

基于LASSO回归构建乳腺癌患者新辅助化疗疗效的预测模型及列线图

赵晓晖, 傅腾超, 吴 珺*

青岛大学附属医院乳腺病诊疗中心, 山东 青岛

收稿日期: 2022年10月14日; 录用日期: 2022年11月10日; 发布日期: 2022年11月17日

摘 要

目的: 探索乳腺癌新辅助化疗反应的影响因素, 通过LASSO回归建立新辅助化疗疗效的临床预测模型, 指导临床上诊疗方案的选择。方法: 回顾性分析2020年1月~2021年12月435例乳腺癌新辅助化疗患者的临床病理资料, 根据新辅助化疗后的病理反应分为有效组和无效组。采用 χ^2 检验对两组的临床病理指标进行单因素分析; 将有统计学意义的指标纳入LASSO回归分析, 筛选出与疗效相关的显著变量并由此构建新辅助化疗疗效的临床预测模型。应用受试者工作特性曲线(receiver operating characteristic curve, ROC)评价该模型的预测性能。结果: LASSO回归筛选出了月经状态、BMI、肿瘤大小、ER状态、HER2状态、分子分型及四周期疗效评价七个预测变量, 基于上述七个变量绘制列线图及ROC曲线, 训练集和验证集的ROC曲线下面积(area under curve, AUC)分别为0.858和0.798, 内部验证显示列线图具有较好的预测能力。结论: 乳腺癌新辅助化疗有效的综合预测模型对新辅助化疗疗效有较好的预测能力, 此模型可为指导患者个体化的治疗提供参考。

关键词

乳腺癌, 新辅助化疗, LASSO, 预测模型, 列线图

Establishment of a Prognostic Model and Nomogram in Breast Cancer Based on LASSO

Xiaohui Zhao, Tengchao Fu, Li Wu*

Breast Center, The Affiliated Hospital of Qingdao University, Qingdao Shandong

Received: Oct. 14th, 2022; accepted: Nov. 10th, 2022; published: Nov. 17th, 2022

*通讯作者 Email: wuliqd@163.com

文章引用: 赵晓晖, 傅腾超, 吴珺. 基于 LASSO 回归构建乳腺癌患者新辅助化疗疗效的预测模型及列线图[J]. 临床医学进展, 2022, 12(11): 10290-10298. DOI: 10.12677/acm.2022.12111484

Abstract

Objective: To explore the influencing factors of neoadjuvant chemotherapy for breast cancer, we establish a clinical prediction model and a nomogram based on LASSO for the efficacy of neoadjuvant chemotherapy, to guide the choice of clinical operation. **Methods:** By a retrospective analysis of the clinicopathological data of 435 patients with neoadjuvant chemotherapy for breast cancer who were newly treated from January 2020 to December 2021, according to the pathological response after neoadjuvant chemotherapy, they are divided into effective group and ineffective group. The χ^2 test was used to conduct single-factor analysis on the clinicopathological indicators of the two groups: statistically significant indicators were included in the LASSO algorithm, LASSO regression was used to screen the predictors, and a clinical prediction model of neoadjuvant chemotherapy efficacy was constructed. The receiver operating characteristic curve (ROC) was used to evaluate the predictive performance of the model. **Results:** LASSO regression screened seven indexes such as menstrual status, BMI, tumor size, ER status, HER2 status, molecular classification and four cycles of RECIST1.1 efficacy evaluation. Based on the above seven variables, the nomogram and ROC curve were drawn. The areas under the ROC curve (AUC) are 0.858 and 0.798 for the training and testing groups. Internal verification showed that the nomogram had high predictive ability. **Conclusion:** The comprehensive prediction model of effective response of breast cancer after neoadjuvant chemotherapy has a good ability to predict the response of breast cancer after neoadjuvant chemotherapy, which can provide a reference for guiding patients' individualized treatment.

Keywords

Breast Cancer, Neoadjuvant Chemotherapy, LASSO, Prediction Model, Nomogram

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

新辅助化疗(neoadjuvant chemotherapy, NAC)作为全身性治疗方法之一,是指在手术治疗前进行的全身性、系统性的细胞毒性药物治疗[1]。新辅助化疗的优势在于对肿瘤及受累淋巴结进行降期治疗,使保乳治疗成为可能,同时对于化疗反应的监测使其有机会更换治疗方案[2]。虽然 NAC 能够使大多数患者从中获益,但是仍有 10%~30%的患者对 NAC 不敏感,可以出现耐药和疾病的进展,延误治疗的时机,失去对疾病的永久控制[3]。乳腺癌新辅助化疗疗效评价的金标准是术后病理,即病理评估系统 Miller-payne (MP 系统),该评价体系依据新辅助化疗后肿瘤细胞密度减少的比例,将化疗反应分为 5 级,在疗效判定方面,1~2 级被视为化疗无效,3~5 级被视为化疗有效。在新辅助化疗过程中,影像学等无创可重复检查为新辅助化疗疗效的评估提供了依据[4]。因此,为进一步明确乳腺癌患者新辅助治疗的效果及预后,并减少不当治疗引起的不良事件[4],如化疗药物引起的心脏损害、骨髓抑制、肝功能损害、胃肠道反应、手足综合征等,联合影像学评估寻找合适的临床指标预测新辅助治疗的疗效是必要的[5]。

列线图可将复杂的回归方程转化为简单直观的图形,在医学研究和临床实践中得到越来越多的重视和应用[6] [7]。本研究通过回顾性分析 435 例 NAC 患者的临床资料、术前病理资料、术后病理资料及影像资料,从而建立一个综合病理资料和临床资料的列线图预测模型,此模型旨在为合理确定诊治及手术

方案、减少手术后相关并发症等提供临床上的指导。

2. 资料与方法

2.1. 研究对象

回顾性分析 2020 年 1 月~2021 年 12 月初治的 II、III 期乳腺癌新辅助化疗患者的临床病理资料, 依据纳入和排除标准, 筛选出符合标准的患者共 435 例。纳入标准: 1) 新辅助化疗前在超声引导下行乳腺肿物穿刺活检术, 病理确诊为乳腺癌, 并行免疫组化检查; 2) 全部患者在新辅助化疗前未行任何肿瘤相关治疗; 3) 新辅助化疗至少完成 4 周期, 并且在化疗结束后和手术后均评估化疗疗效; 4) 化疗前、化疗过程中和手术前均行乳腺超声检查; 5) 具有完整的临床病理资料; 6) 无严重的全身性疾病, 能够耐受新辅助化疗和手术治疗。排除标准: 1) 同时合并有其他部位恶性肿瘤; 2) 隐匿性乳腺癌、炎性乳腺癌、双侧乳腺癌、哺乳期和妊娠期乳腺癌; 3) 新辅助化疗期间同时接受了内分泌治疗; 4) 男性乳腺癌; 5) 伴有远处转移者。本研究获得本院伦理委员会同意, 所有患者均签署知情同意书。

2.2. 研究方法

通过查阅病例, 采集患者的临床病例信息: 年龄、月经状态、肿瘤大小、淋巴结状态、化疗周期、体重指数(Body Mass Index, BMI)、穿刺病理及术后病理(病理类型、免疫组化指标等)、化疗期间影像学检查结果(乳腺彩超、乳腺 MRI、骨扫描等)等内容。比较两组(有效组和无效组)患者的临床指标(年龄、月经状态、肿瘤大小、化疗周期、化疗方案等)和病理指标(ER、PR、HER-2、Ki-67、分子分型)及 4 周期后影像评估(RECIST 标准)是否存在差异。

乳腺癌新辅助化疗的临床疗效评价参考的是实体瘤治疗疗效评价标准(response evaluation criteria in solid tumors, RECIST) [8], 根据化疗开始前和化疗结束后所测量的肿瘤最大径之和的变化情况分为以下几类: 完全缓解(complete response, CR): 所有目标病灶消失, 无新病灶出现, 并维持 4 周以上; 部分缓解(partial response, PR): 所有目标病灶最长径总和减少 $\geq 30\%$, 并维持 4 周以上; 疾病稳定(stable disease, SD): 所有目标病灶最长径总和缩小但未达 PR, 或增大但未达 PD; 疾病进展(progressive disease, PD): 所有目标病灶最长径总和增大 $\geq 20\%$, 或出现新的病灶。将 CR 和 PR 归为有效, SD 和 PD 归为无效。

2.3. 统计学分析

本次数据分析采用 SPSS 25.0 统计软件和 R 软件(3.6.1 版本)。计数资料采用率(%)表示, 单因素分析采用 χ^2 检验及 Fisher 确切概率法, 多因素分析用 LASSO 回归分析法筛选出与疗效独立相关的显著变量, 根据入选的变量构建预测模型。检验水准 $\alpha = 0.05$ 。将 LASSO 回归分析筛选出的显著变量引入 R 软件, 完成列线图预测模型构建。利用受试者工作特性曲线(receiver operating characteristic curve, ROC)评估预测模型对 NAC 疗效的诊断价值。通常情况下, 曲线下面积(area under curve, AUC) > 0.75 说明模型预备较好的预测能力。

3. 结果

3.1. 患者临床特征和病理特征描述及单因素分析

单因素分析结果显示, 新辅助化疗疗效与年龄($P = 0.001$)、月经状态($P = 0.001$)、BMI ($P < 0.001$)、肿瘤大小($P = 0.049$)、ER 状态($P = 0.001$)、PR 状态($P = 0.002$)、HER 状态($P < 0.001$)、分子分型($P < 0.001$)及四周期疗效评价($P < 0.001$)有关, 而与淋巴结状态($P = 0.571$)、组织学分级($P = 0.095$)、化疗周期($P = 0.593$)及 Ki-67 ($P = 0.077$)无关, 见表 1。

Table 1. Univariate analysis of breast cancer response to neoadjuvant chemotherapy n (%)
表 1. 新辅助化疗后乳腺癌患者疗效的单因素分析 n (%)

	无效	有效	χ^2	P
年龄(岁)			10.662	0.001
≤50	44 (38.6)	181 (56.4)		
>50	70 (61.4)	140 (43.6)		
月经状态			11.524	0.001
已绝经	45 (39.5)	186 (57.9)		
未绝经	69 (60.5)	135 (42.1)		
BMI (kg/m ²)			19.425	<0.001
<24	31 (27.2)	164 (51.1)		
≥24	83 (72.8)	157 (48.9)		
淋巴结状态			0.388	0.571
阴性	18 (15.8)	59 (18.4)		
阳性	96 (84.2)	262 (81.6)		
肿瘤大小			7.841	0.049
cT1	9 (7.9)	28 (8.7)		
cT2	49 (43.0)	181 (56.4)		
cT3	30 (36.3)	66 (20.6)		
cT4	26 (22.8)	46 (14.3)		
化疗周期			1.037	0.593
≤6	63 (55.3)	192 (59.8)		
7~8	48 (42.1)	118 (36.8)		
≥9	3 (2.6)	11 (3.4)		
ER 状态			12.334	0.001
阴性	24 (21.1)	136 (39.3)		
阳性	90 (78.9)	195 (60.7)		
PR 状态			9.560	0.002
阴性	37 (32.5)	158 (49.2)		
阳性	77 (67.5)	163 (50.8)		
HER 状态			38.906	<0.001
阴性	101 (88.6)	180 (56.1)		
阳性	13 (11.4)	141 (43.9)		
Ki-67			3.212	0.077
低表达(<14%)	55 (48.2)	124 (38.6)		
高表达(≥14%)	59 (51.8)	197 (61.4)		

Continued

分子分型			39.414	<0.001
Luminal A 型	43 (37.7)	46 (14.3)		
Luminal B 型	48 (42.1)	164 (51.1)		
HER-2 过表达型	3 (2.6)	61 (19.0)		
三阴性	20 (17.5)	50 (15.6)		
组织学分级			4.690	0.095
I 级	5 (4.4)	6 (2.3)		
II 级	73 (64.0)	198 (74.4)		
III 级	36 (31.6)	62 (23.3)		
NA	65 (14.6)			
四周期疗效评价				
CR	0 (0.0)	20 (6.2)	64.442	<0.001
PR	44 (38.6)	221 (68.8)		
SD	54 (47.4)	77 (24.0)		
PD	16 (14.0)	3 (0.9)		

3.2. 研究因素的 LASSO 回归分析及回归模型建立

利用 LASSO 回归对单因素分析中差异有统计学意义的年龄、月经状态、BMI、肿瘤大小、ER 状态、PR 状态、HER2 状态、分子分型及四周期疗效评价 9 个变量进行筛选, 得到交叉验证的 LASSO 拟合 MSE (图 1) 及 LASSO 拟合系数轨迹图(图 2), 最终纳入模型的变量为月经状态、BMI、肿瘤大小、ER 状态、HER2 状态、分子分型及四周期疗效评价。所构建模型如下: $P = 6.804 + 0.865 * \text{月经状态} - 0.899 * \text{BMI} - 0.692 * \text{肿瘤大小} + 0.138 * \text{ER} - 2.282 * \text{HER2} - 0.241 * \text{分子分型} - 0.199 * \text{四周期疗效评价}$ 。

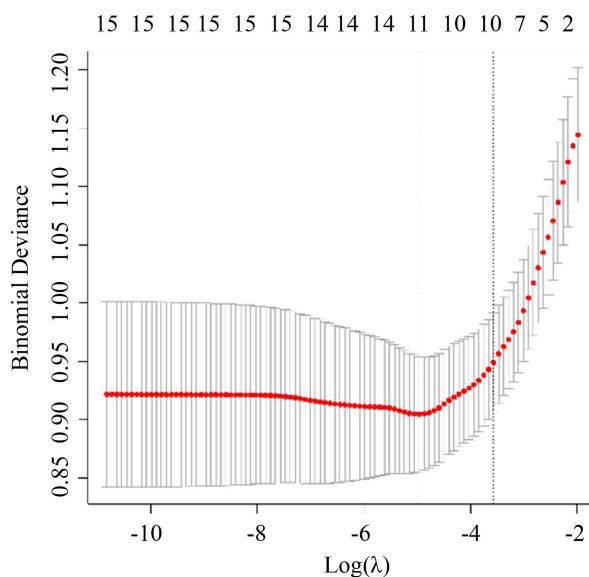


Figure 1. The regression fit of LASSO MSE

图 1. LASSO 回归拟合 MSE

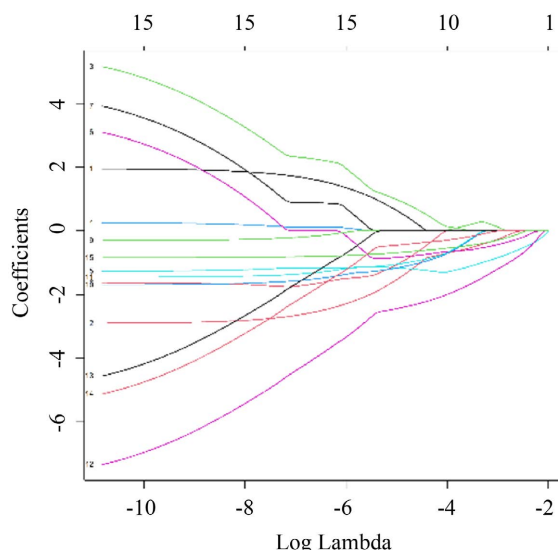


Figure 2. Fitting coefficient locus

图 2. 拟合系数轨迹图

3.3. 列线图预测模型的建立

根据 LASSO 分析结果，使用 R 软件构建了乳腺癌患者新辅助化疗有效的列线图模型(图 3)。列线图是 LASSO 回归分析的可视化过程，参照每个危险因素上方的标准量尺(points)，可以得到该因素的分值，总分为各项分数相加，从总分向下一一对应，可取到对应的乳腺癌患者新辅助化疗有效的概率。通过对每个相关因素单项评分相加得到总分(total points)，总分所对应的值即为模型新辅助化疗有效的发生概率。总分越大代表新辅助化疗有效的概率越高。以此方法，根据每个患者的月经状态、BMI、肿瘤大小、ER、HER2、分子分型及四周期疗效评价的特征，可得到相应的新辅助化疗有效的概率。

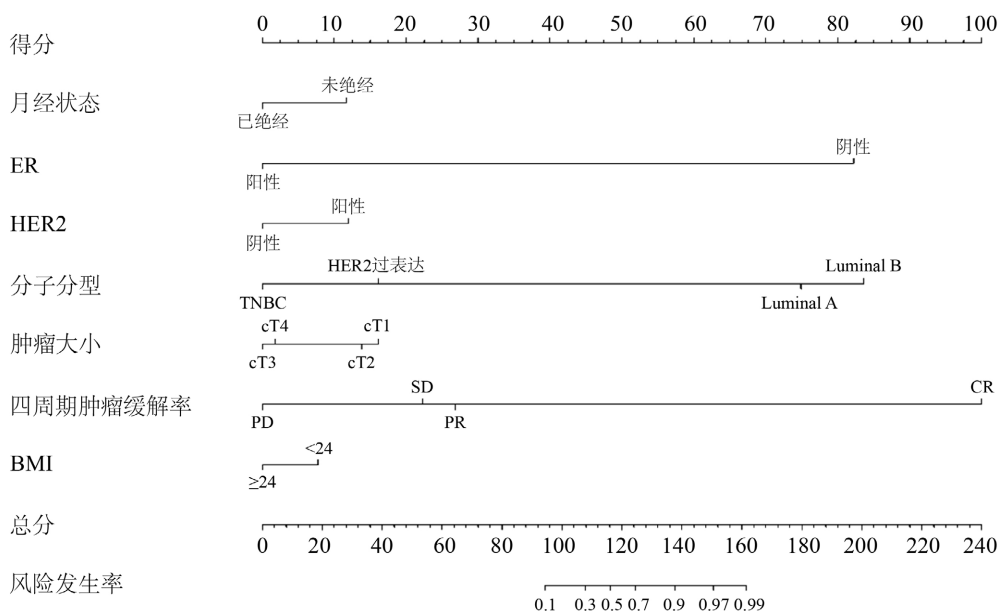


Figure 3. Nomogram was constructed according to predictive factors

图 3. 基于预测因子建立列线图

3.4. 列线图预测模型的评价

通过绘制 ROC 曲线，对列线图模型的预测能力进行评价(图 4)，结果显示：训练集 AUC 为 0.858，截断值为 0.757，敏感度为 73.57%；特异度为 83.33%；内部验证采用 Bootstrap 方法，从原始数据随机抽取同样样本量的数据，进行 1000 次 Bootstrap 重复，结果显示：测试集 AUC 为 0.798，截断值为 0.608，敏感度为 91.50%，特异度为 58.33% (图 5)，结果表明预测模型具有较好的预测性能。列线图预测模型的 C-index 值为 0.794，说明模型具有较好的分辨率。

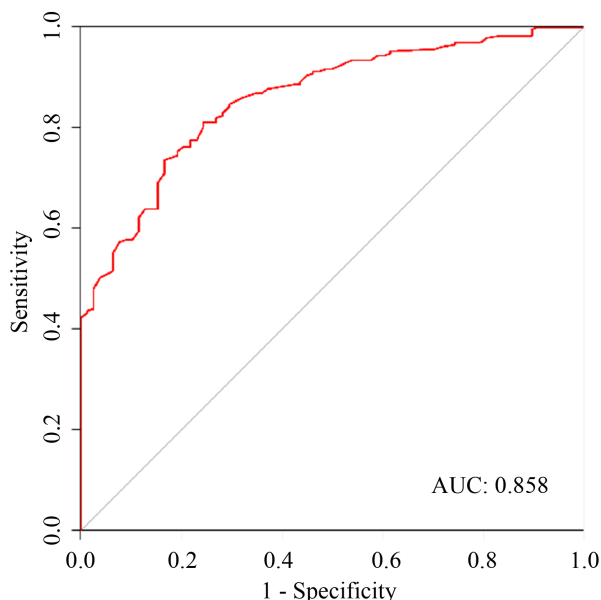


Figure 4. ROC curve of training group
图 4. 基于训练集绘制 ROC 曲线

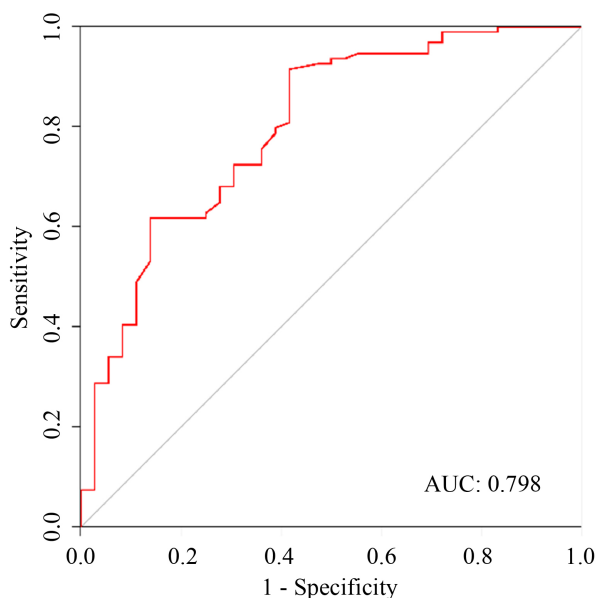


Figure 5. ROC curve of testing group
图 5. 基于测试集绘制 ROC 曲线

4. 讨论

本研究中单因素及 LASSO 回归分析结果均提示月经状态、BMI、肿瘤大小、ER 状态、HER 状态、分子分型及四周期疗效评价与 NAC 后的化疗反应相关, 具有统计学意义($P < 0.05$)。因此这七个因素均可作为 NAC 后为新辅助化疗疗效的预测因素。模型的 AUC 及 C-index 证明模型的区分度较好, 模型拟合优度较高, 说明纳入这七个指标有意义能作为临床参考。

新辅助化疗在诊治乳腺癌的过程中优势在于其使降期治疗成为可能, 同时为有意愿情况未达标的患者提供保乳的可能性。目前评价 NAC 疗效的方法主要包括临床评估和病理学评估。病理完全缓解(pCR)被认为是 NAC 最理想的结局, NAC 后获得 pCR 的乳腺癌患者将具有更好的生存预后, 尤其是三阴性和 HER-2 阳性乳腺癌患者[9]。虽然病理检查是衡量 NAC 反应的金标准, 但这需要在 NAC 后的肿瘤手术切除之后进行, 对 NAC 疗效的评价具有滞后性, 无法早期判断患者对药物的敏感性而及时调整化疗方案。而临床评估有助于在 NAC 早期评估化疗反应, 识别无反应患者, 为选择替代治疗方案提供有用的信息, 避免对无反应的患者产生不必要的药物毒性, 延误治疗时机。有研究表明, 肥胖可增加雌激素及胰岛素水平, 促进微小肿瘤细胞向远处转移, 从而影响乳腺癌患者的预后[10]。激素受体状态是乳腺癌新辅助治疗后 pCR 的独立生物学指标[11]。既往的研究显示, 雌激素低表达与高 pCR 率相关, ER 阳性/HER-2 阴性者的 pCR 率不到 20%, 而三阴型患者的 pCR 率可达 30%~40% [12]。在新辅助化疗过程中, 不同分子分型的乳腺癌患者对化疗的敏感性也相差较多[13]。周怡君等人[14]对 563 例接受 4~8 个周期新辅助化疗的乳腺癌患者进行了研究, 结果激素受体阴性/HER-2 阴性亚型(HR-/HER-2-)、HER-2+亚型和 HR+/HER-2-亚型患者的 pCR 率分别为 38.9%、17.9%和 8.3%, 差异有统计学意义($P < 0.001$); 多因素分析显示, 分子亚型是 pCR 的独立影响因素, 并且在 HR+/HER-2-亚型中, Ki-67 高表达组($Ki-67 \geq 20\%$)的 pCR 率较高($P = 0.004$); 结果表明 HER-2+和 HR-/HER-2-亚型患者更可能对 NAC 有反应。此外, 在 HR+/HER-2-亚型中, Ki-67 高表达的患者往往受益于新辅助化疗。可见, 乳腺癌分子分型可作为 NAC 反应的预测因子, 为我们的治疗决策提供参考。

本研究具有一定的局限性, 首先本研究为回顾性研究。可能具有一定的选择偏倚; 其次为了确保研究准确性, 本研究选取了单发病例; 本研究的研究病例较少, 且为单中心研究, 因此在未来的研究中, 应增加样本量, 并将多发灶患者纳入, 进行多中心、前瞻性研究验证该研究结果的准确性及可靠性。

本研究回顾性分析临床病理相关因素与新辅助治疗后疗效的相关性, 找出与新辅助化疗有效相关的独立预测因子, 建立临床预测模型, 为后续的临床治疗及手术方式的选择提供依据及参考, 同时也为新辅助化疗患者手术后化疗疗效的评估提供了一个可靠的依靠。在对乳腺癌患者行临床干预之前, 可筛选出新辅助治疗最大受益人群, 并进一步预测评估新辅助治疗疗效, 减少不必要的治疗反应及毒性, 进一步优化个体治疗方案。

参考文献

- [1] Haque, W., Verma, V., Hatch, S., et al. (2018) Response Rates and Pathologic Complete Response by Breast Cancer Molecular Subtype Following Neoadjuvant Chemotherapy. *Breast Cancer Research and Treatment*, **170**, 559-567. <https://doi.org/10.1007/s10549-018-4801-3>
- [2] 方依寒, 赵毅. 乳腺癌新辅助化疗后腋窝淋巴结转移综合预测模型的建立[J]. 现代肿瘤医学, 2020, 28(1): 53-57.
- [3] Rao, N.Y., et al. (2017) Significance of Tumor-Infiltrating Lymphocytes and the Expression of Topoisomerase II Alpha in the Prediction of the Clinical Outcome of Patients with Triple-Negative Breast Cancer after Taxane-Anthracycline-Based Neoadjuvant Chemotherapy. *Chemotherapy: International Journal of Experimental and Clinical Chemotherapy*, **62**, 246-255. <https://doi.org/10.1159/000470900>
- [4] Gu, Y.L., Pan, S.M., Ren, J., et al. (2017) Role of Magnetic Resonance Imaging in Detection of Pathologic Complete Remission in Breast Cancer Patients Treated with Neoadjuvant Chemotherapy: A Meta-Analysis. *Clinical Breast*

- Cancer*, **17**, 245-255. <https://doi.org/10.1016/j.clbc.2016.12.010>
- [5] Garutti, M., Pelizzari, G., Bartoletti, M., *et al.* (2019) Platinum Salts in Patients with Breast Cancer: A Focus on Predictive Factors. *International Journal of Molecular Sciences*, **20**, 3390. <https://doi.org/10.3390/ijms20143390>
- [6] Liang, W.H., *et al.* (2015) Development and Validation of a Nomogram for Predicting Survival in Patients with Resected Non-Small-Cell Lung Cancer. *Journal of Clinical Oncology*, **33**, 861-869.
- [7] Sun, W., Jiang, Y.Z., Liu, Y.R., *et al.* (2016) Nomograms to Estimate Long-Term Overall Survival and Breast Cancer-Specific Survival of Patients with Luminal Breast Cancer. *Oncotarget*, **7**, 20496-20506. <https://doi.org/10.18632/oncotarget.7975>
- [8] 杨学宁, 吴一龙. 实体瘤治疗疗效评价标准-RECIST [J]. 循证医学, 2004, 4(2): 85-90+111.
- [9] Spring, L.M., Fell, G., Arfe, A., *et al.* (2020) Pathologic Complete Response after Neoadjuvant Chemotherapy and Impact on Breast Cancer Recurrence and Survival: A Comprehensive Meta-Analysis. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, **26**, 2838-2848. <https://doi.org/10.1158/1078-0432.CCR-19-3492>
- [10] Modugno, F., Kip, K.E. and Cochrane, B. (2007) Obesity, Hormone Therapy, Estrogen Metabolism and Risk of Postmenopausal Breast Cancer. *Breast Diseases*, **17**, 337. [https://doi.org/10.1016/S1043-321X\(07\)80111-1](https://doi.org/10.1016/S1043-321X(07)80111-1)
- [11] Del Prete, S., Caraglia, M., Luce, A., *et al.* (2019) Clinical and Pathological Factors Predictive of Response to Neoadjuvant Chemotherapy in Breast Cancer: A Single Center Experience. *Oncology Letters*, **18**, 3873-3879. <https://doi.org/10.3892/ol.2019.10729>
- [12] Houssami, N., Macaskill, P., Vonminckwitz, G., *et al.* (2012) Meta-Analysis of the Association of Breast Cancer Subtype and Pathologic Complete Response to Neoadjuvant Chemotherapy. *European Journal of Cancer: Official Journal for European Organization for Research and Treatment of Cancer (EORTC) and European Association for Cancer Research (EACR)*, **48**, 3342-3354. <https://doi.org/10.1016/j.ejca.2012.05.023>
- [13] 牛海飞, 魏丽娟, 连镇, 等. 乳腺癌新辅助化疗疗效与分子分型和预后的关系[J]. 中华肿瘤杂志, 2016, 38(3): 190-196. <https://doi.org/10.3760/cma.j.issn.0253-3766.2016.03.006>
- [14] 周怡君, 英旻, 何英剑, 等. 原发性乳腺癌分子分型与新辅助化疗疗效的相关性分析[J]. 中华医学杂志, 2013, 93(22): 1711-1715. <https://doi.org/10.3760/cma.j.issn.0376-2491.2013.22.008>