将更加可靠和稳定。其次，在分析过程中，本研究也同样独立地运用了LASSO、经典的单因素、多因素Cox比例风险模型，并且利用方差扩展因子对模型的多重共线性问题进行了定量的评价与分析。最后，本研究总共纳入了367例病人，352（95.91%）例病人发生死亡结局事件，其中325（88.56%）病人出现肿瘤特异性死亡，因此，研究的样本含量和事件结局数，可以为统计分析提供了足够大的检验效能。

然而，研究仍然存在一定的局限性。首先，作为回顾性研究，虽然在设计时制定了严格的纳入和排除标准，但是选择偏倚仍是不可避免的。其次，SEER数据库是大型肿瘤数据库，容易受编码错误及删失值的影响。再次，由于SEER数据库数据获取权限的限制，部分数据（如：化疗、手术和放疗的

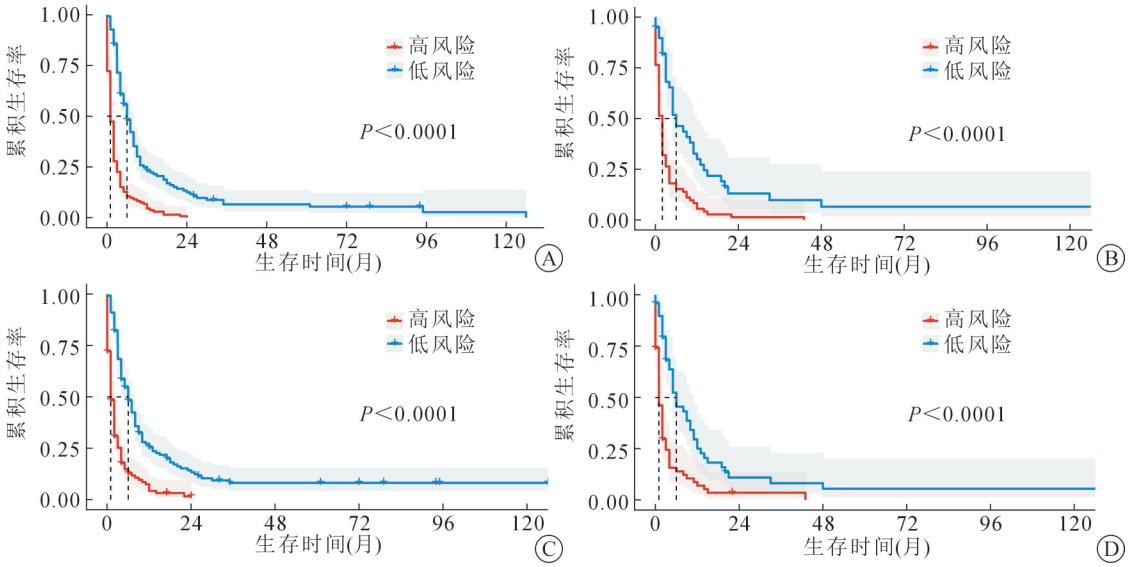



**图8**　不同风险病的总生存期和肿瘤特异生存期生存曲线　A. 训练集总生存期；B. 验证集总生存期；C. 训练集肿瘤特异生存期；D. 验证集肿瘤特异生存期

具体方案，肿瘤复发信息等等）是无法获取的，而这些数据有可能会对分析结果产生影响。最后，纳入研究的病例数中，部分组别病例数较少、构成比不均衡，可能影响检验效能。

综上所述，笔者认为年龄、肿瘤部位、M 分期、手术和化疗是病人预后的独立影响因素。基于这些因素所构建的预后预测模型对临床实践有较好的参考作用。

**利益冲突**　所有作者均声明不存在利益冲突

## 参　考　文　献


[1] 王健力，俞泽元，陈志涛，等. TRIM21 与胰腺癌临床病理特征及免疫浸润的相关性分析[J]. 重庆医科大学学报，2022，47（2）：201-208. DOI：10. 13406/j. cnki. cyxb. 002737.

[2] Timmer FEF，Geboers B，Nieuwenhuizen S，et al. Locoregional treatment of metastatic pancreatic cancer utilizing resection，ablation and embolization：a systematic review [J]. Cancers，2021，13（7）：1608. DOI：10. 3390/cancers13071608.

[3] Zhou WT，Wang DS，Lou WH. Current role of surgery in pancreatic cancer with synchronous liver metastasis [J]. Cancer Control，2020，27（1）：1073274820976593. DOI：10. 1177/1073274820976593.

[4] Siegel RL，Miller KD，Jemal A. Cancer statistics，2016[J]. CA Cancer J Clin，2016，66（1）：7-30. DOI：10. 3322/caac. 21332.

[5] 肖文博，朱广涵，朱云，等. 1990-2019 年中国主要消化系统恶性肿瘤发病变化分析[J]. 中国肿瘤，2022，31（9）：693-700. DOI：10. 11735/j. issn. 1004-0242. 2022. 09. A003.

[6] 张敏，段朝晖，徐杰茹，等. 基于 GBD 数据分析与预测 2000—2030 年中国胰腺癌发病与死亡趋势[J]. 中国肿瘤，2022，31（11）：862-868. DOI：10. 11735/j. issn. 1004-0242. 2022. 11. A003.

[7] Luo GP，Fan ZY，Gong YT，et al. Characteristics and outcomes of pancreatic cancer by histological subtypes[J]. Pancreas，2019，48（6）：817-822. DOI：10. 1097/MPA. 0000000000001338.

[8] 黎美青，张永俊，刘争进. 胰腺原发性透明细胞鳞状细胞癌一例[J]. 中华病理学杂志，2021，50（10）：1186-1188. DOI：10. 3760/cma. j. cn112151-20210722-00524.

[9] Chen MS，Liu PC，Yi JZ，et al. Development and validation of nomograms for predicting survival in patients with *de novo* metastatic triple-negative breast cancer [J]. Sci Rep，2022，12（1）：14659. DOI：10. 1038/s41598-022-18727-2.

[10] Huang C，Yu QP，Li H，et al. A novel nomogram model to predict the overall survival of patients with retroperitoneal leiomyosarcoma：a large cohort retrospective study[J]. Sci Rep，2022，12（1）：11851. DOI：10. 1038/s41598-022-16055-z.

[11] Zhang SL，Wang ZM，Wang WR，et al. Novel nomograms individually predict the survival of patients with soft tissue sarcomas after surgery[J]. Cancer Manag Res，2019，11：3215-3225. DOI：10. 2147/CMAR. S195123.

[12] Wang W，Hong JQ，Meng JH，et al. Nomograms predict cancer-specific and overall survival of patients with primary limb leiomyosarcoma [J]. J Orthop Res，2019，37（7）：1649-1657. DOI：10. 1002/jor. 24298.

[13] 刘红枝，林海涛，林昭旺，等. 机器学习算法在肝细胞癌微血管侵犯术前预测中的应用价值[J]. 中华消化外科杂志，2020，19（2）：156-165. DOI：10. 3760/cma. j. issn. 1673-9752. 2020. 02. 008.

[14] 黄坤，赵平武，白斗，等. 不同病理学类型的腹膜后脂肪肉瘤患者的预后分析及临床评价[J]. 中国普通外科杂志，2021，30（10）：1212-1228. DOI：10. 7659/j. issn. 1005-6947. 2021. 10. 011.

[15] 陈茂山，李芳芳，杨宏伟，等. 基于 SEER 数据库分析 142007 例乳腺癌诊断时婚姻状态与预后的关系[J]. 重庆医科大学学报，2020，45（11）：1567-1572. DOI：10. 13406/j. cnki. cyxb. 002325.

[16] 黄坤，赵平武，白斗，等. 不同病理学类型的腹膜后脂肪肉瘤患者的预后分析及临床评价[J]. 中国普通外科杂志，2021，7（10）：1212-1228. DOI：10. 7659/j. issn. 1005-6947. 2021. 10. 011.

[17] 黄坤,何运胜,张红梅,等. 手术治疗对胆囊小细胞癌患者预后价值的SEER数据库分析[J]. 中国普通外科杂志,2022,8(8):997-1005. DOI:10.7659/j.issn.1005-6947.2022.08.002.

[18] 刘千琳,李秋. 肝细胞癌靶向治疗联合免疫治疗的研究剖析[J]. 中国普外基础与临床杂志,2023,30(4):392-396. DOI:10.7507/1007-9424.202303055.

[19] 黄坤,何运胜,赵攀,等. 手术治疗对直肠小细胞神经内分泌癌患者预后影响的SEER数据库分析[J]. 中国普外基础与临床杂志,2023,30(2):179-184. DOI:10.7507/1007-9424.202208041.

[20] 程杨,陈小彬,魏志鸿,等. 术前淋巴细胞与C反应蛋白比值在预测胆囊癌术后复发中的应用及其临床价值[J]. 中国普通外科杂志,2023,32(2):190-199. DOI:10.7659/j.issn.1005-6947.2023.02.004.

[21] 王晓晓,陶立元,裴敏玥,等. 连续变量转换为分类变量的几种方法[J]. 中华儿科杂志,2022,60(5):420. DOI:10.3760/cma.j.cn112140-20220317-00212.

[22] 黄坤,何运胜,张红梅,等. 手术治疗对胆囊小细胞癌患者预后价值的SEER数据库分析[J]. 中国普通外科杂志,2022,31(8):997-1005. DOI:10.7659/j.issn.1005-6947.2022.08.002.

[23] Su BB, Bai DS, Yu JQ, et al. Can patients with pancreatic cancer and liver metastases obtain survival benefit from surgery? A population-based study[J]. J Cancer, 2021, 12(2): 539-552. DOI:10.7150/jca.51218.

[24] Pu N, Chen QD, Gan W, et al. Lymph node metastatic patterns and survival predictors based on tumor size in pancreatic ductal adenocarcinoma[J]. Adv Ther, 2021, 38(8): 4258-4270. DOI:10.1007/s12325-021-01819-2.

[25] 陈茂山,李芳芳,杨宏伟,等. 基于SEER数据库分析142007乳腺癌诊断时婚姻状态与预后的关系[J]. 2020,45(11):1567-1572. DOI:10.13406/j.cnki.cyxb.002325.

[26] Kim Y, Margonis GA, Prescott JD, et al. Nomograms to predict recurrence-free and overall survival after curative resection of adrenocortical carcinoma[J]. JAMA Surg, 2016, 151(4): 365-373. DOI:10.1001/jamasurg.2015.4516.

[27] 陈哲,许恒敏,李哲轩,等. 随机生存森林:基于机器学习算法的生存分析模型[J]. 中华预防医学杂志,2021,55(1):104-109. DOI:10.3760/cma.j.cn112150-20200911-01197.

[28] Wang YQ, Zhou YW, Chen YY, et al. Epidemiology, treatment, and outcome of pancreatic squamous cell carcinoma and pancreatic adenocarcinoma: a propensity score-matching analysis based on SEER-database [J]. Technol Cancer Res Treat, 2022, 21: 15330338221106533. DOI: 10.1177/15330338221106533.

[29] Makarova-Rusher OV, Ulahannan S, Greten TF, et al. Pancreatic squamous cell carcinoma: a population-based study of epidemiology, clinicopathologic characteristics and outcomes [J]. Pancreas, 2016, 45(10): 1432-1437. DOI:10.1097/MPA.0000000000000658.

[30] 江珍珍,刘夏天,侯传玲. 胰腺原发性鳞状细胞癌一例[J]. 中华胰腺病杂志,2018,18(2):127. DOI:10.3760/cma.j.issn.1674-1935.2018.02.018.

[31] 何睿哲,付学良,刘德军,等. 胰腺鳞状细胞癌一例[J]. 中华普通外科杂志,2018,33(9):741. DOI:10.3760/cma.j.issn.1007-631X.2018.09.008.

[32] 金圣博,丁雪丽,于亚男,等. 化疗联合经皮肝动脉化疗栓塞术治疗胰腺鳞状细胞癌并肝转移一例[J]. 中华医学杂志,2019,99(46):3667-3668. DOI:10.3760/cma.j.issn.0376-2491.2019.46.015.

[33] Qin WX, Wu Y, Liu J, et al. Primary squamous cell carcinoma of pancreas: a population-based study[J]. Gland Surg, 2021, 10(3): 1029-1037. DOI:10.21037/gs-20-317.

[34] Xu L, Wen N, Qiu JJ, et al. Predicting survival benefit of sparing sentinel lymph node biopsy in low-risk elderly patients with early breast cancer: a population-based analysis [J]. Front Oncol, 2020, 10: 1718. DOI:10.3389/fonc.2020.01718.

[35] Xiang JX, Zhang XF, Weiss M, et al. Multi-institutional development and external validation of a nomogram predicting recurrence after curative liver resection for neuroendocrine liver metastasis[J]. Ann Surg Oncol, 2020, 27(10): 3717-3726. DOI:10.1245/s10434-020-08620-5.

[36] 程晨,吴云桦,徐正水,等. Ⅱ～Ⅲ期结肠癌根治术后复发危险因素分析及其列线图预测模型的应用价值[J]. 中华消化外科杂志,2021,20(3):331-338. DOI:10.3760/cma.j.cn115610-20210120-00034.

[37] 邓家仲,荚卫东. 肝细胞癌微血管侵犯危险因素分析及术前预测列线图模型构建[J]. 中国普通外科杂志,2021,30(7):772-779. DOI:10.7659/j.issn.1005-6947.2021.07.003.

[38] 李吴寒,张营,潘晶晶,等. 胃癌患者预后相关影响因素的列线图模型构建及验证[J]. 中国普通外科杂志,2022,31(10):1381-1388. DOI:10.7659/j.issn.1005-6947.2022.10.014.

[39] Peng F, Qin TT, Wang M, et al. Development and validation of a nomogram to predict survival in pancreatic head ductal adenocarcinoma after pancreaticoduodenectomy [J]. Front Oncol, 2021, 11: 734673. DOI:10.3389/fonc.2021.734673.

（收稿日期:2023-10-20）