

• 临床研究 •

广东地区妇女乳腺癌预测模型的研究

夏建红 王颀 张安秦 李文萍 许娟 陈中扬 杨剑敏 翁秀琼

【摘要】 目的 建立广东地区妇女乳腺癌预测模型,筛选高危人群,为乳腺癌的早期干预提供线索。**方法** 在门诊患者和正常体检的妇女中,采用简单随机的方法选取 35~60 岁乳腺癌病例 284 例,以 1:2 成组配对原则在相应年龄段内选取非乳腺癌妇女 673 例,进行回顾性病例对照研究。分析数据时,首先将数据按绝经前后分层;以是否为乳腺癌患者为结果变量,其余变量为自变量,分别通过 logistic 单因素、多因素回归分析筛选出绝经前、后乳腺癌的危险因素,并以此建立绝经前、后妇女乳腺癌预测模型;用判别分析法分别对绝经前、后预测模型的预测能力进行判别。**结果** 绝经前妇女患乳腺癌的主要危险因素包括服用避孕药史、亲属中有乳腺癌患者、有不良情绪、有不良事件、劳动强度大、无乳腺组织活检史、很少参加运动等。绝经前妇女患乳腺癌的预测模型为: $1/(1 + \exp[-(-4.314 + 1.333 \text{ 服用避孕药} + 1.167 \text{ 亲属中有乳腺癌患者} + 0.404 \text{ 有不良情绪} + 0.613 \text{ 有不良事件} + 0.754 \text{ 劳动强度大} - 0.716 \text{ 有乳腺组织活检史} - 0.839 \text{ 经常运动})])$ 。此模型有统计学意义($\chi^2 = 165.83, P = 0.000$),并且其预测不发生乳腺癌的正确率为 74.2%,预测发生乳腺癌的正确率为 58.4%。绝经后妇女患乳腺癌的主要危险因素包括亲属中有乳腺癌患者、文化程度低、情绪调节能力差等。绝经后妇女患乳腺癌的预测模型为: $1/(1 + \exp[-(1.211 + 2.269 \text{ 亲属中有乳腺癌患者} - 1.617 \text{ 文化程度高} - 0.635 \text{ 情绪调节能力强})])$ 。此模型有统计学意义($\chi^2 = 116.32, P = 0.000$),并且其预测发生乳腺癌的正确率为 78.9%,不发生乳腺癌的正确率为 55.2%。**结论** 绝经前、后妇女患乳腺癌的影响因素不完全相同。绝经后妇女乳腺癌预测模型对预测发生乳腺癌有一定的参考价值;绝经前妇女乳腺癌预测模型对预测不发生乳腺癌有一定的参考价值。有效的乳腺癌预测模型还需要整合更多的临床和生物学指标。

【关键词】 乳腺肿瘤;预测模型;绝经前;绝经后;判别分析

【中图分类号】 R737.9

【文献标识码】 A

A study of predicting model for breast cancer in Guangdong area XIA Jian-hong, WANG Qi, ZHANG An-qin, LI Wen-ping, XU Juan, CHEN Zhong-yang, YANG Jian-min, WENG Xiu-qiong. Breast Disease Center, Guangdong Provincial Hospital for Woman and Child Health, Guangzhou 510010, China

【Abstract】 Objective To establish a predicting model for breast cancer in order to screen high risk women and provide clue for early intervention of breast cancer in Guangdong. **Methods** A retrospective case-control study was used to identify risk factors of breast cancer. From outpatients and women of physical examination, 284 breast cancer patients aged from 35 to 60 years were randomly selected and 673 non-breast cancer patients at the same age from the same population were selected accord-

作者单位:510010 广州,广东省妇幼保健院乳腺病防治中心

通信作者:王颀, E-mail: wangqigdmch@21cn.com

ing to 1 : 2 paired grouping principle. Breast cancer was dependent variable, and all other variables were independent variables. For data analysis, all cases were firstly divided into a premenopausal group and a postmenopausal group. Simple logistic regression and multivariate logistic regression models were applied to filter the risk factors of breast cancer respectively. Then these important factors were used to establish the premenopausal and postmenopausal risk forecast models respectively. The two models were evaluated by discriminant analysis. **Results** In the premenopausal women group, the main risk factors of breast cancer included medicine-contraception, breast cancer-family history, bad emotion, undesirable life, excessive working, absence of breast biopsy, less exercise, etc. The equation of logistic regression was: $1/\{1 + \exp[-(-4.314 + 1.333 \text{ medicine-contraception} + 1.167 \text{ breast cancer-family history} + 0.404 \text{ bad emotion} + 0.613 \text{ undesirable life} + 0.754 \text{ excessive working} - 0.716 \text{ breast biopsy history} - 0.839 \text{ regular exercise})]\}$. This model was significant in predicting breast cancer ($\chi^2 = 165.83, P = 0.000 < 0.010$). The correct rate of predicting non-breast cancer was 74.2%, and the correct rate of predicting breast cancer was 58.4%. In the postmenopausal group, the main risk factors included breast cancer-family history, low education, and the absence of emotion-adjusting ability. The equations of logistic regression was: $1/\{1 + \exp[-(1.211 + 2.269 \text{ family history} - 1.617 \text{ high education} - 0.635 \text{ high ability of emotion adjustment})]\}$. This model was significant in predicting breast cancer ($\chi^2 = 116.32, P = 0.000$). The correct rate of predicting breast cancer was 78.9%, and the correct rate of predicting non-breast cancer was 55.2%. **Conclusions** There were different influencing factors for breast cancer between premenopausal and postmenopausal women. The predictive model for postmenopausal women has value to predict risks of breast cancer; the predictive model for premenopausal women has value to predict non-breast cancer; however a efficacious predicting model of breast cancer still needs integrate biology and clinic index.

【Key words】 Breast neoplasms; Predicting model; Premenopause; Postmenopause ; Discriminant analysis

乳腺癌已成为女性癌症的第二大死因,是严重威胁妇女生命和健康的恶性肿瘤之一。近年来,广东地区妇女病普查资料显示,乳腺癌的检出率呈逐年上升趋势^[1],而早期预防对改善乳腺癌的预后具有重要意义。为此,本研究分别对广东地区妇女绝经前后乳腺癌人群和非乳腺癌人群的危险因素进行了病例对照研究,建立预测模型,用以早期筛选乳腺癌的高危人群。

1 资料和方法

1.1 资料来源

研究对象来源于 2006 年 5 月至 2007 年 2 月参加正常体检或在乳腺门诊就诊、年龄在 35~60 岁的广东地区妇女。在分别进行了调查员培训和筛查技

术培训后,通过简单随机的方法选取患乳腺癌的妇女 284 例作为病例组,全部经医院组织病理学诊断确诊。按照 1:2 成组配对的原则,在 35~60 岁非乳腺癌妇女随机选取 673 例作为对照组,对两组分别进行了问卷调查。问卷调查的内容包括社会经济人口学特征、生理生育情况、遗传生物学情况、心理精神状况、吸烟与饮酒、日常生活与工作等六个部分的指标。

1.2 研究方法

本研究采用 1:2 成组配对的回顾性病例对照研究。根据国内外文献^[2-4],同时考虑绝经对乳腺癌发生的重大影响,数据分析前,先将数据按照绝经前、后分层,绝经的定义采用 NCCN2008 年第 1 版的定义。绝经前和绝经后妇女乳腺癌的高危因素筛选工作分别从检查每个变量与结果变量之间的二元关系着手,以是否为乳腺癌患者为结果变量,其余变量为自变量,对两组分别进行 logistic 单因素、多因素的回归分析,分别筛选出高危因素,并在此基础上建立绝经前、后妇女乳腺癌的预测模型,并用判别分析法对绝经前、后预测模型的预测能力分别进行判别。

1.3 统计学处理

采用 Epi Data 软件对搜集的数据进行双录入,统一管理;采用 SPSS11.0 软件对数据进行 Logistic 回归分析和判别分析,进入变量标准为 $P=0.100$,剔除变量标准为 $P=0.120$ 。

2 结果

2.1 绝经前妇女患乳腺癌的危险因素及预报模型

2.1.1 绝经前妇女乳腺癌病例组和对照组的一般情况:收集的数据经有无绝经分层后,绝经前妇女共有 747 例,其中乳腺癌病例 161 例,非乳腺癌病例 586 例,两组的平均年龄分别为 (40.61 ± 3.005) 岁、 (40.90 ± 3.037) 岁,二者差异无统计学意义($t=1.091$, $P=0.276$)。

2.1.2 绝经前妇女患乳腺癌的危险因素分析:绝经前妇女患乳腺癌的危险因素经 logistic 单因素回归分析,共筛选出教育程度高、服用过避孕药、无乳腺组织活检史、亲属中有乳腺癌患者、有不良情绪、有不良事件、情绪调节能力差、劳动强度大、很少剧烈运动、很少运动等 10 项危险因素,而初潮年龄、初产年龄、有否母乳喂养、有否流产等因素在病例组和对照组间差异无统计学意义($P>0.050$)。在单因素分析的基础上采用逐步后退法用多变量非条件 logistic 回归分析进一步筛选变量。有 7 种因素与绝经前患乳腺癌有关,其中服用避孕药、亲属中有乳腺癌患者、有不良情绪、有不良事件、劳动强度大是危险因素,有乳腺组织活检史、经常运动是保护因素(表 1)。

表 1 绝经前组患乳腺癌的多因素非条件 Logistic 回归结果

| 影响因素 | B 值 | Wald | P 值 | OR 值(95%CI) |
|--------------------|--------|--------|-------|--------------------|
| 服用避孕药 | 1.333 | 9.184 | 0.002 | 3.791(1.601~8.977) |
| 有乳腺组织活检史 | -0.716 | 7.585 | 0.006 | 0.489(0.294~0.813) |
| 亲属中有乳腺癌患者 | 1.167 | 8.389 | 0.004 | 3.213(1.458~7.079) |
| 有不良情绪 | 0.404 | 4.456 | 0.035 | 1.498(1.029~2.180) |
| 有不良事件 | 0.613 | 8.510 | 0.004 | 1.846(1.223~2.787) |
| 劳动强度大 ^a | 0.754 | 12.725 | 0.000 | 2.126(1.405~3.219) |
| 经常运动 ^b | -0.839 | 11.808 | 0.001 | 0.432(0.268~0.697) |
| 常数项 | -4.314 | 24.006 | 0.000 | 0.013 |

a:劳动强度分为轻度、中度和重度。轻度:以静坐、站立为主,如坐办公室、营业员、学生;中度:如电工、木工、管道工,常温下不易出汗;重度:如建筑、采矿、装卸等,常温下容易出汗。b:经常运动为近 5 年内每周有 3 次以上的中轻度锻炼,如散步、慢跑、扭秧歌、太极拳等。

2.1.3 绝经前妇女乳腺癌预报模型及预测能力判别:绝经前 7 种因素与患乳腺癌的 Logistic 多元回归的概率模型(P)为 $P = 1/(1 + y)$; $y = \exp[-(-4.314 + 1.333 \text{ 服用避孕药} + 1.167 \text{ 亲属中有乳腺癌患者} + 0.404 \text{ 有不良情绪} + 0.613 \text{ 有不良事件} + 0.754 \text{ 劳动强度大} - 0.716 \text{ 有乳腺组织活检史} - 0.839 \text{ 有经常运动})]$

对该模型方程进行显著性检验,结果显示差异有统计学意义($\chi^2 = 165.83, P = 0.000$)。经判别分析,该预测模型对预测可能不发生乳腺癌的正确率为 74.2%,发生乳腺癌的正确率为 58.4%,全组总正确率为 70.7%。

2.2 绝经后妇女患乳腺癌的危险因素及预报模型

2.2.1 绝经后妇女乳腺癌病例组和对照组一般情况:收集的数据经有无绝经分层后,绝经后妇女共 210 例,其中乳腺癌妇女 123 人,非乳腺癌妇女 87 人,两组的平均年龄分别为(56.72±3.417)岁和(56.56±2.876)岁,二者差异无统计学意义($t = 0.357, P = 0.721$);平均绝经年龄分别为(49.82±1.589)岁和(49.55±1.854)岁,二者差异无统计学意义($t = 1.129, P = 0.260$)。

2.2.2 绝经后妇女患乳腺癌的危险因素分析:采用单因素 logistic 回归方法对绝经后妇女患乳腺癌的危险因素进行分析,共筛选出教育程度低、无乳腺组织活检史、亲属中有乳腺癌患者、情绪调节能力差 4 种危险因素,而初潮年龄、初产年龄、绝经年龄、有否母乳喂养、有否流产等因素在病例组和对照组间差异无统计学意义($P > 0.050$)。在单因素分析的基础上,采用逐步后退法用多变量非条件 logistic 回归分析进一步筛选变量,有 3 种因素与绝经后患乳腺癌有关,其中亲属中有乳腺癌患者为危险因素,文化程度高、情绪调节能力强是保护因素(表 2)。

表 2 绝经后组患乳腺癌的多因素非条件 Logistic 回归结果

| 影响因素 | B 值 | Wald | P 值 | OR 值(95%CI) |
|-----------|--------|-------|-------|---------------------|
| 文化程度高 | -1.617 | 4.643 | 0.031 | 0.198(0.046~0.864) |
| 亲属中有乳腺癌患者 | 2.269 | 4.399 | 0.036 | 9.670(1.160~90.591) |
| 情绪调节能力强 | -0.635 | 5.988 | 0.014 | 0.530(0.318~0.881) |
| 常数项 | 1.211 | 3.724 | 0.043 | 1.358 |

2.2.3 绝经后妇女乳腺癌预报模型及预测能力判别:绝经后3种因素与患乳腺癌的 Logistic 多元回归概率模型(p)为 $p=1/(1+y)$; $y=\exp[-(1.211+2.269 \text{ 亲属中有乳腺癌患者}-1.617 \text{ 文化程度高}-0.635 \text{ 情绪调节能力强})]$

对该模型方程进行显著性检验,结果显示差异有统计学意义($\chi^2=116.32, P=0.000$)。经判别分析,该预测模型对预测可能发生乳腺癌的正确率为78.9%,不发生乳腺癌的正确率为55.2%,全组总正确率为72.3%。

3 讨论

乳腺癌的发生是多因素作用的结果。国内外的一些研究均证实乳腺癌的发生与年龄及绝经史有着密切的关系^[2-5]。亚洲地区妇女(以日本为例)发生乳腺癌的高峰年龄大多在40~50岁,约70%的乳腺癌患者年龄小于50岁^[2-5]。绝经前后乳腺癌的危险因素也不相同^[2]。为研究除年龄和绝经史以外因素的影响,更有针对性的做好目标人群的乳腺癌防治工作,本研究选择年龄在35~60岁的妇女,并且分成绝经前组和绝经后组,分别对两组乳腺癌妇女和非乳腺癌妇女进行了病例对照研究。在分析绝经前、后妇女危险因素基础上,分别建立了绝经前和绝经后患乳腺癌风险的预测模型。

3.1 妇女绝经前、后患乳腺癌的危险因素分析

在绝经前妇女中,国内外的许多研究已证实抑郁等有密切的关系^[3-4,6-7]。本研究也发现有服用避孕药史、亲属中有乳腺癌患者、有不良情绪、有不良事件、劳动强度大等因素是广东地区绝经前妇女患乳腺癌的主要危险因素。另外,有研究显示长期的雌激素暴露,如初潮早、绝经晚、行经时间长、月经不规律等可能与乳腺癌有着密切的关系^[4,7-9],而有研究显示初潮年龄、初产年龄、是否哺乳与乳腺癌的发生无相关性^[10]。另外,国外部分研究证实多次行乳腺组织活检增加了患乳腺癌的危险,另一部分研究却证实乳腺组织活检不会增加患乳腺癌的风险^[11]。本研究显示,曾有乳腺组织活检史是患乳腺癌的保护因素,可能是曾有乳腺组织活检史的妇女,能够得到早期的干预而防止了某些乳腺疾病进一步转化为乳腺癌。本研究还发现经常的体育活动能够减少乳腺癌发生,这与国内外的研究一致^[2-4],主要是运动能够减少脂肪的堆积和减重,改善代谢,促进女性内分泌系统的稳定。

在绝经后妇女中,本研究显示亲属中有乳腺癌患者是患乳腺癌的主要危险因素。西方国家的研究表明妇女的肥胖、腹部脂肪堆积、血清雌激素水平较高是西方妇女绝经后的主要危险因素。这与西方女性摄食过多的热量,缺乏体育锻炼导致的异常肥胖有关^[2]。而本研究中未发现绝经后妇女有过度肥胖者,一定范围内体重指数的大小与乳腺癌的发生没有相关性。另外,国外一些学者指出,绝经后体内雌激素水平过高是患乳腺癌的一个危险信号^[2-4]。本

研究因为条件所限没有对绝经后妇女的雌激素水平进行测定,其还有待更进一步的研究。本研究还发现文化程度高、情绪调节能力强是绝经后妇女患乳腺癌的一个保护因素,说明良好的社会心理适应能力对预防绝经后乳腺癌的发生具有重要意义^[6-7]。

综上所述,绝经前、后妇女患乳腺癌的危险因素和保护因素不完全相同,这与国外的有关研究结论一致。但总体而言,不论是绝经前或绝经后的妇女,发生乳腺癌的风险都与个体的社会、心理、行为有着密切的关系。由此说明,在进行乳腺癌早防早治的过程中,都要注意对社会、心理、行为因素的早期干预。

3.2 预测乳腺癌风险的正确率分析

早期筛查出乳腺癌的高危人群,进行早期的干预,对乳腺癌的防治具有重要意义。因此,乳腺癌风险预测模型的研究,一直是研究的热点问题^[12-14]。1989 年,为了给临床咨询提供协助,Dr. Mitchell Gail 研制出以他名字命名的评估乳腺癌患病概率的 Gail 模型。如今,Gail 模型被广泛运用于乳腺癌危险度的预测,它的准确性也在一系列的大规模的人群研究中得到了验证。但是 Gail 模型是在西方妇女人群中得出的模型,不一定适合中国的妇女。本研究通过分析广东地区绝经前后妇女乳腺癌的危险因素,分别建立了乳腺癌风险的预测模型。

对于绝经前的妇女,预测模型总的预测正确率为 70.7%,其中对预测可能不发生乳腺癌的正确率为 74.2%,即对于无避孕药服用史、无乳腺癌家族史、无不良情绪、无不良生活事件、劳动强度适中、有乳腺活检史、有适当的运动等行为因素的妇女,预测其不发生乳腺癌的正确率有一定的参考价值。而预测乳腺癌发生的正确率不高,仅有 58.4%。对于绝经后的妇女,模型总的预测正确率为 72.3%,其中对预测发生乳腺癌的正确率为 78.9%,即如果个体有乳腺癌家族史、文化程度低、情绪调节能力差,预测其发生乳腺癌的正确率也有一定的参考价值,而预测不发生乳腺癌的正确率偏低,仅为 55.2%。目前国内外大多数乳腺癌危险因素的研究主要是人群常规描述性因素及社会心理行为因素,缺少生物学特征性的因素,如血清肿瘤标志物、个体雌孕激素及受体水平以及病理学、影像学特征等^[15]。这可能是此模型预测正确率不高的主要原因之一。因此,还有待进一步结合生物学特征因素及临床实际进行研究,以便为乳腺癌的早期干预提供更充分的线索。

参考文献

- [1] 武秋林,曾芳玲,曾晓琴,等. 广州市 1994~2003 年乳腺癌筛查情况分析. 中国初级卫生保健, 2004,18:52-54.
- [2] Stoll A. Adiposity as a risk determinant for postmenopausal breast cancer. Intl J Obes, 2000,

24: 527—533.

- [3] 刘山, 黄晓蓉, 易瑛, 等. 近 20 年乳腺癌发病的常见危险因素变化分析. 华西医学, 2007, 22: 729—731.
- [4] 戴月, 杜玉开. 女性乳腺癌发病危险因素探析. 中国妇幼健康研究, 2007, 18: 71—74.
- [5] 胡崇珠, 方志沂. 中国乳腺癌普查开展现状. 中华乳腺病杂志: 电子版, 2007, 1: 177—179.
- [6] Price M A, Tennant C, Butow P N, *et al.* The role of psychosocial factors in the development of breast carcinoma: Part II life event stress, social support, defence style, and emotional control and their interactions. *Cancer*, 2001, 91: 686—697.
- [7] Candova M J, Cunningham L L, Carlson C R, *et al.* Social constraints, cognitive processing, and adjustment to breast cancer. *Consult Clin psychol*, 2001, 69: 706—711.
- [8] 代丽萍. 乳腺癌的危险因素分析及预防. 中国初级卫生保健, 2005, 19: 84.
- [9] 戴月, 袁宝君, 史祖民, 等. 乳腺癌影响因素的病例对照研究. 江苏卫生保健, 2006, 8: 4—5.
- [10] Mokbel K, Kirkpatrick K L. Recent advances in breast cancer (the Twenty-fourth San Antonio Breast Cancer Symposium, December, 2001). *Curr Med Res Opin*, 2002, 18: 26—29.
- [11] Ohta K, Dohden K, Morishita M, *et al.* Clinical examination of safety and effectiveness of primary chemotherapy with CEF followed by docetaxel in preoperative breast cancer. *Gan To Kagaku Ryoho*, 2006, 33: 199—202.
- [12] 潘瑞芹, 陈平. 加强对乳腺癌易患人群的研究. 中国普通外科杂志, 2005, 14: 241—242.
- [13] 钟晓妮, 周燕荣. 女性乳腺癌预报模型研究. 数理医药学杂志, 2002, 15: 4—6.
- [14] 宋金霞, 于兰贞, 宁丽霞. 乳腺癌风险因素评估与健康教育的研究. 国外医学·护理学分册, 2004, 23: 246—249.
- [15] 耿翠芝. 与乳腺癌相关的肿瘤标志物研究现状及评价. 中华乳腺病杂志: 电子版, 2009, 3: 6—10.

(收稿日期: 2009-03-02)

(本文编辑: 罗承丽)

夏建红, 王颀, 张安秦, 等. 广东地区妇女乳腺癌预测模型的研究[J/CD]. 中华乳腺病杂志: 电子版, 2009, 3(5): 503—509.