

男性乳腺癌死亡风险预测模型的构建及外部验证研究

田林¹, 朱荔¹, 张丽萍², 孙嘉鸿¹, 王建东¹

¹解放军总医院第一医学中心普通外科医学部, 北京 100853; ²中南大学湘雅医院普通外科, 湖南长沙 410008

摘要: **背景** 男性乳腺癌(male breast cancer, MBC)在临床上罕见, 但其年发病率近年来有所增加, 且预后不佳。目的 构建预测MBC患者预后的临床预测模型并进行外部验证。**方法** 选取1990—2019年SEER数据库中MBC患者, 按照7:3的比例随机分为训练集和验证集, 以5年病死率作为结局指标, 使用Lasso回归、单因素、多因素Cox比例模型筛选相关预测特征, 采用ROC曲线、校准曲线评估MBC临床预测模型的区分度和校准度, 采用列线图对临床预测模型进行可视化, 并以解放军总医院第一医学中心的病例进行外部验证。**结果** 本研究模型构建纳入SEER数据库中1 614例患者, 将其分为训练集(n=1 129)与验证集(n=485), 随访期间训练集和验证集中分别有274例和111例患者死亡。LASSO回归筛选出9个与死亡风险相关联的临床特征——诊断时年龄、组织学分级、T分期、N分期、M分期、骨转移、是否接受手术、ER状态和PR状态, 多因素Cox风险比例模型分析确定这9个指标与死亡独立关联($P<0.05$), 基于此构建的预测模型在训练集预测1年、3年、5年死亡风险的ROC曲线下面积分别是0.895、0.853和0.829, C-index值为0.811; 内部验证集1年、3年、5年死亡风险预测模型的ROC曲线下面积分别为0.858、0.858、0.848, 验证集的C-index值为0.798。对解放军总医院第一医学中心的44例MBC患者进行外部验证, 5年死亡风险的AUC为0.731, 10年死亡风险的AUC为0.797。**结论** 本研究建立了基于诊断年龄、组织学分级、T分期、N分期、M分期、骨转移状态、手术干预状态、ER状态和PR状态预测模型, 对MBC患者死亡风险具有较高预测价值, 为临床医师精准评估提供了重要参考。

关键词: 男性; 乳腺癌; 临床预测模型; 机器学习; 列线图; 预后; 死亡

中图分类号: R737.9; R655.8

文献标志码: A

文章编号: 2095-5227(2025)08-0784-08

DOI: 10.12435/j.issn.2095-5227.24122502

引用本文: 田林, 朱荔, 张丽萍, 等. 男性乳腺癌死亡风险预测模型的构建及外部验证研究 [J]. 解放军医学院学报, 2025, 46 (8): 784-791.

Construction and external validation of a mortality prediction model for male breast cancer

TIAN Lin¹, ZHU Li¹, ZHANG Liping², SUN Jiahong¹, WANG Jiandong¹

¹Senior Department of General Surgery, the First Medical Center of PLA General Hospital, Beijing 100853, China; ²Department of General Surgery, XiangYa Hospital of Central South University, Changsha 410008, Hunan Province, China

Corresponding author: WANG Jiandong. Email: Vicky1968@163.com

Abstract: Background Male breast cancer (MBC) is clinically rare, but its annual incidence has increased in recent years and the prognosis is poor. **Objective** To develop a clinical prediction model for prognosis in male breast cancer patients and externally validate it using data from MBC patients at our institution. **Methods** Patients with MBC from the SEER database from 1990 to 2019 were selected and randomly divided into a training set and a validation set at a ratio of 7:3. The overall 5-year death of MBC patients was used as the outcome indicator. Lasso regression, univariate, and multivariate Cox proportional hazards models were employed to screen for predictive features. The discriminative ability and calibration of the clinical prediction model were evaluated using receiver operating characteristic (ROC) curves, calibration curves, and external validation analysis. A nomogram was constructed to visualize the clinical prediction model. **Results** This study included 1 614 patients from the SEER database as the study subjects and divided them into a training set (n=1 129) and a validation set (n=485). Follow-up results showed that 274 died in the training set and 111 patients died in the validation set. LASSO regression identified 9 clinical features: age at diagnosis, histological grade, T stage, N stage, M stage, bone metastasis status, surgical intervention status, estrogen receptor (ER) status, and progesterone receptor (PR) status. Univariate Cox proportional hazards model analysis revealed that all variables were significantly associated with risk of death in MBC ($P<0.05$). Multivariate Cox proportional hazards model analysis demonstrated that age, tumor grade, T stage, N stage, M stage, and bone metastasis were independent risk factors for risk of death. At the same time, surgical intervention, ER positivity, and PR positivity served as protective factors ($P<0.05$). In the development set, the AUC values for the 1-, 3-, and 5-year risk of death prediction models were 0.895, 0.853, and 0.829, respectively, with a C-index of 0.811. In the

收稿日期: 2024-12-25

第一作者: 田林, 在读硕士, 医师。Email: tl13595255679@163.com

通信作者: 王建东, 主任医师, 教授, 博士生导师。Email: Vicky1968@163.com

validation set, the corresponding AUC values were 0.858, 0.858, and 0.848, with a C-index of 0.798. External validation was performed on 44 patients with MBC in our hospital. The AUC for 5-year risk of death was 0.731, and the 10-year risk of death was 0.797. **Conclusion** This study establishes a prediction model based on diagnostic age, histological grading, T staging, N staging, M staging, bone metastasis status, surgical intervention status, ER status, and PR status, which has high predictive value for risk of death in MBC patients and provides a reliable tool for clinical doctors.

Keywords: male; breast cancer; clinical prediction model; machine learning; nomogram; prognosis; death

Cited as: Tian L, Zhu L, Zhang LP, et al. Construction and external validation of a mortality prediction model for male breast cancer[J]. Acad J Chin PLA Med Sch, 2025, 46(8): 784-791.

男性乳腺癌(male breast cancer, MBC)在临床上很罕见, 约占所有乳腺癌病例的1%, 近年来其年发病率有所增加^[1-2]。由于MBC的发病率远低于女性乳腺癌, 大多数乳腺癌临床研究只包括女性, 几乎没有前瞻性数据可指导MBC的临床治疗。尽管近年来乳腺癌患者的死亡率有所下降, 但远处转移的乳腺癌患者预后仍较差, 5年死亡率为73%^[3]。一些研究表明, 与女性相比, MBC患者的预后更差, 这可能是由于诊断时期较晚、诊断时年龄较大, 或是不良预后的亚型, 如三阴性乳腺癌^[4-7]。因此, 开发一种用于预测MBC患者死亡风险的工具/模型, 将有助于提高对癌症预防的认识。在过去几年中, 医学领域广泛利用机器学习进行多种预测, 如癌症发病率预测^[8]、癌症检测^[9]、癌症死亡预测^[10]和骨转移风险预测^[11]。本研究旨在构建疾病预测模型以预测MBC患者的死亡风险, 并分析该模型的预测能力, 为MBC进行危险分层提供依据, 以指导个体化治疗。

1 对象与方法

1.1 数据来源

建模样本提取自美国国立癌症研究所监测、流行病学及结局项目数据库(Surveillance, Epidemiology and End Results Program, SEER)。该数据库是公认的癌症信息的权威信息源库, 覆盖了美国约34.6%的人口。基于估计发生第二原发癌症(second primary malignancy, SPM)的风险, 本研究使用SEER*Stat 8.4.1版本。采集过程中, 以国际肿瘤学疾病分类第3版(ICD-O-3)的专用站点代码C50(包括C50.1~C50.9)来识别/标记乳腺恶性肿瘤。设置3个关键变量——诊断年份、序列号、患者原位/恶性肿瘤总数, 以确定SPM的状态。

收集1990—2019年的MBC患者, 纳入标准:(1)男性;(2)诊断为乳腺癌;(3)至少具备5年随访数据。排除标准:(1)初诊2个月又诊断为SPM同步癌症;(2)第一原发恶性肿瘤并非乳腺恶性肿瘤。

外部验证集样本来自2001—2019年解放军总医院第一医学中心的MBC患者。纳入及排除标准同建模数据集。本研究在入组前已获得全部受试者的知情同意, 并经解放军总医院第一医学中心医学伦理委员会批准(伦审第S2024-202-01号)。

1.2 研究方法

建模样本被随机分为训练队列(70%)和内部验证队列(30%), 用于开发和验证预后预测风险模型。另外收集解放军总医院第一医学中心2001—2019年的MBC患者用于外部验证。

本研究选择的结局指标是1年、3年、5年死亡风险。根据乳腺癌NCCN指南^[12], 结合SEER数据库, 筛选可能对MBC患者预后产生影响的临床指标, 包括年龄、肿瘤位置、组织学分级、雌激素受体(estrogen receptor, ER)、孕激素受体(progesterone receptor, PR)、HER2、侵犯神经、脉管浸润、淋巴结阳性个数、Ki-67、淋巴结转移情况和病理分期。使用Lasso回归、单因素、多因素Cox比例模型筛选相关预测特征。

SEER数据库将组织学类型分为浸润性导管癌、腺癌、其他。激素受体(hormone receptor, HR)状态分为HR阳性[ER和(或)PR阳性]和HR阴性(ER和PR均为阴性)。补充分析中, SEER数据库将TMN分期调整为第6版AJCC分期, 记录SPM的部位和诊断时间。

1.3 统计学方法

采用SPSS 27.0、R 4.1.3软件、Python 3.9.7和MedCalc 19.0.7进行数据分析。分类变量以例数(百分比)表示, 连续变量以 $M(IQR)$ 表示。使用非参数检验分析组间连续变量差异, χ^2 检验分析分类变量组间差异。采用Lasso回归、单因素和多因素Cox回归筛选模型预测特征。根据训练集中多变量Cox回归分析的结果构建了诺莫图。使用训练集、验证集和外部验证集的AUC和一致性指数(C-index)评估区分度。校准通过校准曲线进行评估。以 $P < 0.05$ 为差异有统计学意义。

2 结果

2.1 样本一般资料(训练集 vs 验证集)

SEER数据库单原发病灶的男性乳腺癌病例共计9 470例,死亡原因不明2 685例,临床病理资料缺失5 032例,乳腺癌分期缺失133例,生存时间缺失或者为0者6例,最终1 614例纳入后续分析。

1 614例MBC患者按7:3的比例,随机分为训练集(n=1 129)和验证集(n=485)。两组年龄、婚姻状态、种族、初始肿瘤大小、肿瘤等级、肿瘤T分期、肿瘤N分期、肿瘤M分期、骨转移情况、肝转移情况、肺转移情况、肿瘤治疗方法、肿瘤亚分期、ER、HER2等差异无统计学意义($P>0.05$)。见表1。

表1 训练集和验证集均衡性检验

Tab. 1 Balance check between training set and testing set

临床特征	总人数 (n=1 614)	训练集 (n=1 129)	验证集 (n=485)	Z/ χ^2 值	P值	临床特征	总人数 (n=1 614)	训练集 (n=1 129)	验证集 (n=485)	Z/ χ^2 值	P值
中位年龄/ [岁, $M(IQR)$]	65.00 (57.00 ~ 72.00)	65.00 (57.00 ~ 72.00)	66.00 (57.00 ~ 72.00)	0.805	0.421	骨转移/(例, %)				0.018	0.894
生存时间/ [月, $M(IQR)$]	92.00 (73.00 ~ 115.00)	92.00 (74.00 ~ 114.00)	91.00 (72.00 ~ 116.00)	0.202	0.840	无/未知	1 526(94.55)	1 068(94.60)	458(94.43)		
婚姻状况/(例, %)				1.827	0.401	有	88(5.45)	61(5.40)	27(5.57)		
已婚	1 086(67.29)	748(66.25)	338(69.69)			肝转移/(例, %)				<0.001	0.997
未婚	296(18.34)	214(18.95)	82(16.91)			无/未知	1 599(99.07)	1 118(99.03)	481(99.18)		
其他	232(14.37)	167(14.79)	65(13.40)			有	15(0.93)	11(0.97)	4(0.82)		
种族/(例, %)				2.056	0.358	肺转移/(例, %)				2.180	0.140
白种人	201(12.45)	140(12.40)	61(12.58)			无/未知	1 569(97.21)	1 102(97.61)	467(96.29)		
黑种人	99(6.13)	63(5.58)	36(7.42)			有	45(2.79)	27(2.39)	18(3.71)		
其他	1 314(81.41)	926(82.02)	388(80.00)			手术/(例, %)				2.000	0.157
肿瘤位置/(例, %)				0.799	0.850	无/未知	99(6.13)	63(5.58)	36(7.42)		
中央区	671(41.57)	476(42.16)	195(40.21)			有	1 515(93.87)	1 066(94.42)	449(92.58)		
乳房外象限	195(12.08)	133(11.78)	62(12.78)			放疗/(例, %)				0.631	0.427
乳房内象限	283(17.53)	199(17.63)	84(17.32)			无/未知	1 119(69.33)	776(68.73)	343(70.72)		
其他	465(28.81)	321(28.43)	144(29.69)			有	495(30.67)	353(31.27)	142(29.28)		
组织学分级/(例, %)				0.024	0.999	化疗/(例, %)				2.258	0.133
I	194(12.02)	136(12.05)	58(11.96)			无/未知	896(55.51)	613(54.30)	283(58.35)		
II	785(48.64)	550(48.72)	235(48.45)			有	718(44.49)	516(45.70)	202(41.65)		
III	527(32.65)	368(32.60)	159(32.78)			肿瘤亚型/(例, %)				1.339	0.512
未知	108(6.69)	75(6.64)	33(6.80)			HR+/HER2-	188(11.65)	129(11.43)	59(12.16)		
T分期/(例, %)				2.754	0.431	HER2+	34(2.11)	21(1.86)	13(2.68)		
T1	672(41.64)	462(40.92)	210(43.30)			HR-/HER2-	1392(86.25)	979(86.71)	413(85.15)		
T2	758(46.96)	542(48.01)	216(44.54)			ER状态/(例, %)				1.075	0.300
T3	56(3.47)	41(3.63)	15(3.09)			阴性	46(2.85)	29(2.57)	17(3.51)		
T4	128(7.93)	84(7.44)	44(9.07)			阳性	1568(97.15)	1100(97.43)	468(96.49)		
N分期/(例, %)				2.347	0.504	PR状态/(例, %)				4.174	0.041
N0	882(54.65)	616(54.56)	266(54.85)			阴性	147(9.11)	92(8.15)	55(11.34)		
N1	499(30.92)	341(30.20)	158(32.58)			阳性	1467(90.89)	1037(91.85)	430(88.66)		
N2	138(8.55)	101(8.95)	37(7.63)			HER2状态/(例, %)				0.180	0.671
N3	95(5.89)	71(6.29)	24(4.95)			阴性	1 426(88.35)	1000(88.57)	426(87.84)		
M分期/(例, %)				<0.001	0.994	阳性	188(11.65)	129(11.43)	59(12.16)		
M0	1 491(92.38)	1 043(92.38)	448(92.37)			缺失状态/(例, %)				0.357	0.550
M1	123(7.62)	86(7.62)	37(7.63)			0	1 229(76.15)	855(75.73)	374(77.11)		
						1	385(23.85)	274(24.27)	111(22.89)		

2.2 训练集生存/死亡两组相关资料比较

1 129 例的训练集中, 5 年生存(生存组)855 例, 死亡(死亡组)274 例。对比两组人口学特征及肿瘤相关因素, 年龄、婚姻状况、种族、肿瘤分级、TNM

分期、肿瘤转移状态、是否手术治疗、是否化疗、肿瘤亚分型、ER、PR、HER2 等指标的差异有统计学意义($P < 0.05$)。内部验证集中死亡 111 例, 生存 374 例, 两组比较结果与训练集基本一致。见表 2。

表 2 训练和内部验证的死亡组与生存组基线资料比较

Tab. 2 Basic characteristics comparison between survival and death groups in training set and internal testing set

临床特征	训练集(n=1 129)		P 值	内部验证集(n=485)		P 值
	生存(n=855)	死亡(n=274)		生存(n=374)	死亡(n=111)	
年龄[岁, <i>M(IQR)</i>]	65(57.0 ~ 71.0)	67(58.25 ~ 74)	0.003	65(54 ~ 76)	65(52 ~ 78)	0.240
婚姻状况/(例, %)			<0.001			<0.001
已婚	598(69.94)	150(54.74)		276(73.80)	62(55.86)	
未婚	147(17.19)	67(24.45)		59(15.78)	23(20.72)	
其他	110(12.87)	57(20.80)		39(10.43)	26(23.42)	
种族/(例, %)			0.009			0.008
白种人	93(10.88)	47(17.15)		38(10.16)	23(20.72)	
黑种人	53(6.20)	10(3.65)		31(8.29)	5(4.50)	
其他	709(82.92)	217(79.20)		305(81.55)	83(74.77)	
肿瘤位置/(例, %)			0.077			0.411
中央区	353(41.29)	123(44.89)		156(41.71)	39(35.14)	
乳房外象限	111(12.98)	22(8.03)		50(13.37)	12(10.81)	
乳房内象限	156(18.25)	43(15.69)		62(16.58)	22(19.82)	
其他	235(27.49)	86(31.39)		106(28.34)	38(34.23)	
组织学分级/(例, %)			<0.001			<0.001
I	123(14.39)	13(4.74)		51(13.64)	7(6.31)	
II	445(52.05)	105(38.32)		204(54.55)	31(27.93)	
III	237(27.72)	131(47.81)		94(25.13)	65(58.56)	
未知	50(5.85)	25(9.12)		25(6.68)	8(7.21)	
T 分期/(例, %)			<0.001			<0.001
T1	418(48.89)	44(16.06)		188(50.27)	22(19.82)	
T2	375(43.86)	167(60.95)		164(43.85)	52(46.85)	
T3	19(2.22)	22(8.03)		9(2.41)	6(5.41)	
T4	43(5.03)	41(14.96)		13(3.48)	31(27.93)	
N 分期/(例, %)			<0.001			<0.001
N0	518(60.58)	98(35.77)		232(62.03)	34(30.63)	
N1	247(28.89)	94(34.31)		107(28.61)	51(45.95)	
N2	58(6.78)	43(15.69)		25(6.68)	12(10.81)	
N3	32(3.74)	39(14.23)		10(2.67)	14(12.61)	
M 分期/(例, %)			<0.001			<0.001
M0	845(98.83)	198(72.26)		371(99.20)	77(69.37)	
M1	10(1.17)	76(27.74)		3(0.80)	34(30.63)	
骨转移/(例, %)			<0.001			<0.001
无/未知	850(99.42)	218(79.56)		372(99.47)	86(77.48)	
有	5(0.58)	56(20.44)		2(0.53)	25(22.52)	
肝转移/(例, %)			<0.001			<0.001
无/未知	855(100)	263(95.99)		374(100)	107(96.40)	
有	0(0)	11(4.01)		0(0)	4(3.60)	
肺转移/(例, %)			<0.001			<0.001
无/未知	851(99.53)	251(91.61)		374(100)	93(83.78)	
有	4(0.47)	23(8.39)		0(0)	18(16.22)	

续表

临床特征	训练集(n=1 129)			内部验证集(n=485)		
	生存(n=855)	死亡(n=274)	P值	生存(n=374)	死亡(n=111)	P值
手术/(例,%)			<0.001			<0.001
否/未知	9(1.05)	54(19.71)		10(2.67)	26(23.42)	
是	846(98.95)	220(80.29)		364(97.33)	85(76.58)	
放疗/(例,%)			0.272			0.024
无/未知	595(69.59)	181(66.06)		274(73.26)	69(62.16)	
有	260(30.41)	93(33.94)		100(26.74)	42(37.84)	
化疗/(例,%)			0.040			0.001
无/未知	479(56.02)	134(48.91)		233(62.30)	50(45.05)	
有	376(43.98)	140(51.09)		141(37.70)	61(54.95)	
肿瘤亚分期/(例,%)			<0.001			0.010
HR+/HER2-	88(10.29)	41(14.96)		39(10.43)	20(18.02)	
HER2+	10(1.17)	11(4.01)		7(1.87)	6(5.41)	
HR-/HER2-	757(88.54)	222(81.02)		328(87.70)	85(76.58)	
ER状态/(例,%)			0.002			0.215
阴性	15(1.75)	14(5.11)		11(2.94)	6(5.41)	
阳性	840(98.25)	260(94.89)		363(97.06)	105(94.59)	
PR状态/(例,%)			0.003			0.012
阴性	58(6.78)	34(12.41)		35(9.36)	20(18.02)	
阳性	797(93.22)	240(87.59)		339(90.64)	91(81.98)	
HER2状态/(例,%)			0.034			0.032
阴性	767(89.71)	233(85.04)		335(89.57)	91(81.98)	
阳性	88(10.29)	41(14.96)		39(10.43)	20(18.02)	

2.3 死亡结局预测模型特征筛选

Lasso 回归初筛选出年龄、肿瘤组织学分级、T分期、N分期、M分期、骨转移情况、是否手术、ER状态、PR状态等9个因素，见图1。单因素Cox回归结果发现9个因素全部有统计学差异，与MBC患者死亡风险显著关联($P<0.05$)。多因素Cox回归结果亦发现9个因素全部具有统计学差异($P<0.05$)，见表3。

2.4 MBC预后预测模型效能评估

ROC分析显示，由训练集所构建的1年、3年、5年死亡预测模型的ROC曲线下面积分别为0.895、0.853、0.829，对应的C-index值为0.811；由内部验证集所构建的1年、3年、5年死亡预测模型的ROC曲线下面积分别为0.858、0.858、0.848，对应的C-index值为0.798。见图2。

校准曲线评估显示校准曲线均呈近似直线，提示预测值与实际值间有较好的线性关系。各数据点和校准曲线贴合紧密，分布偏离程度较小，提示数据的拟合度较高(图3)。

列线图构建及解析可见预测模型中9个因子的总分为425分，各因子得分40~100。其中年龄的权重最大，TNM分期及是否手术占权重也较大，

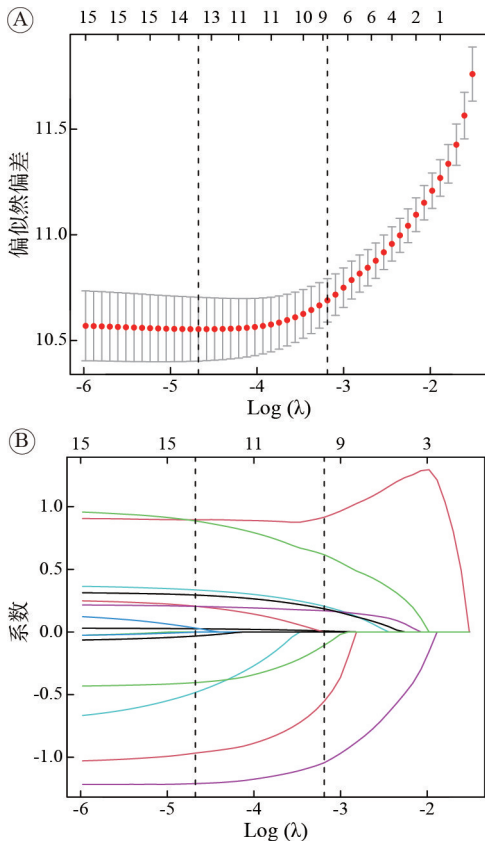
提示它们对模型及患者1年、3年、5年的死亡风险影响较大(图4)。

2.5 MBC预后预测模型的外部验证

建模集与外部验证集年龄、肿瘤位置、组织学分级、ER、PR、HER2、侵犯神经、脉管浸润、淋巴结阳性个数、Ki-67、淋巴结转移情况、病理分期和5年死亡等资料均无统计学差异($P<0.05$)。对上述预后预测模型进行外部验证，预测5年死亡风险的AUC为0.731；预测10年死亡风险的AUC为0.797(图5)。

3 讨论

由于MBC相对罕见，目前医师和患者对MBC的认知均存在不足，先前亦有研究表明，首诊时MBC患者为晚期的比例高于女性乳腺癌患者，这可能与MBC患者对乳腺癌认识不足、筛查不够有关^[13]。此外，MBC与女性乳腺癌的临床病理特征存在显著不同^[13-14]。美国SEER数据库及解放军总医院第一医学中心回顾性分析也有类似发现，MBC患者多为激素受体阳性/HER2阴性(HR+/HER2-)亚型(占比86.25%，共1 392例)，97.15%的患者ER阳性(外部验证中为90.9%)，仅11.7%的患



横坐标轴：标准化后的系数向量；纵坐标：系数的值。

图1 Lasso回归预测训练集临床预测模型的相关因素

Fig. 1 Related factors of Lasso regression prediction training set clinical prediction model

者HER2阳性(外部验证中为38.6%)。不同国际乳腺癌指南中，MBC治疗标准多参照女性乳腺癌^[15-16]。尽管MBC患者可从局部和全身治疗中获益，但因诊断较晚或年龄较大，其预后通常比女性乳腺癌患者更差^[17]。此外，MBC患者发生对侧乳腺癌风险更高，增加了死亡风险^[18]。因此建立针对MBC的危险分层工具，以指导个体化治疗。本研究采用SEER数据库数据建立MBC患者死亡风险预测模型，采用解放军总医院第一医学中心的数据进行外部验证，系统筛选出诊断年龄、组织学分级、T分期、N分期、M分期、骨转移状态、手术干预状态、ER状态和PR状态这9个与MBC预后密切相关的因素，并构建了预测MBC患者死亡的临床模型。该模型为临床医师提供了一种量化评估MBC患者的预后工具，还通过列线图将模型进行可视化，提高了临床决策的准确性。

本模型中纳入了9个变量，即年龄、组织学分级、T、N、M、骨转移、手术、ER、PR。本研究显示，随着年龄的增长，病死率增加，这可能与本研究纳入的MBC患者年龄较大有关，本研究中

表3 单因素与多因素Cox比例风险回归结果

Tab. 3 Univariate and Multivariate Cox proportional hazards regression analysis results

因素	单因素Cox回归		多因素Cox回归	
	P值	HR (95% CI)	P值	HR (95% CI)
年龄	0.001	1.019(1.008 ~ 1.031)	<0.001	1.029(1.018 ~ 1.041)
组织学分级				
I		1.000(Reference)		1.000(Reference)
II	0.018	2.009(1.129 ~ 3.575)	0.336	1.332(0.742 ~ 2.391)
III	<0.001	4.272(2.415 ~ 7.555)	0.002	2.457(1.373 ~ 4.397)
未知	<0.001	3.923(2.007 ~ 7.669)	0.014	2.359(1.190 ~ 4.673)
T分期				
T1		1.000(Reference)		1.000(Reference)
T2	<0.001	3.770(2.704 ~ 5.256)	<0.001	2.486(1.761 ~ 3.509)
T3	<0.001	8.792(5.262 ~ 14.688)	<0.001	3.605(2.085 ~ 6.231)
T4	<0.001	6.977(4.556 ~ 10.684)	0.004	2.069(1.265 ~ 3.383)
N分期				
N0		1.000(Reference)		1.000(Reference)
N1	<0.001	1.855(1.398 ~ 2.462)	0.048	1.347(1.003 ~ 1.810)
N2	<0.001	3.181(2.222 ~ 4.555)	<0.001	1.936(1.313 ~ 2.854)
N3	<0.001	4.451(3.068 ~ 6.458)	<0.001	2.461(1.641 ~ 3.693)
M分期				
M0		1.000(Reference)		1.000(Reference)
M1	<0.001	10.814(8.235 ~ 14.200)	0.004	2.202(1.279 ~ 3.792)
骨转移				
无/未知		1.000(Reference)		1.000(Reference)
有	<0.001	12.085(8.922 ~ 16.370)	<0.001	2.477(1.447 ~ 4.239)
手术				
无/未知		1.000(Reference)		1.000(Reference)
有	<0.001	0.099(0.073 ~ 0.134)	<0.001	0.255(0.168 ~ 0.387)
ER状态				
阴性		1.000(Reference)		1.000(Reference)
阳性	<0.001	0.248(0.145 ~ 0.426)	0.042	0.476(0.233 ~ 0.973)
PR状态				
阴性		1.000(Reference)		1.000(Reference)
阳性	<0.001	0.498(0.348 ~ 0.714)	0.028	0.589(0.368 ~ 0.944)

MBC患者中位数年龄为65岁，随着年龄的增长，各项基础疾病增多，对化疗等系统治疗的耐受性差，因此病死率增加。组织学分级是衡量肿瘤恶性程度的指标，组织学分级越高，提示乳腺癌细胞分化越差，增殖活性、侵袭性越强，死亡风险越高。TNM分期是国际公认的临床分期，其级别越高，提示肿瘤越晚期，死亡风险越高。骨转移转移也是远处转移的一种形式，出现远处转移的患者死亡风险升高。此外，我们发现接受了手术治疗的患者，死亡风险低，而未接受手术或失去手术治疗机会的患者，死亡风险高。有机会接受根治性手术治疗的患者，往往临床分期相对较早，

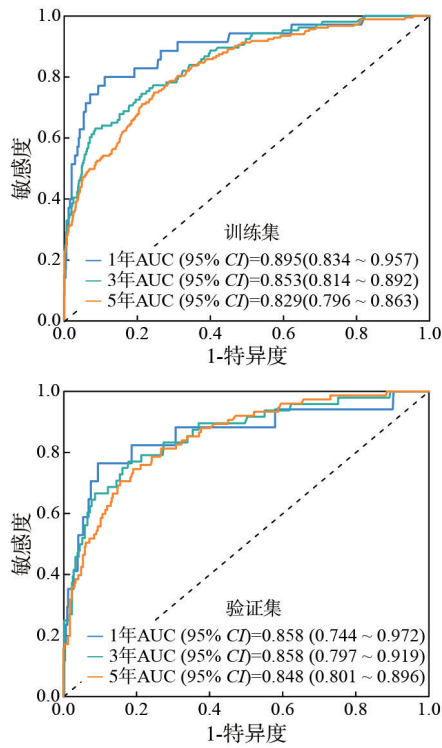


图2 训练集和验证集中临床预测模型的1年、3年、5年ROC分析
 Fig. 2 1-year, 3-year, and 5-year ROC analysis of clinical prediction models in the training and validation sets

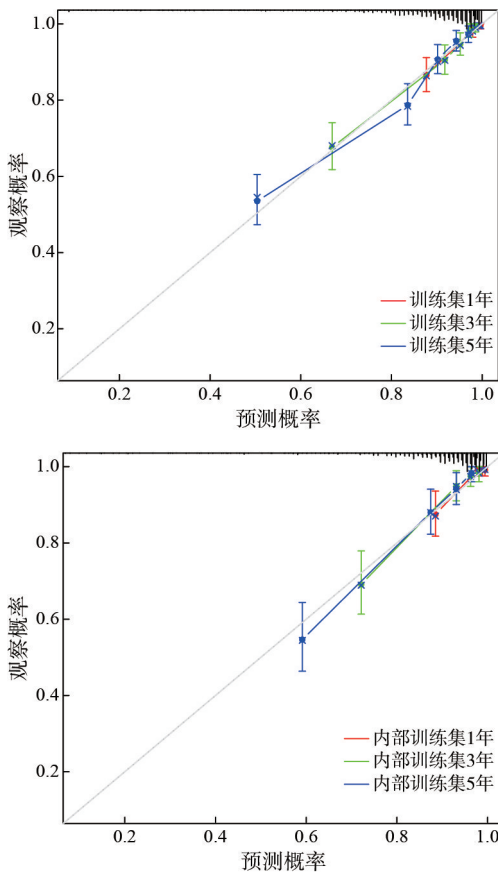


图3 训练集和验证集中临床预测模型的1年、3年、5年校准曲线
 Fig. 3 Calibration curves for clinical prediction models in the training and internal testing sets at 1, 3, and 5 years

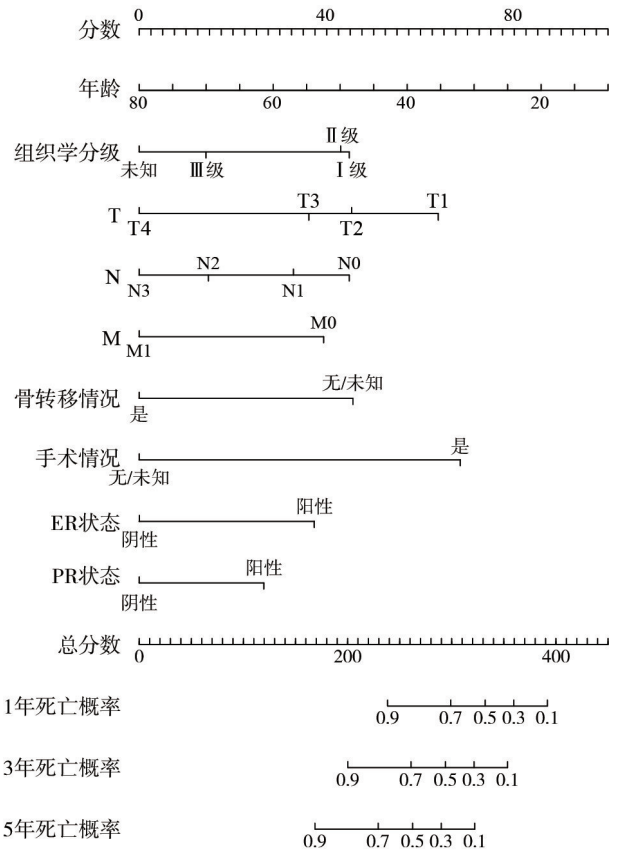


图4 男性乳腺癌患者死亡预测模型列线图
 Fig. 4 Nomogram of mortality prediction model in patients with male breast cancer

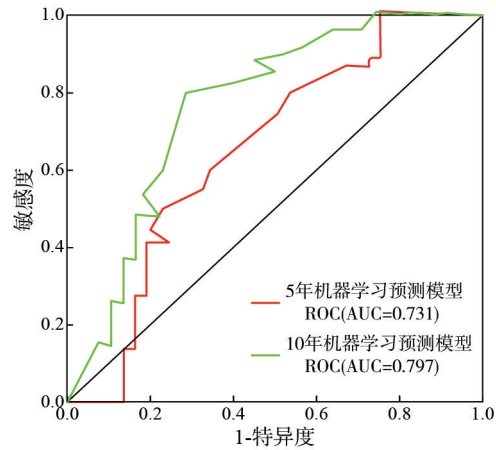


图5 基于外部验证, 男性乳腺癌手术切除死亡风险的机器学习预测模型5年、10年ROC曲线
 Fig. 5 Analysis of ROC curves for the machine learning prediction model of post-surgical mortality in male breast cancer patients constructed from external validation data (5-year and 10-year)

因此死亡风险相对较低。丹麦一项 643 例 MBC 病例的队列研究表明, 年龄大、肿瘤体积大、淋巴结阳性、分级高是影响 MBC 患者死亡的危险因素^[19]。我国一项 152 例 MBC 病例的研究报道, 肿瘤大小是 MBC 患者死亡和无病生存的影响因

素^[20]；另一项77例中国MBC病例的研究则发现，M1分期是死亡的独立危险因素^[21]。以上研究支持本研究，表明年龄增加、组织学分级、TNM分期高等是MBC患者死亡率增加的危险因素。此外，我们发现ER、PR阳性是存活的有利因素。一项65例MBC患者的病例对照研究发现，ER阳性患者病死率更低，而PR状态与患者病死率无关^[22]。一项涉及10 873例MBC病例的研究显示，PR表达阳性与更低的病死率相关^[23-24]。ER、PR阳性的MBC患者，对内分泌治疗较为敏感，因此预后得以改善。

本研究通过划分训练集和验证集评估临床预测模型的区分度和校准度。结果显示，训练集和验证集的ROC曲线下面积均较大，表明模型区分能力良好；校准曲线也显示模型对患者死亡的预测校准度较好。表明本研究构建的临床预测模型在预测MBC患者死亡方面准确可靠。此外，本研究采用列线图对临床预测模型进行可视化提高了临床决策的效率与准确性。同时，对解放军总医院第一医学中心MBC患者的外部验证结果，同样证实了回归模型在预测MBC患者死亡方面具有良好的价值，支持了该临床预测模型在实际临床应用中的可行性和有效性。

本研究也存在一定局限性。(1)因MBC发病率低，研究纳入样本量有限，这可能影响模型的稳定性与泛化能力。未来可增加随样本量和数据，进一步优化模型。(2)本研究主要依托SEER数据库和解放军总医院第一医学中心的数据进行验证。尽管这些数据有一定代表性和可靠性，但不同国家、人种和医院的患者特征及医疗资源存在差异，实际应用时需考虑这些因素对模型预测效能的影响。(3)因素纳入可能不全。变量筛选采用Lasso回归以及单变量和多变量分析，可能遗漏变量间的交互作用，后续可运用贝叶斯模型平均法优化变量，同时扩大样本量，构建更优的MBC疾病风险预测模型。(4)男性乳腺癌在临床上相对罕见，本研究外部验证集的数据来源于单个临床中心，纳入的病例数相对较少，在1年、3年内死亡的人数过少，因此无法对1年、3年的死亡风险进行验证，同时样本量可能也是造成外部验证AUC相对低的原因。

综上所述，本研究的可解释Cox回归模型在可

解释性和处理生存分析问题上具有独特的优势，可为临床医师提供更直观、有价值的信息，便于在实际临床中应用推广。未来研究需要进一步扩大样本量，纳入更多的临床中心数据并进一步探索融合不同模型，充分发挥各自优势，提高MBC患者死亡风险预测的准确性、普适性和可靠性。

作者贡献 田林：数据收集，文章设计，文章撰写；朱荔、王建东：指导文章选题，对文章的知识性内容作批评性审阅；张丽萍、孙嘉鸿：数据收集，支持性贡献。

利益冲突 所有作者声明无利益冲突。

数据共享声明 本论文相关数据可依据合理理由从作者处获取，Email: tl1359525567@163.com。

参考文献

- 1 Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018 [J]. *CA A Cancer J Clin*, 2018, 68 (1): 7-30.
- 2 Chidambaram A, Prabhakaran R, Sivasamy S, et al. Male breast cancer: current scenario and future perspectives [J]. *Technol Cancer Res Treat*, 2024, 23: 15330338241261836.
- 3 DeSantis CE, Ma JM, Gaudet MM, et al. Breast cancer statistics, 2019 [J]. *CA Cancer J Clin*, 2019, 69 (6): 438-451.
- 4 Asgharian M, Moslemi D, Nikbakht HA, et al. Male breast cancer: a 32-year retrospective analysis in radiation therapy referral center in northern Iran [J]. *Ann Med Surg (Lond)*, 2024, 86 (10): 5756-5761.
- 5 Fentiman IS. Prognostic difficulties of men with breast cancer [J]. *Breast J*, 2021, 27 (12): 877-882.
- 6 Yi JR, Shuang ZY, Zhong WJ, et al. Identification of immune-related risk characteristics and prognostic value of immunophenotyping in TNBC [J]. *Front Genet*, 2021, 12: 730442.
- 7 Han YQ, Wang JY, Wang ZJ, et al. Sex-based heterogeneity in the clinicopathological characteristics and prognosis of breast cancer: a population-based analysis [J]. *Front Oncol*, 2021, 11: 642450.
- 8 Xie J, Ying YY, Xu B, et al. Metastasis pattern and prognosis of male breast cancer patients in US: a population-based study from SEER database [J]. *Ther Adv Med Oncol*, 2019, 11: 1758835919889003.
- 9 Scomersi S, Giudici F, Cacciatore G, et al. Comparison between male and female breast cancer survival using propensity score matching analysis [J]. *Sci Rep*, 2021, 11 (1): 11639.
- 10 Foerster R, Foerster FG, Wulff V, et al. Matched-pair analysis of patients with female and male breast cancer: a comparative analysis [J]. *BMC Cancer*, 2011, 11: 335.
- 11 Sekeroglu B, Tuncal K. Prediction of cancer incidence rates for the European continent using machine learning models [J]. *Health Informatics J*, 2021, 27 (1): 1460458220983878.
- 12 Gradishar WJ, Moran MS, Abraham J, et al. Breast cancer, version 3.2024, NCCN clinical practice guidelines in oncology [J]. *J Natl Compr Canc Netw*, 2024, 22 (5): 331-357.
- 13 Montenegro MA, Domingues TD, Garcia TM, et al. Male breast cancer in Portugal: a descriptive analysis of a 20-year cohort [J]. *Eur J Breast Health*, 2025, 21 (2): 154-161.