

临床研究

DOI: 10.13406/j.cnki.cyx.003180

前列腺癌风险预测模型的构建与验证

路 帅¹, 李文杰¹, 徐紫薇¹, 张浩轩², 陆 进²

(1. 蚌埠医学院临床医学院2019级, 蚌埠 233030; 2. 蚌埠医学院人体解剖学教研室, 蚌埠 233030)

【摘要】目的:筛选前列腺癌(prostatic cancer, PCa)发病的影响因素, 构建PCa风险预测模型并进行验证。**方法:**利用国家临床医学科学数据中心《前列腺肿瘤预警数据集》, 对数据处理后将数据按7:3随机分为建模组和验证组; 使用最小绝对值收缩和选择算子(least absolute shrink-age and selection operator, LASSO)对建模组进行筛选, 得到PCa特征指标; 对特征指标进行多因素logistic回归分析, 并利用其分析结果对建模组数据构建PCa风险预测模型, 同时利用建模组数据进行内部评价及验证组数据内部验证。**结果:**共纳入880例样本数据, 其中建模组616例, 验证组264例; 通过LASSO回归分析对筛选得到的14个特征指标进行多因素logistic回归分析, 结果显示球蛋白($OR=1.112$, $95\%CI=1.044\sim1.185$)、无机磷($OR=65.167$, $95\%CI=20.437\sim207.796$)、总前列腺特异性抗原(total prostate specific antigen, tPSA) ($OR=1.026$, $95\%CI=1.014\sim1.037$)与血清尿酸($OR=0.997$, $95\%CI=0.994\sim0.999$)的差异具有统计学意义($P<0.05$), 是PCa发病的独立影响因素; 利用其构建的PCa风险预测模型内部评价和内部验证的校准曲线准确度较高; 模型内部评价的曲线下面积(area under curve, AUC)为0.766($95\%CI=0.728\sim0.804$), 患者的决策曲线分析(decision curve analysis, DCA)净获益率为9%~72%; 而模型内部验证的AUC为0.704($95\%CI=0.639\sim0.768$), 患者的DCA净获益率为18%~59%及63%~64%。**结论:**球蛋白、无机磷、tPSA与血清尿酸是PCa的独立影响因素, 通过其构建的风险预测模型具有良好预测作用。

【关键词】前列腺癌; 前列腺增生; 风险预测模型**【中图分类号】**R604**【文献标志码】**A**【收稿日期】**2022-08-16

Construction and validation of a prostatic cancer risk prediction model

Lu Shuai¹, Li Wenjie¹, Xu Ziwei¹, Zhang Haoxuan², Lu Jin²

(1. Batch 2019, College of Clinical Medicine, Bengbu Medical College;

2. Teaching and Research Section of Human Anatomy, Bengbu Medical College)

【Abstract】Objective: To screen the influencing factors of the incidence of prostatic cancer (PCa), build a risk prediction model for PCa and validate it. **Methods:** Based on the Prostatic Cancer Early Warning Data Set of National Clinical Medical Science Data Center, and after the data processing, the data were randomly divided into a modeling group and a verification group according to 7:3. Least absolute shrink-age and selection operator (LASSO) regression was used to screen the PCa characteristic indicators of the modeling group, multifactor logistic regression analysis was carried out to analyze the characteristic indicators, and the analysis results were used to build a PCa risk prediction model for the data of the modeling group. At the same time, the data of the modeling group were used for internal evaluation and the data of the validation group for internal verification. **Results:** A total of 880 sample data were included, including 616 in the modeling group and 264 in the validation group. The 14 characteristic indexes screened by LASSO regression analysis were used for multivariate logistic regression analysis. The results showed that only globulin ($OR=1.112$, $95\%CI=1.044\sim1.185$), inorganic phosphorus ($OR=65.167$, $95\%CI=20.437\sim207.796$), total prostate specific antigen (tPSA) ($OR=1.026$, $95\%CI=1.014\sim1.037$) and serum uric acid ($OR=0.997$, $95\%CI=0.994\sim0.999$) had significant differences ($P<0.05$), and they were independent influencing factors for PCa. The calibration curve of the internal evaluation and internal verification of the PCa risk prediction model had high accuracy. The area under curve (AUC) of internal evaluation of the model was 0.766 ($95\%CI=0.728\sim0.804$), and the net benefit rate of decision curve analysis (DCA) of patients was 9%~72%; the AUC of internal validation of the model was 0.704 ($95\%CI=0.639\sim0.768$), and the net benefit rate of DCA of patients was 18%~59% and 63%~64%. **Conclusion:** Globulin, inorganic phosphorus, tPSA and serum uric acid are independent influencing factors of PCa. The risk prediction model constructed by them has a good prediction effect.

【Key words】prostatic cancer; prostatic hyperplasia; risk prediction model

作者介绍: 路 帅, Email: 1694161507@qq.com,

研究方向: 前列腺癌风险分析。

通信作者: 陆 进, Email: 0100197@bbmc.edu.cn。

基金项目: 2021年教育部产学研合作协同育人资助项目(编号: 202101160001)。

优先出版: <https://kns.cnki.net/kcms/detail/50.1046.R.20230227.1644.014.html>
(2023-02-28)

前列腺癌(prostatic cancer, PCa)是指发生在前列腺上皮的恶性肿瘤,是全球男性最常见的癌症,严重威胁老年男性健康^[1]。2020年世界癌症报告数据显示,PCa在男性恶性肿瘤中发病率居第6位,死亡率居第9位^[2]。虽然PCa在我国人口中的发病率相对较低,但随着我国经济水平不断提高和人均寿命延长以及饮食结构和生活习惯与西方的差异逐渐缩小,我国PCa的发病率不断上升^[3]。

由于前列腺解剖学位置较隐蔽,且PCa发病早期缺乏典型的临床表现,患者一经诊断已是中晚期,其中80%伴有骨转移^[4],从而失去根治机会。由于传统的PCa临床诊断指标血清前列腺特异性抗原(prostate specific antigen, PSA)特异性较低,在良性前列腺疾病与泌尿系统感染性疾病中表达水平异常升高^[5]。因此,PSA不能准确判断和评估患者的PCa进展和预后。临床上亟需寻找一些具有更高敏感性和特异性的生物标记物,且通过检测PCa特异性生化指标可减少不必要的穿刺活检给患者带来痛苦。本研究从PCa临床大数据出发,通过R语言算法,筛选PCa发病的独立影响因素,并构建风险预测模型,为早期诊断与治疗PCa提供可靠依据。

1 资料与方法

1.1 数据来源

对国家人口健康科学数据中心数据仓储平台临床医学科学数据中心(解放军总医院)《前列腺肿瘤预警数据集》进行回顾性病例对照研究,数据集中前列腺增生583例、前列腺癌303例、同时有前列腺癌和前列腺增生114例,数据集中包含球蛋白、总前列腺特异性抗原(total prostate specific antigen, tPSA)、游离前列腺特异性抗原(free prostate specific antigen, fPSA)、无机磷、血清尿酸等共33项信息。详细数据处理与分析流程图如图1所示。

1.2 数据处理

数据为真实病历数据,记录数据存在缺失值、异常值与记录不一致的问题,所以首先须对数据进行处理。将原始数据中的33项变量缺失比例 $\geq 25\%$ 的删除,每位患者的检查项目中缺失数 ≥ 5 项的删除,经过清洗后剩余25项变量。其中部分病例的生化指标中存在汉字、特殊字符、空格情况,将生化结果中汉字、特殊字符删除,空格数据用该字段变量均值进行替换。原始数据中前列腺预测结果包括3种,现将预测结果重新分类定义为有无PCa,其中前列腺增生归为无PCa组,前列腺癌与前列腺增生同时伴有前列腺癌归为PCa组。

1.3 数据整合

经过数据处理和数据插补后共删除120例样本数据,最终纳入的数据共880例,其中无PCa的患者511例,占58.1%;有PCa的患者369例,占41.9%,对其中23个变量进行赋值(表1)。

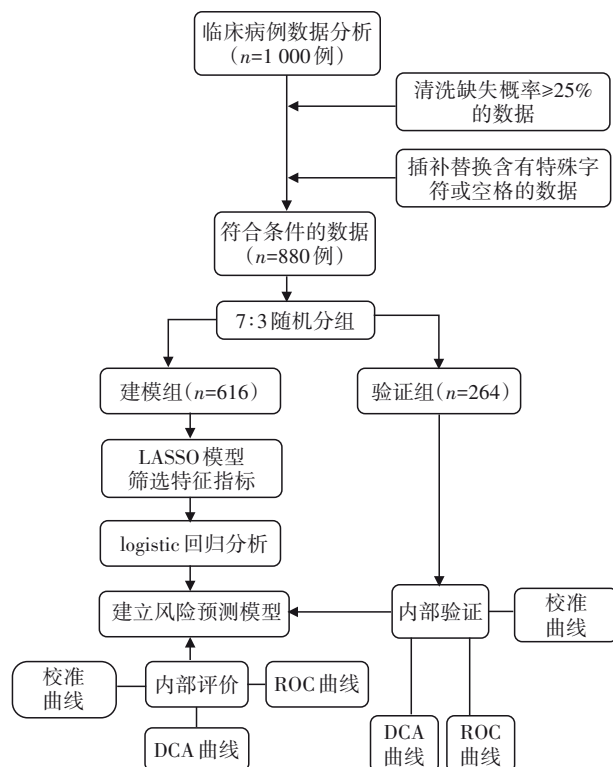


图1 研究流程图

表1 变量名及其赋值

分类	赋值
有无PCa	0=无PCa; 1=PCa
年龄/岁	1=(<45); 2=(45~60); 3=(≥ 60)
BMI/($\text{kg}\cdot\text{m}^{-2}$)	1=(<18.5); 2=(18.5~23.9); 3=(≥ 23.9)
球蛋白/($\text{g}\cdot\text{L}^{-1}$)	1=(<40); 2=(≥ 40)
碱性磷酸酶/($\text{U}\cdot\text{L}^{-1}$)	1=(<45); 2=(45~135); 3=(≥ 135)
肌酸激酶同工酶/($\text{U}\cdot\text{L}^{-1}$)	1=(<18); 2=(≥ 18)
fPSA/($\text{ng}\cdot\text{mL}^{-1}$)	1=(<0.18); 2=(0.18~0.27); 3=(0.27~0.43); 4=(≥ 0.43)
tPSA/($\text{ng}\cdot\text{mL}^{-1}$)	1=(<4); 2=(4~10); 3=(≥ 10)
钠/($\text{mmol}\cdot\text{L}^{-1}$)	1=(<135); 2=(135~145); 3=(≥ 145)
钙/($\text{mmol}\cdot\text{L}^{-1}$)	1=(<2.2); 2=(≥ 2.2)
氯化物/($\text{mmol}\cdot\text{L}^{-1}$)	1=(<95); 2=(95~105); 3=(≥ 105)
无机磷/($\text{mmol}\cdot\text{L}^{-1}$)	1=(<0.97); 2=(0.97~1.61); 3=(≥ 1.61)
游离钙/($\text{mmol}\cdot\text{L}^{-1}$)	1=(<1.1); 2=(≥ 1.1)
乳酸脱氢酶/($\text{U}\cdot\text{L}^{-1}$)	1=(<252); 2=(≥ 252)
肌酸激酶/($\text{U}\cdot\text{L}^{-1}$)	1=(<50); 2=(50~310); 3=(≥ 310)
肌酐/($\mu\text{mol}\cdot\text{L}^{-1}$)	1=(<54); 2=(54~133); 3=(≥ 133)
血清尿酸/($\mu\text{mol}\cdot\text{L}^{-1}$)	1=(<416); 2=(≥ 416)
甘油三酯/($\text{mmol}\cdot\text{L}^{-1}$)	1=(<0.45); 2=(0.45~1.69); 3=(≥ 1.69)
高密度脂蛋白胆固醇/ ($\text{mmol}\cdot\text{L}^{-1}$)	1=(<1.16); 2=(1.16~1.55); 3=(≥ 1.55)
低密度脂蛋白胆固醇/ ($\text{mmol}\cdot\text{L}^{-1}$)	1=(<2.6); 2=(2.6~3.4); 3=(≥ 3.4)
载脂蛋白A1/($\text{g}\cdot\text{L}^{-1}$)	1=(<1.2); 2=(1.2~1.6); 3=(≥ 1.6)
载脂蛋白B/($\text{g}\cdot\text{L}^{-1}$)	1=(<0.42); 2=(0.42~1.14); 3=(≥ 1.14)
钾/($\text{mmol}\cdot\text{L}^{-1}$)	1=(<3.5); 2=(3.5~5.5); 3=(≥ 5.5)

1.4 数据分组

首先利用 SPSS 25.0 对赋值后的 2 组研究对象进行差异性分析,结果显示年龄、球蛋白、碱性磷酸酶、肌酸激酶同工酶、tPSA、钙、无机磷、乳酸脱氢酶、肌酸激酶、血清尿酸、低密度脂蛋白胆固醇及载脂蛋白 B 等在 2 组之间的差异具有统计学意义 ($P<0.05$),见表 2。再按 7:3 随机分为建模组(616 例)和验证组(264 例)对 2 组数据进行差异分析,结果显示 2 组数据整体具有可比性,见表 3。

表 2 无 Pa 组和 Pa 组基线资料比较

变量	赋值	无 PCa 组 (n=511)	PCa 组 (n=369)	χ^2 值	P 值
年龄	1	0(0.0)	6(0.7)	11.771	0.003
	2	62(7.0)	60(6.9)		
	3	449(51.0)	303(34.4)		
BMI	1	17(1.9)	10(1.2)	3.260	0.196
	2	205(23.3)	128(14.5)		
	3	289(32.8)	231(26.3)		
球蛋白	1	233(26.5)	128(14.5)	10.539	0.001
	2	278(31.6)	241(27.4)		
碱性磷酸酶	1	38(4.3)	24(2.7)	20.148	<0.001
	2	469(53.3)	323(36.7)		
	3	4(0.5)	22(2.5)		
肌酸激酶同工酶	1	476(54.1)	324(36.8)	7.409	0.006
	2	35(4.0)	45(5.1)		
fPSA	1	24(2.7)	28(3.2)	4.922	0.178
	2	21(2.4)	17(1.9)		
	3	43(4.9)	22(2.5)		
tPSA	1	161(18.3)	94(10.7)	20.596	<0.001
	2	135(15.3)	64(7.3)		
	3	215(24.4)	211(24)		
钠	1	5(0.6)	4(0.4)	0.084	0.959
	2	414(47.0)	301(34.3)		
	3	92(10.5)	64(7.2)		
钙	1	236(26.8)	131(14.9)	10.058	0.002
	2	275(31.3)	238(27.0)		
氯化物	1	4(0.4)	2(0.3)	0.343	0.842
	2	196(22.3)	137(15.5)		
	3	311(35.3)	230(26.2)		
无机磷	1	101(11.5)	41(4.6)	17.122	<0.001
	2	408(46.4)	320(36.3)		
	3	2(0.2)	8(1.0)		
游离钙	1	132(15.0)	86(9.8)	0.733	0.392
	2	379(43.1)	283(32.1)		
乳酸脱氢酶	1	505(57.4)	356(40.4)	5.596	0.018
	2	6(0.7)	13(1.5)		
肌酸激酶	1	96(10.9)	43(4.9)	8.700	0.013
	2	412(46.8)	322(36.6)		
	3	3(0.3)	4(0.5)		
肌酐	1	10(1.1)	10(1.2)	3.986	0.136
	2	474(53.9)	349(39.6)		
	3	27(3.1)	10(1.1)		
血清尿酸	1	425(48.3)	325(36.9)	4.096	0.043
	2	86(9.8)	44(5.0)		

续表 2

变量	赋值	无 PCa 组 (n=511)	PCa 组 (n=369)	χ^2 值	P 值
甘油三酯	1	6(0.7)	2(0.2)	5.504	0.064
	2	415(47.2)	280(31.8)		
	3	90(10.2)	87(9.9)		
高密度脂蛋白胆固醇	1	257(29.2)	184(20.9)	0.018	0.991
	2	189(21.5)	138(15.7)		
	3	65(7.4)	47(5.3)		
低密度脂蛋白胆固醇	1	270(30.7)	159(18.1)	9.632	0.008
	2	170(19.3)	137(15.6)		
	3	71(8.1)	73(8.2)		
载脂蛋白 A1	1	248(28.2)	153(17.4)	4.335	0.114
	2	230(26.1)	188(21.4)		
	3	33(3.8)	28(3.1)		
载脂蛋白 B	1	4(0.5)	1(0.1)	13.804	0.001
	2	484(55.0)	328(37.2)		
	3	23(2.6)	40(4.6)		
钾	1	28(3.2)	19(2.2)	1.429	0.490
	2	483(54.9)	349(39.6)		
	3	0(0.0)	1(0.1)		

表 3 建模组与验证组的基线资料比较

变量	赋值	建模组 (n=616)	验证组 (n=264)	χ^2 值	P 值
年龄	1	3(0.3)	3(0.4)	2.557	0.278
	2	80(9.0)	42(4.9)		
	3	533(60.5)	219(24.9)		
BMI	1	21(2.4)	6(0.7)	1.862	0.394
	2	239(27.2)	94(10.6)		
	3	356(40.5)	164(18.6)		
球蛋白	1	246(27.9)	115(13.1)	1.004	0.316
	2	370(42.1)	149(16.9)		
碱性磷酸酶	1	45(5.1)	17(1.9)	1.085	0.581
	2	555(63.1)	237(26.9)		
	3	16(1.8)	10(1.2)		
肌酸激酶同工酶	1	559(63.5)	241(27.4)	0.065	0.798
	2	57(6.5)	23(2.6)		
fPSA	1	33(3.7)	19(2.2)	2.687	0.442
	2	30(3.4)	8(0.9)		
	3	44(5.0)	21(2.4)		
tPSA	1	509(57.8)	216(24.6)	0.339	0.844
	2	180(20.5)	75(8.5)		
	3	136(15.5)	63(7.1)		
钠	1	300(34.1)	126(14.3)	1.568	0.457
	2	7(0.8)	2(0.2)		
	3	494(56.1)	221(25.2)		
钙	1	115(13.1)	41(4.6)	1.843	0.175
	2	266(30.2)	101(11.5)		
	3	350(39.8)	163(18.5)		
氯化物	1	5(0.6)	1(0.1)	4.908	0.086
	2	219(24.9)	114(12.9)		
	3	392(44.6)	149(16.9)		

续表 3

变量	赋值	建模组 (n=616)	验证组 (n=264)	χ^2 值	P 值
无机磷	1	108(12.3)	34(3.8)	2.964	0.227
	2	501(56.9)	227(25.8)		
	3	7(0.8)	3(0.4)		
游离钙	1	158(18.0)	60(6.8)	0.847	0.357
	2	458(52.0)	204(23.2)		
乳酸脱氢酶	1	602(68.4)	259(29.4)	0.126	0.723
	2	14(1.6)	5(0.6)		
肌酸激酶	1	95(10.8)	44(5.0)	0.219	0.896
	2	516(58.6)	218(24.8)		
	3	5(0.6)	2(0.2)		
肌酐	1	16(1.8)	4(0.5)	0.979	0.613
	2	574(65.2)	249(28.3)		
	3	26(3.0)	11(1.2)		
血清尿酸	1	531(60.3)	219(24.9)	1.547	0.214
	2	85(9.7)	45(5.1)		
甘油三酯	1	8(0.9)	0(0.0)	5.315	0.070
	2	492(55.9)	203(23.1)		
	3	116(13.2)	61(6.9)		
高密度脂蛋白胆固醇	1	304(34.5)	137(15.6)	0.807	0.668
	2	230(26.1)	97(11.1)		
	3	82(9.3)	30(3.4)		
低密度脂蛋白胆固醇	1	294(33.4)	135(15.4)	1.984	0.371
	2	224(25.5)	83(9.4)		
	3	98(11.1)	46(5.2)		
载脂蛋白 A1	1	273(31.0)	128(14.6)	1.775	0.412
	2	297(33.8)	121(13.7)		
	3	46(5.2)	15(1.7)		
载脂蛋白 B	1	4(0.5)	1(0.1)	1.004	0.605
	2	571(64.9)	241(27.3)		
	3	41(4.7)	22(2.5)		
钾	1	35(4.0)	12(1.4)	0.908	0.635
	2	580(65.9)	252(28.6)		
	3	1(0.1)	0(0.0)		

1.5 特征指标筛选

使用最小绝对值收敛和选择算子(least absolute shrinkage and selection operator, LASSO)对建模组数据进行分析。LASSO 回归分析利用 R 语言的“glmnet”包进行,通过十折交叉验证得到最佳调整参数 λ_{1se} ,当 λ 取值为 λ_{1se} 即 0.099 时得到特征指标。

1.6 logistic 回归分析

利用 SPSS 25.0 对 LASSO 回归筛选得到的特征指标进行多因素 logistic 回归分析,筛选 PCa 的独立影响因素进行下一步风险预测模型构建,筛选标准为 $P < 0.05$ 。

1.7 风险预测模型的构建与验证

利用多因素 logistic 回归分析筛选得到的独立影响因素,通过 R 语言的“rms”包对建模组数据构建 PCa 风险预测模型,并绘制决策曲线分析(decision curve analysis, DCA)、

校准曲线以及利用 SPSS 25.0 绘制受试者工作特征(receiver operating characteristic, ROC)曲线对风险预测模型进行内部评价,同时利用验证组数据进行内部验证。

1.8 统计学处理

利用 SPSS 25.0 软件,对原始数据中缺失比例 $\geq 25\%$ 的变量及检查项目中缺失数 ≥ 5 的样本进行删除;再利用均值对缺失的变量进行填补;然后,对变量进行赋值,并通过卡方检验对数据进行描述及组间差异分析;通过 R 语言进行 LASSO 回归筛选影响因素,再采用多因素 logistic 回归分析筛选独立影响因素,并使用其构建 PCa 风险列线图预测模型,同时对建模组和验证组数据绘制 DCA 曲线、ROC 曲线及校正曲线,对风险预测模型进行内部评价及内部验证。检验水准 $\alpha = 0.05$ 。

2 结果

2.1 LASSO 回归分析结果

利用 LASSO 回归从特征变量中进行筛选,通过十折交叉验证得到最佳调整参数 λ_{1se} ,当 λ 取值为 λ_{1se} 即 0.099 时得到 14 个最精简的模型系数,分别为年龄、BMI、球蛋白、碱性磷酸酶、肌酸激酶同工酶、tPSA、无机磷、游离钙、肌酸激酶、血清尿酸、低密度脂蛋白胆固醇、载脂蛋白 A1、载脂蛋白 B 和钾(图 2)。

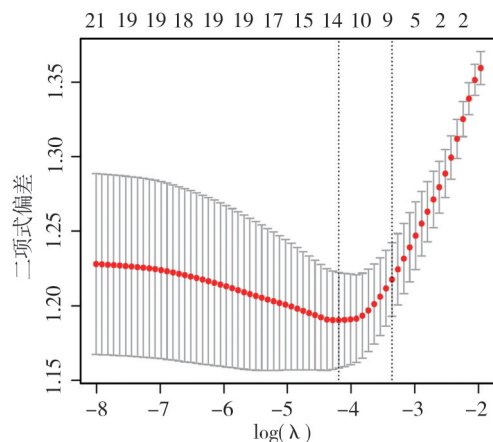


图 2 LOSSA 回归变量筛选图

2.2 logistic 回归模型分析

利用 616 例建模组数据,将有无 PCa 作为因变量,将上述筛选出的 14 个特征指标作为自变量,进行多因素 logistic 回归分析,结果显示球蛋白($OR = 1.112, 95\%CI = 1.044 \sim 1.185$)、无机磷($OR = 65.167, 95\%CI = 20.437 \sim 207.796$)、tPSA($OR = 1.026, 95\%CI = 1.014 \sim 1.037$)与血清尿酸($OR = 0.997, 95\%CI = 0.994 \sim 0.999$)的差异具有统计学意义($P < 0.05$),可作为 PCa 的独立影响因素(表 4)。

2.3 构建风险预测模型

利用多因素 logistic 回归分析得到的独立影响因素构建风险列线图预测模型,指标包括球蛋白、无机磷、tPSA 与血

表 4 多因素 logistic 回归分析

变量	β	标准误差	瓦尔德	自由度	P	OR	95%CI	
							下限	上限
年龄	0.018	0.013	2.140	1	0.143	1.019	0.994	1.044
BMI	0.031	0.031	1.015	1	0.314	1.032	0.971	1.097
球蛋白	0.106	0.032	10.713	1	0.001	1.112	1.044	1.185
碱性磷酸酶	0.005	0.003	3.005	1	0.083	1.005	0.999	1.011
肌酸激酶同工酶	0.030	0.016	3.633	1	0.057	1.031	0.999	1.064
tPSA	0.025	0.006	19.682	1	0.000	1.026	1.014	1.037
无机磷	4.177	0.592	49.842	1	0.000	65.167	20.437	207.796
游离钙	3.177	1.844	2.970	1	0.085	23.981	0.646	889.573
肌酸激酶	0.001	0.002	0.179	1	0.672	1.001	0.998	1.004
血清尿酸	-0.003	0.001	9.285	1	0.002	0.997	0.994	0.999
低密度脂蛋白胆固醇	0.150	0.257	0.338	1	0.561	1.161	0.701	1.923
载脂蛋白 A1	0.603	0.447	1.819	1	0.177	1.828	0.761	4.395
载脂蛋白 B	0.572	0.977	0.343	1	0.558	1.772	0.261	12.035
钾	-0.418	0.290	2.071	1	0.150	0.659	0.373	1.163

清尿酸。图中每个指标都对应至上方分数标尺的分值,每位患者可以根据实际情况,读出各项指标对应分值,通过相加得到总分数,然后对应下方的PCa发病风险概率(图3)。

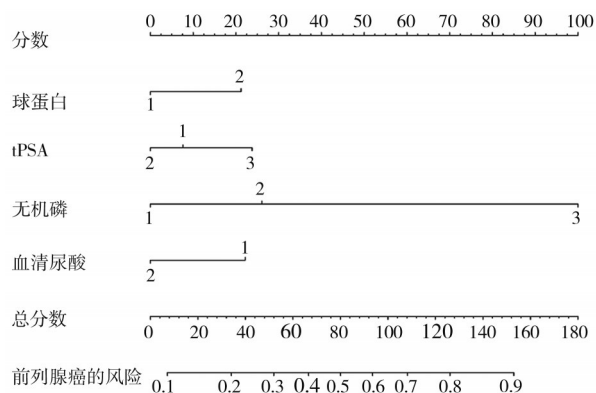


图3 预测模型的列线图

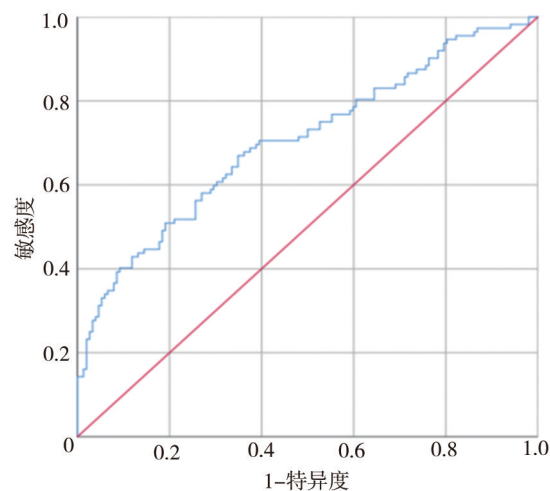
2.4 风险预测模型的内部评价与内部验证

分别利用ROC、DCA及校准曲线对建模组数据构建的风险预测模型进行内部评价;同时利用验证组数据对模型进行内部验证。

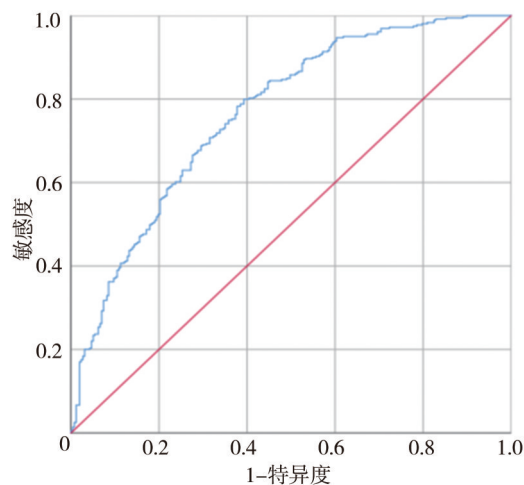
2.4.1 ROC曲线 ROC分析结果显示,建模组的曲线下面积(area under curve, AUC)为0.766(95%CI=0.728~0.804);验证组的AUC为0.704(95%CI=0.639~0.768)。2组的ROC显示了较好的区分能力,表明风险预测模型对PCa与非PCa患者具有很好的区分作用(图4)。

2.4.2 DCA曲线 DCA分析结果显示,建模组患者的阈概率为9%~72%,而验证组患者的阈概率为18%~59%及63%~64%;所以模型具有一定的临床应用价值(图5)。

2.4.3 校准曲线 校准曲线分析结果显示,2组的校准曲线与预测曲线拟合程度较高,且与理想曲线偏差较小,说明预测模型能够准确地预测患PCa风险的概率(图6)。



A. 建模组



B. 验证组

图4 建模组和验证组的ROC曲线

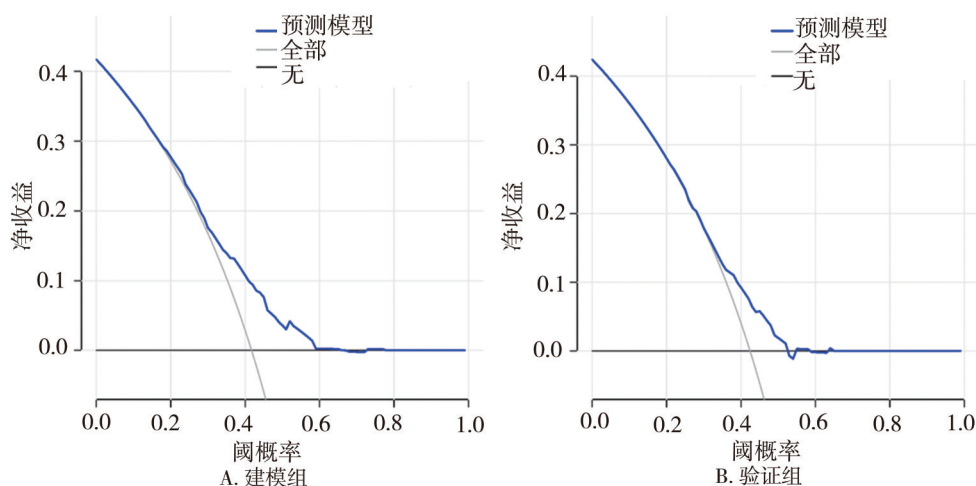


图5 建模组和验证组的DCA曲线

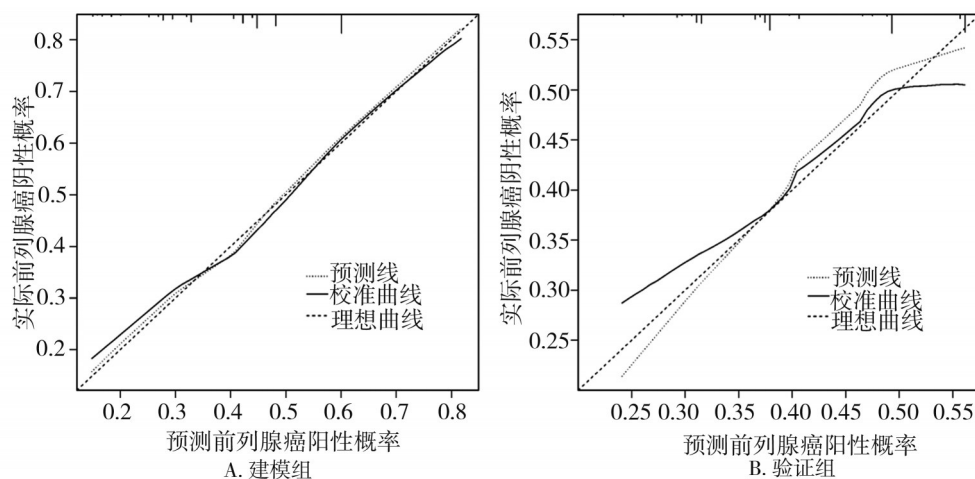


图6 建模组和验证组的校准曲线

3 讨论

PCa的主要危险因素包括性行为、生理、环境以及遗传等因素^[6],随着生活方式的改变和预期寿命的延长,我国PCa的发病率呈现上升趋势。然而,传统的生化指标血清PSA用于临床诊断PCa特异性较低,大部分PCa患者早期诊断困难,确诊时已至中晚期,从而失去根治的机会,导致PCa患者生存预后不良。因此,临床上迫切需要寻找一种高效且准确的生物标记物。而风险预测模型在预测疾病发病风险与预后方面具有独特的临床应用价值。

本研究利用LASSO回归进行特征指标筛选,对特征指标进行logistic回归分析后,筛选PCa独立影响因素,并使用R语言构建PCa风险预测模型,通过内部评价和内部验证后,提示风险预测模型对患PCa的风险有良好的预测作用。其中球蛋白、tPSA与无机磷的表达量与PCa的发病风险呈正相关,而

血清尿酸的表达量与PCa的发病风险呈负相关。

PSA是一种由前列腺上皮细胞分泌的蛋白酶,是目前临床上用于诊断PCa的首选标志物^[7]。PSA在外周血中存在结合型和游离型2种形式,两者合称tPSA。正常情况下,tPSA在精液中的浓度较高,在外周血中的浓度较低,当发生PCa时会破坏前列腺屏障导致进入外周血的tPSA增多,但是其他因素同样会导致进入血液的tPSA增多,如前列腺炎、前列腺增生、前列腺穿刺活检等,提示tPSA单独作为诊断PCa的标志物特异性并不高^[8-9]。近年研究表明,tPSA联合其他检测手段能够有效提高PCa诊断的灵敏性和特异性^[10-13]。此外,还有研究发现分级越高的PCa患者外周血中的tPSA越高,因此tPSA可为鉴别不同级别的PCa提供临床线索^[14]。

尿酸是体内DNA的黄嘌呤和RNA的次黄嘌呤在黄嘌呤氧化还原酶的催化作用下产生的代谢产物。尿酸在体内的产生和排出于一种平衡状态,当这种平衡被打破时,就可能导致高尿酸血症和痛

风。癌症是细胞异常增殖的疾病,往往伴随着异常代谢模式,其中嘌呤代谢也会增高,因此推测尿酸也会随之变化。尿酸在体内具有抗氧化作用,能够通过抑制自由基的产生而对 PCa 起一定的保护作用^[15]。另有研究表明,尿酸在细胞内发挥促炎症效应^[16],在 PCa 的发生机制中发挥重要作用。尿酸与癌症的发生发展关系复杂,以往国内外对它们的研究结论不一,甚至互相矛盾,因此需要更深入的研究去探索两者之间的关系。

球蛋白是存在于人体血清中的一类具有免疫功能的蛋白,其中 β_2 微球蛋白主要由淋巴细胞产生。有研究发现, β_2 微球蛋白与癌症的病理进程密切相关^[17]。而且 β_2 微球蛋白在骨代谢中起重要的调控作用,与 PCa 晚期骨转移密切相关。此外,炎症反应在 PCa 发生发展过程中的作用已得到广泛认可,白蛋白与球蛋白的比值可以较灵敏地反映机体的慢性炎症反应,因此两者比值可预测晚期 PCa 和其他癌症患者的预后^[18-19]。

人体中大部分磷集中于骨骼和牙齿,只有少部分分布于血液,而血磷增加极可能与骨损伤有关,当恶性肿瘤骨转移时骨质遭到破坏,导致血磷异常升高,而骨转移是 PCa 晚期常见的并发症,因此血磷的检测对 PCa 的治疗与预后具有重要临床意义。另有研究发现,高磷饮食可能增加患 PCa 的风险^[20],可与本文结果相互验证。

综上所述,本文通过对前列腺肿瘤数据集深入挖掘,成功构建了 PCa 的风险预测模型,并且通过验证显示了其良好的预测作用,同时筛选出 PCa 的独立影响因素,对 PCa 的诊断具有一定的临床意义。但是,由于本文为回顾性研究且处理原始数据方法不同以及缺少外部验证的缘故,可能会在一定程度上影响结论。因此,该结论能否应用于临床还有待进一步验证。

参 考 文 献

- [1] 李 斌,刘 旭,王洪颖,等. 小分子多肽抗前列腺癌的研究进展[J]. 中国实验方剂学杂志,2021,27(10):222-230.
- Li B, Liu X, Wang HY, et al. Research progress of polypeptides against prostate cancer[J]. Chin J Exp Tradit Med Formulae, 2021, 27(10): 222-230.
- [2] World cancer report: cancer research for cancer prevention[EB/OL]. (2020-06-20). <https://publications.iarc.fr/586>.
- [3] 前列腺癌筛查中国专家共识(2021 年版)[J]. 中国癌症杂志, 2021,31(5):435-440.
- China expert consensus on prostate cancer screening (2021 edition)[J]. China Oncol, 2021, 31(5):435-440.
- [4] Dyrberg E, Hendel HW, Huynh THV, et al. ^{68}Ga -PSMA-PET/CT in comparison with ^{18}F -fluoride-PET/CT and whole-body MRI for the detection of bone metastases in patients with prostate cancer: a prospective diagnostic accuracy study[J]. Eur Radiol, 2019, 29(3):1221-1230.
- [5] 王茹月,叶 雨. 前列腺癌早期诊断相关外周血循环游离核酸

的研究进展[J]. 实用医学杂志,2019,35(5):831-834.

Wang RY, Ye Y. Research progress of free nucleic acid in peripheral blood circulation related to early diagnosis of prostate cancer[J]. J Pract Med, 2019, 35(5):831-834.

[6] 孙殿钦,雷 林,蔡 颖,等. 前列腺癌相关危险因素的研究进展[J]. 中国肿瘤,2020,29(4):292-298.

Sun DQ, Lei L, Cai Y, et al. Research advances in risk factors for prostate cancer[J]. China Cancer, 2020, 29(4):292-298.

[7] Kohaar I, Petrovics G, Srivastava S. A rich array of prostate cancer molecular biomarkers: opportunities and challenges[J]. Int J Mol Sci, 2019, 20(8):1813.

[8] Hashimoto T, Ohori M, Shimodaira K, et al. Prostate-specific antigen screening impacts on biochemical recurrence in patients with clinically localized prostate cancer[J]. Int J Urol, 2018, 25(6):561-567.

[9] Duffy MJ. Biomarkers for prostate cancer: prostate-specific antigen and beyond[J]. Clin Chem Lab Med, 2020, 58(3):326-339.

[10] Stefancu A, Moisoiu V, Couti R, et al. Combining SERS analysis of serum with PSA levels for improving the detection of prostate cancer[J]. Nanomedicine(Lond), 2018, 13(19):2455-2467.

[11] 刘加成,刘 群,管荣平,等. T1、T2 值联合血清总前列腺特异性抗原对移行带前列腺癌的诊断价值[J]. 中国临床医学影像杂志, 2022,33(8):546-550,556.

Liu JC, Liu Q, Guan RP, et al. T1, T2 value combined with serum tPSA in diagnosis of prostate cancer in transition zone[J]. J China Clin Med Imaging, 2022, 33(8):546-550,556.

[12] Mello-Grand M, Gregnanin I, Sacchetto L, et al. Circulating microRNAs combined with PSA for accurate and non-invasive prostate cancer detection[J]. Carcinogenesis, 2019, 40(2):246-253.

[13] Saout K, Zambon A, Nguyen TA, et al. Impact of multiparametric MRI and PSA density on the initial indication or the maintaining in active surveillance during follow-up in low-risk prostate cancer[J]. Clin Genitourin Cancer, 2022, 20(3):e244-e252.

[14] 红 华,吴国柱,刘举珍,等. 不同级别前列腺癌经直肠超声造影参数分析[J]. 中国超声医学杂志,2020,36(2):170-174.

Hong H, Wu GZ, Liu JZ, et al. Analysis of transrectal contrast-enhanced ultrasound parameters in different grades of prostate cancer[J]. Chin J Ultrasound Med, 2020, 36(2):170-174.

[15] Strasak AM, Rapp K, Hilbe W, et al. The role of serum uric acid as an antioxidant protecting against cancer: prospective study in more than 28 000 older Austrian women[J]. Ann Oncol, 2007, 18(11):1893-1897.

[16] Fabbrini E, Serafini M, Colic Baric I, et al. Effect of plasma uric acid on antioxidant capacity, oxidative stress, and insulin sensitivity in obese subjects[J]. Diabetes, 2014, 63(3):976-981.

[17] Teasdale C, Mander AM, Fifield R, et al. Serum β_2 -microglobulin in controls and cancer patients[J]. Clin Chimica Acta, 1977, 78(1):135-143.

[18] Bozkaya Y, Erdem GU, Demirci NS, et al. Prognostic importance of the albumin to globulin ratio in metastatic gastric cancer patients[J]. Curr Med Res Opin, 2019, 35(2):275-282.

[19] 张华锋,赵 佳,徐道明,等. 全雄激素阻断治疗前前列腺癌患者白蛋白球蛋白比值在预后评估中的研究[J]. 中华老年医学杂志, 2019,38(10):1153-1157.

Zhang HF, Zhao J, Xu DM, et al. Prognostic value of pretreatment albumin to globulin ratio in prostate cancer patients treated with maximal androgen blockade[J]. Chin J Geriatr, 2019, 38(10):1153-1157.

[20] Wilson KM, Shui IM, Mucci LA, et al. Calcium and phosphorus intake and prostate cancer risk: a 24-y follow-up study[J]. Am J Clin Nutr, 2015, 101(1):173-183.

(责任编辑:冉明会)