

# 蚌埠地区妇女乳腺癌风险评估模型的初步研究

李婧婧<sup>1</sup>, 张利<sup>1</sup>, 朋文佳<sup>2</sup>, 程骋<sup>1</sup>

**[摘要]** **目的:** 筛选蚌埠地区妇女患乳腺癌的危险因素, 初步建立符合蚌埠地区的乳腺癌风险评估模型, 探讨乳腺癌低、中、高风险的分界点, 并用模型评估蚌埠地区女性具备特定危险因素情况下患乳腺癌的概率。 **方法:** 采用 logistic 回归筛选乳腺癌的危险因素, 建立乳腺癌风险评估模型。采用 Fisher 判别分析对模型进行评价; 通过 ROC 曲线下面积观察模型效能, 并利用 ROC 曲线寻找乳腺癌低、中、高风险的截断值。 **结果:** 乳腺癌人群的主要危险因素包括家庭经济收入状况、食用豆制品、负面情绪排解、工作强度和乳腺癌筛查 ( $P < 0.05 \sim P < 0.01$ )。通过风险评估模型预测低、中、高危人群, 以预测概率值  $P \leq 0.49$  为低风险性人群,  $P \geq 0.51$  为高风险性人群,  $0.49 < P < 0.51$  为中风险性人群。 **结论:** 建立的模型可评估蚌埠地区女性具备特定危险因素情况下患乳腺癌的风险, 可为建立筛查标准提供一定依据。

**[关键词]** 乳腺肿瘤; 风险评估模型; logistic 回归模型; ROC 曲线

**[中图分类号]** R 737.9

**[文献标志码]** A

**DOI:** 10.13898/j.cnki.issn.1000-2200.2017.04.027

## Preliminary study on the risk assessment model of breast cancer in Bengbu

LI Jing-jing<sup>1</sup>, ZHANG Li<sup>1</sup>, PENG Wen-jia<sup>2</sup>, CHENG Cheng<sup>1</sup>

(1. School of Nursing, 2. School of Public Health, Bengbu Medical College, Bengbu Anhui 233030, China)

**[Abstract]** **Objective:** To screen the risk factors of breast cancer in Bengbu women, establish the risk assessment model of breast cancer, investigate the cut-off points of low, medium and high risk, and assess the probability of breast cancer in women with specific risk factor. **Methods:** The risk factors of breast cancer were screened by logistic regression, and which was used to establish the risk assessment model of breast cancer. The risk model was evaluated using Fisher. The model performance was examined by the area under the ROC curve, and the appropriate low, medium and high risk cut-off points were distinguished by the ROC curve. **Results:** The family economy condition, eating bean product, releasing negative emotion, work intensity and screening breast cancer were the main risk factors of breast cancer ( $P < 0.05$  to  $P < 0.01$ ). The risk model showed that the women with  $P \leq 0.49$ ,  $0.49 < P < 0.51$  and  $P \geq 0.51$  were low, medium and high risk populations, respectively. **Conclusions:** The model can assess the probability of breast cancer in Bengbu women with specific risk factor, and provide certain basis in establishing the screening criteria.

**[Key words]** breast neoplasms; risk assessment model; logistic regression model; ROC curve

乳腺癌是世界范围内女性最多发的恶性肿瘤<sup>[1]</sup>。我国每年的乳腺癌发病率以 3% ~ 4% 速度增加, 超出世界平均水平 1% ~ 2%<sup>[2]</sup>。在北京、上海等一线城市, 乳腺癌已成为危害女性健康第一位的恶性肿瘤。随着生活水平的提高, 生活方式、饮食结构和居住环境等改变, 乳腺癌的危险因素也在不断变化, 并呈现出地区性差异。评价和分析地区性的乳腺癌危险因素, 筛选高危人群, 有利于本地的乳腺癌早期防治工作。国外已开展一系列关于乳腺癌风险评估工具的研究, 建立了数个模型来评价乳腺

癌风险度, 如 Gail 模型、IBIS 模型等, 其主要研究对象是白种人群, 而适用我国的乳腺癌风险评估工具的研究尚少。我们基于蚌埠医学院第一、二附属医院和蚌埠市第三人民医院三家三级甲等医院的优势和资源, 开展蚌埠地区乳腺癌危险因素的调查研究, 初步制定符合蚌埠地区特点的女性乳腺癌风险评估模型。现作报道。

### 1 对象与方法

1.1 研究对象 本研究项目主要针对乳腺癌病人和健康女性。乳腺癌组(观察组)纳入标准: (1) 近 2 年内确诊, 知情同意并能够参加调查, 排除其他部位肿瘤及乳腺癌复发或转移病人; (2) 蚌埠市本地居民; (3) 年龄  $\geq 18$  岁, 无精神疾患, 能进行有效沟通。根据纳入标准选取 2015 年 3 - 11 月蚌埠医学院第一、二附属医院和蚌埠市第三人民医院 3 家医院肿瘤外科和普外科住院并经病理确诊的原发性乳

[收稿日期] 2015-11-07

[基金项目] 安徽省教育厅自然科学基金项目(KJ2009B136Z); 蚌埠医学院研究生科研创新计划项目(Byycxz1410)

[作者单位] 蚌埠医学院 1. 护理学院, 2. 公共卫生学院, 安徽 蚌埠 233030

[作者简介] 李婧婧(1986 -), 女, 硕士研究生。

[通信作者] 张利, 硕士研究生导师, 副教授。E-mail: 390026566@qq.com

腺癌病人 71 例, 年龄 27 ~ 62 岁。健康女性(对照组)纳入标准:(1)年龄与病例相近( $\pm 2$ 岁);(2)同一时期在同一家医院体检中心健康体检的女性,知情同意;(3)蚌埠市本地居民。对照组亦在三家医院选取,共 91 名,年龄 25 ~ 60 岁。

1.2 方法 调查内容包括(1)基本情况:年龄、职业、体质指数、家庭收入情况等;(2)生物遗传学因素:乳腺癌家族史、乳腺良性疾病史等;(3)生殖因素:初潮年龄、绝经年龄、流产次数等;(4)行为生活方式:吸烟、喝酒、体育锻炼等;(5)饮食习惯:豆制品、奶及奶制品、油炸烧烤类食品、蔬菜水果等的摄入;(6)睡眠情况:失眠、熬夜等;(7)心理因素:工作压力、负性情绪等;(8)环境因素:居住环境、染发等;(9)疾病认知及筛查行为:是否知道乳腺癌、乳房定期筛查等。

1.3 质量控制 调查结束对资料仔细检查,对缺项、漏项、填错或其他填写不明情况及时更正。

1.4 统计学方法 采用单因素和多因素 logistic 回归、Fisher 判别分析及 ROC 曲线下面积分析和双纵轴寻找截断点法。

## 2 结果

2.1 乳腺癌病人危险因素筛选 筛选出的乳腺癌病人主要危险因素有:家庭收入状况、食用豆制品、负性情绪排解和乳腺癌筛查( $P < 0.05 \sim P < 0.01$ ) (见表 1)。

表 1 乳腺癌病人危险因素多因素 logistic 回归分析

变量	B	SE	Wald	P	OR	95% CI
家庭收入状况	1.366	0.492	7.727	<0.05	3.921	1.496 ~ 10.274
食用豆制品	-0.995	0.495	4.037	<0.05	0.370	0.140 ~ 0.976
负性情绪排解	-1.454	0.485	8.969	<0.05	0.234	0.090 ~ 0.605
乳腺癌筛查	-2.676	0.788	11.542	<0.01	0.069	0.015 ~ 0.322
常量	23.004	10.441.379	0.000	0.998	0.050	—

2.2 建立乳腺癌风险评估模型 结合专业意义和 OR 值精简多因素条件 logistic 回归筛选出的变量,建立新的乳腺癌风险评估模型: $P = 1/(1 + y)$ ;  $y = \exp[-(23.004 + 1.366 \text{ 家庭收入状况} - 0.995 \text{ 豆制品} - 1.454 \text{ 负性情绪排解} - 2.676 \text{ 乳腺癌筛查})]$ 。其中豆制品和乳腺癌筛查等因素为保护性因素。通过此模型可计算某妇女患乳腺癌的可能性。

2.3 模型评价 将乳腺癌风险评估模型中的变量纳入 Fisher 判别分析,采用交叉验证观察此模型的

判别效能,结果显示该判别函数灵敏度为 87.3%,特异度为 92.3%,总符合率 90.5% (见表 2)。

表 2 模型判别效果的 Fisher 检验 [ $n$ ;百分率(%)]

分组	预测组别		合计
	观察组	对照组	
观察组	62 (87.3)	9 (12.7)	71 (100.0)
对照组	7 (7.7)	84 (92.3)	91 (100.0)
合计	69 (42.6)	93 (57.4)	162 (100.0)

2.4 乳腺癌危险人群的分界点 通过 ROC 曲线提供的灵敏度和特异度变化 (见图 1),同时请教乳腺医生,并结合临床知识,所获量表得出女性乳腺癌危险度的评价标准见表 3。预测概率值  $P \leq 0.49$  为低危险人群, $P \geq 0.51$  为高危险人群, $0.49 < P < 0.51$  为中危险人群。

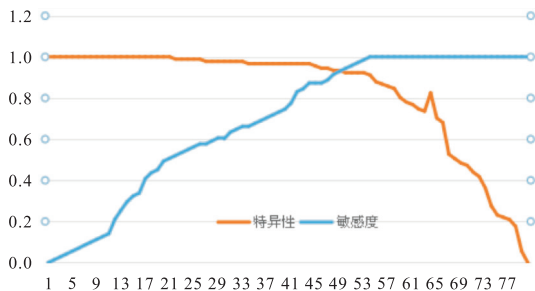


图 1 敏感度/特异性随预测概率的变化趋势

表 3 危险人群最佳分界点及评价指标

最佳截断点	方法	预测可能性	敏感度	特异性
低、中风险截断点	最大灵敏度、特异度	0.49	0.93	0.93
中、高风险截断点	预设定特异性 = 0.95	0.51	0.94	0.95

2.5 模型的诊断效能分析 通过拟合 ROC 曲线,计算模型的 ROC 曲线下面积,观察此模型的效能。其曲线下面积为 0.981,标准误 0.009, $P < 0.01$ ,95% CI 为 0.964 ~ 0.998 (见图 2)。

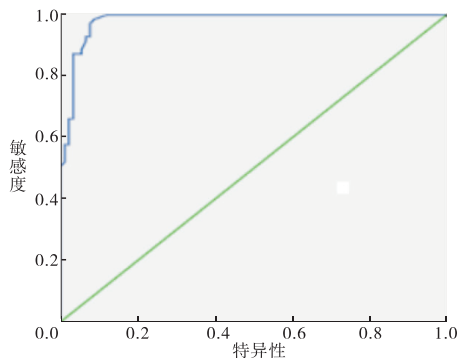


图 2 乳腺癌风险评估模型的 ROC 曲线下面积

### 3 讨论

目前,我国乳腺癌的发病率逐年增长并呈年轻化趋势,引起广泛关注。各个地区的乳腺癌主要危险因素略有不同,因此,开展本地区的乳腺癌风险评估模型研究,对于蚌埠地区女性乳腺癌的防治工作有重要的现实意义。

本研究在前人研究的基础上,通过咨询乳腺外科专家、统计学专家意见,以影响人类健康的四大危险因素为理论框架设计问卷,对蚌埠地区三家三甲医院及体检中心进行调查,经人口学资料分析年龄和婚姻状况无显著差异后,通过单因素和多因素 logistic 回归,筛选出对乳腺癌的主要危险因素,其中家庭收入状况好、常食豆制品、负性情绪能排解和乳腺癌筛查为乳腺癌的保护因素( $P < 0.05 \sim P < 0.01$ )。有研究<sup>[3]</sup>显示,经历生活困境等严重生活事件的妇女患乳腺癌的相对危险性提高 2~3 倍。本研究结果亦显示,家庭收入状况差者患乳腺癌风险增加 13 倍。同时,乳腺癌是基因与环境因素互相作用的产物,可以通过健康的行为生活方式来降低其风险,从而降低发病率。例如本研究结果显示,饮食习惯中常食豆制品为乳腺癌的保护因素,且摄入频率与乳腺癌的发生存在剂量反应关系。这与豆类食品中含有大豆异黄酮,作为植物性雌激素可以与人体雌激素竞争雌激素受体有关<sup>[4]</sup>,而豆制品的中高摄入量对中国女性是有利的<sup>[5]</sup>。

通过分析蚌埠地区乳腺癌的影响因素,本研究初步建立了符合本地区的乳腺癌风险评估模型,通过这个模型界定了乳腺癌低、中、高风险人群的分界点,由该模型可以评估出某一女性患乳腺癌的可能性。假设某女性具备了一些危险因素,代入模型计算得出其患乳腺癌的概率为 0.7,则根据本研究风险分界点,该名女性属于乳腺癌高危人群( $P > 0.51$ )。国外通过分析人群的遗传因素和人口学信息判断女性患乳腺癌风险已经应用于临床,例如众所周知的美国 Gail 模型应用最为广泛,该模型针对白人建立,对亚洲女性有一定的局限性<sup>[6]</sup>。纳入基因因素的模型如 IBIS 模型<sup>[7]</sup>,被用于评估携带易感基因 BRCA1/2 突变的可能性及 BRCA1 的恶性进程和良性表达<sup>[8]</sup>,集合了广泛而深度的危险因素,但

因需要昂贵的基因检测而使该模型无法广泛推广<sup>[9]</sup>。本研究中,乳腺癌风险评估模型的 ROC 曲线下面积达 0.981,提示该模型诊断效能较好,可在对蚌埠地区女性进行乳腺癌风险评估时进行应用。而对于本模型判定为高风险的人群应早期进行健康教育,强调每月 1 次的乳房自检,并对钼靶检查的年龄、频率以及积极健康的生活行为方式做普及和宣教。

综上,本研究通过观察蚌埠地区女性发生乳腺癌的危险因素,初步建立了乳腺癌风险评估模型,有助于乳腺癌的防治。但由于时间、精力和成本的限制,本研究调查的样本量尚少,不能对疾病发生的远期影响进行评价,今后仍需扩大调查范围,以完成对该模型的验证和校正。

#### [ 参 考 文 献 ]

- [1] 郑莹,吴春晓,吴凡. 中国女性乳腺癌死亡现况和发展趋势[J]. 中华预防医学杂志,2011,45(2):150.
- [2] JEMAL A, BRAY F, CENTER MM, *et al.* Global cancer statistics [J]. CA Cancer J Clin, 2011, 61(2):69.
- [3] 荣素英,李君,张焯. 饮食因素与女性乳腺癌关系的病例对照研究[J]. 环境与健康杂志,2008,25(4):337.
- [4] LIU XO, HUANG YB, GAO Y, *et al.* Association between dietary factors and breast cancer risk among Chinese females: a systematic review and meta analysis [J]. Asian Pac J Cancer Prev, 2014, 15(3):1291.
- [5] WANG F, YU L, WANG F, *et al.* Risk factors for breast cancer in women residing in urban and rural areas of eastern China [J]. J Int Med Res, 2015, 43(6):774.
- [6] AMIR E, FREEDMAN OC, SERUGA B, *et al.* Assessing women at high risk of breast cancer: a review of risk assessment models [J]. J Natl Cancer Inst, 2010, 102(10):680.
- [7] TYRER J, DUFFY SW, CUZICK J. A breast cancer prediction model incorporating familial and personal risk factors [J]. Stat Med, 2004, 23(7):1111.
- [8] FEILOTTER HE, MICHEL C, UY P, *et al.* BRCA1 haplo insufficiency lead to altered expression of genes involved in cellular proliferation and development [J]. PLoS One, 2014, 9(6):e100068.
- [9] MOOK S, SCHMIDT MK, VIALE G, *et al.* The 70-gene prognosis-signature predicts disease outcome in breast cancer patients with 1-3 positive lymph nodes in an independent validation study [J]. Breast Cancer Res Treat, 2009, 116(2):295.

(本文编辑 卢玉清)