

引用本文:章锐,阮剑,赵建国.基于SEER数据库构建的远处转移性乳腺癌的预后列线图[J].安徽医药,2022,26(9):1781-1785.DOI:10.3969/j.issn.1009-6469.2022.09.020.



◇临床医学◇

## 基于SEER数据库构建的远处转移性乳腺癌的预后列线图

章锐,阮剑,赵建国

作者单位:武汉市第一医院甲状腺乳腺外科,湖北 武汉 430000

通信作者:赵建国,男,主任医师,硕士生导师,研究方向为甲状腺疾病和乳腺疾病,Email:3040114010@qq.com

基金项目:湖北省卫生健康科研基金资助项目(WJ2021M001);武汉市卫生健康科研基金资助项目(WX20B10)

**摘要:** **目的** 构建远处转移性乳腺癌的有效预后列线图。**方法** 从SEER(The Surveillance, Epidemiology, and End Results)数据库中筛选到2010年1月至2012年12月诊断为远处转移性乳腺癌的3 989例病人,采用整群随机抽样法分为建模组( $n=2 993$ )和验证组( $n=996$ ),使用Cox比例风险回归模型分析建模组预后独立危险因素,并构建预后列线图,利用内部自助重取样和验证组病人验证结果,列线图的预测准确性和判别能力由C指数和校准曲线确定。**结果** 在对建模组病人的多因素分析中,生存的独立危险因素如年龄、婚姻状态、种族、组织学分级、雌激素受体(ER)状态、孕激素受体(PR)状态、人表皮生长因子受体2(Her-2)状态、T分期、原发灶手术、骨转移、脑转移、肝转移和肺转移都被选入列线图。生存概率的校准曲线显示出列线图预测良好的一致性,列线图预测生存的C指数为0.702[95%CI:(0.696, 0.708)],在验证队列中,列线图的C指数为0.707[95%CI:(0.706, 0.708)]。**结论** 该列线图可以为远处转移性乳腺癌病人提供较为准确的预后预测。

**关键词:** 乳腺肿瘤; 肿瘤转移; 列线图; 生存; SEER数据库; 预后

### Prognostic nomogram for metastatic advanced breast cancer constructed based on SEER data

ZHANG Rui, RUAN Jian, ZHAO Jianguo

Author Affiliation: Department of Thyroid and Breast Surgery, Wuhan No.1 Hospital, Wuhan, Hubei 430000, China

**Abstract:** **Objective** To construct and validate an effective prognostic nomogram for metastatic advanced breast cancer. **Methods** The nomogram was based on 3 989 patients who were diagnosed with metastatic advanced breast cancer from January 2010 to December 2012 in SEER (The Surveillance, Epidemiology, and End Results) program data. The data was assigned into the training group ( $n=2 993$ ) and the validation group ( $n=996$ ) by random sampling. The Cox proportional hazard regression model was performed to identify the independent risk factors affecting prognosis of the training group, and the prognostic nomogram was constructed. The results were internally validated using bootstrap resampling and externally with the validation group, the predictive accuracy and discriminative ability of the nomogram were determined by concordance index (C-index) and calibration curve. **Results** On multivariate analysis of the training group, independent factors for survival were age, marital status, race and grade, estrogen receptor (ER) status, progesterone receptor (PR) status, human epidermal Growth factor receptor 2 (Her-2) status, stage T, surgery of primary site, bone metastasis, brain metastasis, liver metastasis and lung metastasis, which were all selected into the nomogram. The calibration curve for probability of survival showed good agreement between prediction by nomogram, the C-index of the nomogram for predicting survival was 0.702 [95%CI: (0.696, 0.708)]. In the validation group, the C-index of nomogram discrimination was 0.707 [95%CI: (0.706, 0.708)]. **Conclusion** The nomogram can provide more accurate prognosis prediction for patients with metastatic advanced breast cancer.

**Key words:** Breast neoplasms; Neoplasm metastasis; Nomogram; Survival; The Surveillance, Epidemiology, and End Results program; Prognosis

乳腺癌是全世界女性中最常见的恶性肿瘤,也是导致癌症死亡的第2大原因<sup>[1]</sup>,从1976年至2009年,50岁以下女性乳腺癌的发病率以每年0.2%的速度增长,而在被确诊为乳腺癌的病人中,远处转移的发病率则以每年约2%的速度增长<sup>[2]</sup>。各期乳腺癌的5年相对生存率为89.7%,而晚期乳腺癌的预后较差,5年生存率仅为26.9%<sup>[3]</sup>。晚期乳腺癌包

括远处转移和局部晚期乳腺癌,目前远处转移性乳腺癌仍然是无法治愈的疾病,激素受体阳性并且人表皮生长因子受体2(Her-2)阴性的晚期乳腺癌比其他类别的晚期乳腺癌具有更好的生存率<sup>[3-4]</sup>,但是晚期乳腺癌(远处转移性乳腺癌)个体化的预后并不明确。

列线图是一个能综合分析预后危险因素,从而

进行可视化、个体化评估的方法,目前被广泛用于各种类型的癌症<sup>[5-6]</sup>,但是在远处转移性乳腺癌方面,仅发现极少量的研究。本研究基于SEER(The Surveillance, Epidemiology, and End Results)数据库首次较好地建立远处转移性乳腺癌的预后列线图,根据SEER数据库中远处转移性乳腺癌病人的临床病理资料,以确定该模型可以较为准确地预测病人的存活率,从而为临床医师对远处转移性乳腺癌预后判断及治疗提供参考。

## 1 资料与方法

**1.1 数据资料获取** 使用SEER\*stat软件从数据库中提取数据,提取条件为“(1)发病部位:乳腺;(2)性别:女性;(3)诊断年份:2010—2012年;(4)既往无肿瘤病史”,提取变量为“病人ID、生存时间、生存状态、年龄、种族、婚姻状态、组织学分级、原发灶是否手术、雌激素受体(ER)状态、孕激素受体(PR)状态、Her-2状态、第七版美国癌症联合委员会(AJCC)T分期、第七版AJCC N分期、有无骨转移、有无脑转移、有无肝转移、有无肺转移”。本研究符合《世界医学协会赫尔辛基宣言》相关要求。

**1.2 纳入、排除标准** 纳入标准:为保证所有病人均随访5年以上,选择诊断年份为2012年前的乳腺癌;远处转移数据库中SEER combined mets at DX-bone /brain/liver/lung(2010+)为2010年后的乳腺癌。共获得临床资料完整的139 584例女性乳腺癌病人。排除标准:生存时间为0,无远处转移。获得2010年1月至2012年12月具有完整临床信息的有效数据3 989例。

**1.3 数据分组** 通过整群随机抽样法<sup>[7]</sup>,以3:1分组,从3 989例病人选2 993例为建模组,996例为验证组。

**1.4 统计学方法** 采用SPSS 25.0统计分析危险因素,变量分组计数后采用 $\chi^2$ 检验,多变量分析采用Cox比例风险回归模型筛选影响预后的独立危险因素,按照文献<sup>[5-6]</sup>中的方法基于多变量分析的结果使用R版本3.6.2(<https://www.r-project.org/>)中的“rms”包和“survival”包绘制列线图。列线图的准确性通过C指数和校准曲线进行衡量,C指数越大(接近1),预后预测越准确。内部验证通过Bootstrap法重复取样1 000次自检完成,对列线图的外部验证通过验证组进行验证。受试者工作特征曲线(ROC曲线)下面积用于评估模型预测远处转移性乳腺癌预后的准确性。 $P<0.05$ 被认为差异有统计学意义。

## 2 结果

**2.1 两组的临床病理资料** 两组临床病理资料差异无统计学意义( $P>0.05$ ),见表1。

**2.2 Cox比例风险回归分析** 分析提示年龄、婚姻状态、种族、组织学分级、ER状态、PR状态、Her-2状态、T分期、原发灶是否手术、骨转移、脑转移、肝转移、肺转移是影响预后的独立危险因素( $P<0.05$ ),其中相对于年轻( $<35$ 岁),年老( $>49$ 岁)是远处转移性乳腺癌病人预后的危险因素( $P<0.05$ );Her-2阴性相对于Her-2阳性是预后的危险因素( $P<0.05$ );原发灶手术相对于原发灶不手术具有更低的风险比( $P<0.05$ );N分期与预后相关性差异无统计学意义( $P>0.05$ )。Cox单因素和多因素回归分析见表2。

**2.3 远处转移性乳腺癌生存预后列线图及其内/外部验证** 基于Cox比例风险回归分析绘制列线图,见图1。内部验证的C指数为0.702[95%CI:(0.696, 0.708)],远处转移性乳腺癌校准曲线提示在观察值与预测值之间的一致性:1年和3年生存率一致性良好,5年生存率一致性较好,见图2。外部验证C指数为0.707[95%CI:(0.706, 0.708)],1年和3年生存率外部验证一致性良好,5年生存率一致性较好,见图3。

**2.4 远处转移性乳腺癌建模组和验证组的ROC曲线** 建模组和验证组的校准图提示该临床预测模型在1年和3年的预测结果与实际观测数据具有良好的 consistency。建模组1年、3年、5年的ROC曲线下面积分别为0.8、0.7、0.7,验证组的1年、3年、5年的ROC曲线下面积分别为0.8、0.8、0.695,也证实了该模型能够准确预测远处转移性乳腺癌的预后,尤其是1年和3年预后。

## 3 讨论

随着乳腺癌药物治疗的研究发展及相关临床试验的开展,乳腺癌的远处转移率及死亡率显著下降,但是远处转移仍然是乳腺癌病人致死的主要原因<sup>[8-9]</sup>,远处转移的预后不良,其5年生存率约为27%,脑转移乳腺癌病人预后最差,中位总生存时间为10个月[95%CI:(9, 12)个月]<sup>[10]</sup>。乳腺癌的随访研究不仅能及早发现复发及转移,并且能从中不断总结经验,完善诊治。SEER数据库涵盖了大约28%的美国人口的癌症治疗随访数据<sup>[11]</sup>,该数据库中获得的研究数据代表了该人群的流行病学状况,在癌症监测科学以及收集、分析、解释和传播可靠的基于人口的统计数据方面处于国家领导地位。SEER数据库收集的癌症数据中免疫组化从2010年开始录入,癌症结局包括死亡与否以及死亡原因,对研究复发、转移或进展有限制,但该数据库对预后研究具有较大的指导意义。

据文献报道,晚期乳腺癌的预后与以下变量有关:肿瘤的分型、年龄<sup>[12]</sup>、生理机能、转移的器官、全身化疗、原发灶手术切除、原发灶的大小、孤立转移/

表1 两组远处转移性乳腺癌病人的临床病理资料比较/例(%)

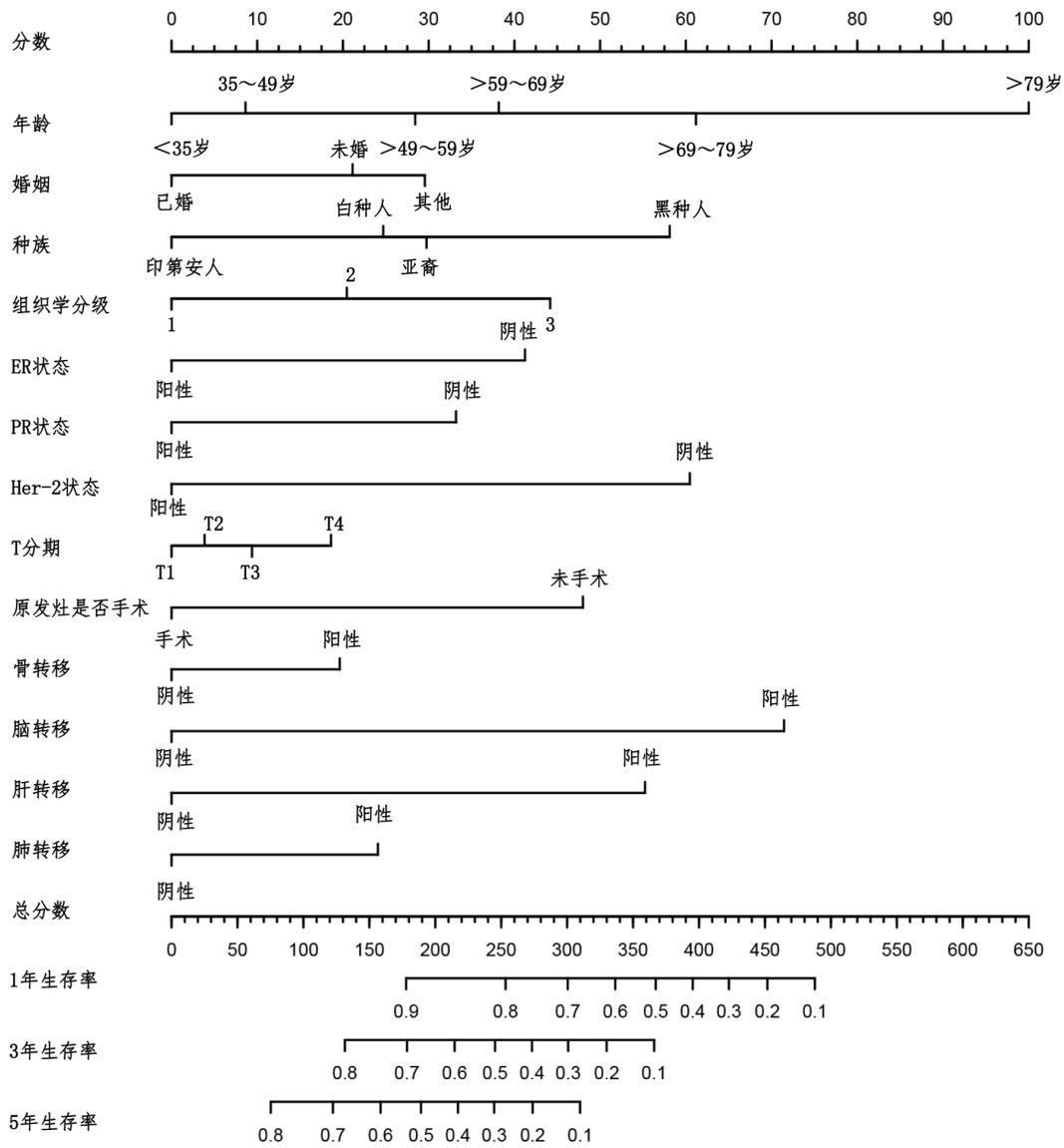
项目	建模组 (n=2 993)	验证组 (n=996)	$\chi^2$ 值	P值	项目	建模组 (n=2 993)	验证组 (n=996)	$\chi^2$ 值	P值
年龄			1.70	0.888	Her-2状态			0.49	0.482
<35岁	98(3.3)	34(3.4)			阳性	739(24.7)	257(25.8)		
35~49岁	613(20.5)	198(19.9)			阴性	2 254(75.3)	739(74.2)		
>49~59岁	785(26.2)	256(25.7)			组织学分级			0.91	0.634
>59~69岁	738(24.7)	246(24.7)			1	242(8.1)	72(7.2)		
>69~79岁	469(15.7)	152(15.3)			2	1 236(41.3)	422(42.4)		
>79岁	290(9.6)	110(11.0)			3	1 515(50.6)	502(50.4)		
婚姻状态			1.27	0.530	T分期			1.24	0.742
未婚	652(21.8)	208(20.9)			T1	435(14.5)	131(13.2)		
已婚	1 385(46.3)	451(45.3)			T2	1 032(34.5)	348(34.9)		
其他	956(31.9)	337(33.8)			T3	544(18.2)	181(18.2)		
种族			0.46	0.927	T4	982(32.8)	336(33.7)		
白种人	2 278(76.1)	766(76.9)			N分期			2.26	0.520
黑种人	510(17.0)	162(16.3)			N0	732(24.5)	249(25.0)		
印第安人	18(0.6)	5(0.5)			N1	1 350(45.1)	463(46.5)		
亚裔	187(6.3)	63(6.3)			N2	390(13.0)	131(13.2)		
ER状态			0.40	0.526	N3	521(17.4)	153(15.3)		
阳性	2 257(75.4)	761(76.4)			肿瘤原发灶手术	1 638(54.7)	525(52.7)	1.22	0.268
阴性	736(24.6)	235(23.6)			有骨转移	2 171(72.5)	728(73.1)	0.12	0.733
PR状态			1.28	0.258	有脑转移	199(6.6)	66(6.6)	0.00	0.980
阳性	1 839(61.4)	632(63.5)			有肝转移	823(27.5)	280(28.1)	0.14	0.707
阴性	1 154(38.6)	364(36.5)			有肺转移	996(33.3)	301(30.2)	3.18	0.074

注:ER为雌激素受体,PR为孕激素受体,Her-2为人表皮生长因子受体2。

表2 建模组2 993例远处转移性乳腺癌病人预后影响因素Cox单因素和多因素回归分析

变量	单因素分析			多因素分析			变量	单因素分析			多因素分析		
	P值	HR值	95%CI	P值	HR值	95%CI		P值	HR值	95%CI	P值	HR值	95%CI
<35岁		参照			参照		Her-2阳性		参照			参照	
35~49岁	0.718	1.05	(0.79,1.39)				Her-2阴性	<0.001	1.24	(1.12,1.37)	<0.001	1.83	(1.63,2.05)
>49~59岁	0.025	1.36	(1.04,1.78)	0.047	1.32	(1.00,1.74)	T1期		参照			参照	
>59~69岁	0.011	1.42	(1.08,1.86)	0.008	1.46	(1.10,1.93)	T2期	0.32	1.07	(0.93,1.23)			
>69~79岁	<0.001	1.85	(1.40,2.44)	<0.001	1.82	(1.37,2.42)	T3期	0.018	1.20	(1.03,1.40)			
>79岁	<0.001	2.67	(2.00,3.54)	<0.001	2.70	(2.00,3.64)	T4期	<0.001	1.52	(1.32,1.74)	0.026	1.17	(1.02,1.36)
未婚		参照			参照		N0		参照			参照	
已婚	<0.001	0.79	(0.71,0.88)	<0.001	0.81	(0.72,0.90)	N1	0.381	0.95	(0.86,1.06)			
其他	<0.001	1.13	(1.13,1.42)				N2	0.384	0.94	(0.81,1.08)			
白种人		参照			参照		N3	0.484	1.05	(0.92,1.20)			
黑种人	<0.001	1.45	(1.30,1.61)	<0.001	1.38	(1.23,1.55)	原发灶未手术		参照			参照	
印第安人	0.204	0.67	(0.36,1.25)				原发灶手术	<0.001	0.58	(0.53,0.66)	<0.001	0.60	(0.55,0.66)
亚裔	0.746	0.97	(0.81,1.16)				无骨转移		参照			参照	
组织学1级		参照			参照		有骨转移	<0.001	0.08	(0.74,0.90)	<0.001	1.22	(1.09,1.35)
组织学2级	0.017	1.24	(1.04,1.47)	0.020	1.23	(1.03,1.47)	无脑转移		参照			参照	
组织学3级	<0.001	1.71	(1.43,2.02)	<0.001	1.54	(1.29,1.85)	有脑转移	<0.001	2.06	(1.77,2.40)	<0.001	2.02	(1.73,2.38)
ER阳性		参照			参照		无肝转移		参照			参照	
ER阴性	<0.001	1.84	(1.67,2.02)	<0.001	1.51	(1.32,1.72)	有肝转移	<0.001	1.52	(1.39,1.67)	<0.001	1.74	(1.57,1.93)
PR阳性		参照			参照		无肺转移		参照			参照	
PR阴性	<0.001	1.68	(1.54,1.83)	<0.001	1.40	(1.24,1.58)	有肺转移	<0.001	1.47	(1.35,1.61)	<0.001	1.27	(1.16,1.40)

注:ER为雌激素受体,PR为孕激素受体,Her-2为人表皮生长因子受体2。



注:ER为雌激素受体,PR为孕激素受体,Her-2为人表皮生长因子受体2。

图1 远处转移性乳腺癌预后列线图

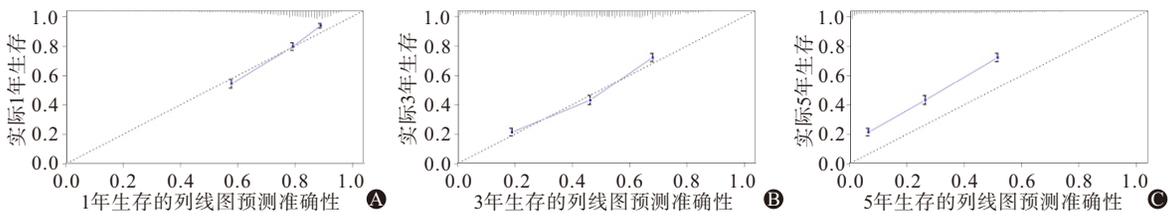


图2 建模组远处转移性乳腺癌1年(A)、3年(B)、5年(C)预测生存的校准曲线

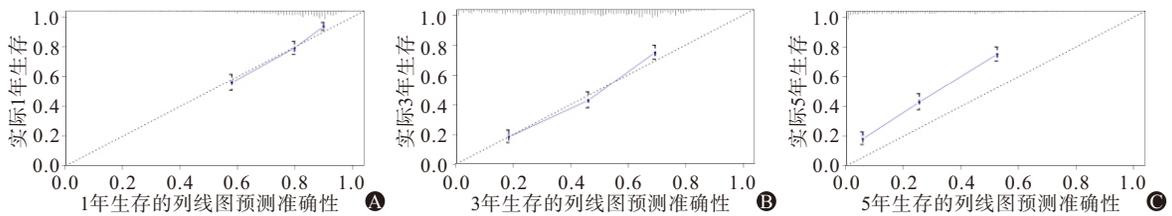


图3 验证组远处转移性乳腺癌1年(A)、3年(B)、5年(C)预测生存的校准曲线

多发转移,因此此次模型建立时纳入了在晚期乳腺癌的生存方面存在差异的因素。在文献中,年龄<50岁较>50岁,局部无复发生存更高,>65岁是晚期

乳腺癌生存的不利因素<sup>[12]</sup>。在本研究中证实年龄同样是晚期乳腺癌预后的重要因素,较大年龄(>49岁)是转移性乳腺癌的危险因素,而在早期乳腺癌

复发风险分层中, <35岁作为复发的危险因素之一, 此次研究侧重于转移性乳腺癌的预后分析, 虽然年轻病人复发转移率较高, 但其对后期治疗的耐受性更好, 年龄越大(>49岁)出现治疗不能耐受的可能越大, 所以其预后更差。Her-2阳性乳腺癌的预后优于三阴性乳腺癌<sup>[4,13]</sup>, Her-2阳性晚期乳腺癌的4年总生存为31%~70%<sup>[14]</sup>, 三阴性晚期乳腺癌的中位生存时间仅为18~25个月<sup>[15]</sup>, 我们分析发现Her-2阳性远处转移性乳腺癌病人预后优于Her-2阴性远处转移性乳腺癌<sup>[16]</sup>, 可能得益于Her-2阳性乳腺癌靶向治疗的多样性、有效性、安全性<sup>[17-19]</sup>。我们构建了预测远处转移性乳腺癌生存预后的列线图, 列线图通过建模组的内部验证和验证组的外部验证, 并且通过校准和判别评估性能。校准曲线是用于评估列线图的估计风险和观察风险之间的一致性, 保证所建立的列线图的可靠性。在远处转移性乳腺癌预后列线图的部分文献中报道其C指数<0.700<sup>[13]</sup>, 而本研究列线图的内部验证和外部验证的C指数分别是0.702和0.707, 高于之前的研究, 并且3年生存率的校准曲线显示生存一致性良好, 表明列线图对转移性乳腺癌具有很高的预测能力, 而5年生存率的校准曲线一般的主要原因在于远处转移性乳腺癌病人(肝、肺、脑转移)5年生存率极低。

尽管我们提供了一个准确性较高的列线图模型, 但是该模型仍有较多不足。因临床资料来源于SEER数据库, 而针对乳腺癌的化疗、靶向治疗、放疗等关键性的治疗方案并不能从该数据库中获得, 所以在一致性上只能达到0.7左右, 如果能够将乳腺癌的相关治疗纳入其中, 可能预测效果更好。

综上所述, 年龄、种族、婚姻状态、组织学分级、原发灶是否手术、ER状态、PR状态、Her-2状态、T分期、骨/脑/肝/肺转移是影响预后的独立危险因素, 基于以上因素建立的列线图能较为准确、直观地个体化预测远处转移性乳腺癌病人的生存率。

### 参考文献

- [1] TURNER NC, SLAMON DJ, RO J, et al. Overall survival with palbociclib and fulvestrant in advanced breast cancer[J]. *N Engl J Med*, 2018, 379(20):1926-1936.
- [2] RUGO HS. Achieving improved survival outcomes in advanced breast cancer[J]. *N Engl J Med*, 2019, 381(4):371-372.
- [3] SAYED REL, JAMAL LEL, ISKANDARANI SEL, et al. Endocrine and targeted therapy for hormone-receptor-positive, HER2-negative advanced breast cancer: insights to sequencing treatment and overcoming resistance based on clinical trials[J]. *Front Oncol*, 2019, 9:510.DOI: 10.3389/fonc.2019.00510.
- [4] LIM HS, SUN W, PARIVAR K, et al. Predicting overall survival and progression-free survival using tumor dynamics in advanced breast cancer patients[J]. *AAPS J*, 2019, 21(2):22.
- [5] 杨梅, 莫永洋, 韩晓蓉, 等. 分子标记列线图模型建立及评估三阴性乳腺癌预后的研究[J]. *中华肿瘤防治杂志*, 2016, 23(15):988-995.
- [6] 陈志航, 陈泽斌, 周倩, 等. 预测肝癌肝切除术后肝衰竭的影像组学列线图模型建立[J/CD]. *中华肝脏外科手术学电子杂志*, 2019, 8(2):143-148.DOI: 10.3877/cma.j.issn.2095-3232.2019.02.014.
- [7] 李皇保, 周俊, 吴晓俊, 等. 预测胆囊癌根治性切除术预后的列线图的构建及验证[J]. *肝胆胰外科杂志*, 2019, 31(4):226-232.
- [8] AHMAD A. Breast cancer statistics: recent trends[J]. *Adv Exp Med Biol*, 2019, 1152:1-7.
- [9] SLAMON DJ, NEVEN P, CHIA S, et al. Overall survival with ribociclib plus fulvestrant in advanced breast cancer[J]. *N Engl J Med*, 2020, 382(6):514-524.
- [10] DESANTIS CE, MA J, GAUDET MM, et al. Breast cancer statistics, 2019[J]. *CA Cancer J Clin*, 2019, 69(6):438-451.
- [11] SHEN K, YAO L, WEI J, et al. Worse characteristics can predict survival effectively in bilateral primary breast cancer: a competing risk nomogram using the SEER database[J]. *Cancer Med*, 2019, 8(18):7890-7902.
- [12] CHOU HH, KUO WL, YU CC, et al. Impact of age on pathological complete response and locoregional recurrence in locally advanced breast cancer after neoadjuvant chemotherapy[J]. *Biomed J*, 2019, 42(1):66-74.
- [13] XIONG Y, CAO H, ZHANG Y, et al. Nomogram-predicted survival of breast cancer brain metastasis: a SEER-based population study [J/OL]. *World Neurosurg*, 2019, 128: e823-e834. DOI: 10.1016/j.wneu.2019.04.262.
- [14] HOPKINS AM, ROWLAND A, MCKINNON RA, et al. Predictors of long-term disease control and survival for HER2-positive advanced breast cancer patients treated with pertuzumab, trastuzumab, and docetaxel [J]. *Front Oncol*, 2019, 9: 789. DOI: 10.3389/fonc.2019.00789.
- [15] SCHMID P, RUGO HS, ADAMS S, et al. Atezolizumab plus nab-paclitaxel as first-line treatment for unresectable, locally advanced or metastatic triple-negative breast cancer (IMpassion130): updated efficacy results from a randomised, double-blind, placebo-controlled, phase 3 trial[J]. *Lancet Oncol*, 2020, 21(1):44-59.
- [16] ANG F, ROWLAND A, MODI ND, et al. Early adverse events predict survival outcomes in HER2-positive advanced breast cancer patients treated with lapatinib plus capecitabine[J]. *J Cancer*, 2020, 11(11):3327-3333.
- [17] REINERT T, BARRIOS CH. Overall survival and progression-free survival with endocrine therapy for hormone receptor-positive, HER2-negative advanced breast cancer: review [J]. *Ther Adv Med Oncol*, 2017, 9(11):693-709.
- [18] DIERAS V, MILES D, VERMA S, et al. Trastuzumab emtansine versus capecitabine plus lapatinib in patients with previously treated HER2-positive advanced breast cancer (EMILIA): a descriptive analysis of final overall survival results from a randomised, open-label, phase 3 trial[J]. *Lancet Oncol*, 2017, 18(6):732-742.
- [19] HOPKINS AM, ROWLAND A, LOGAN JM, et al. Primary predictors of survival outcomes for HER2-positive advanced breast cancer patients initiating ado-trastuzumab emtansine [J]. *Breast*, 2019, 46:90-94.

(收稿日期:2021-05-21,修回日期:2021-07-12)