

## 新疆哈萨克族食管癌风险的预测

李婧, 秦江梅

### ■背景资料

食管癌的死亡率不同地区、不同民族差异很大, 新疆哈萨克族是我国食管癌的高发民族。对慢性疾病和复杂性疾病建立风险预测模型, 对高危人群和易感个体进行早期预测、早期诊断、早期治疗是有效的。

李婧, 石河子大学医学院预防医学系 新疆维吾尔自治区石河子市 832002

秦江梅, 卫生部卫生发展研究中心 北京市 100191

李婧, 主要从事肿瘤分子流行病学研究。

国家自然科学基金资助项目, No. 30660161

作者贡献分布: 此课题由秦江梅设计; 数据分析由李婧完成; 本论文写作由李婧与秦江梅完成。

通讯作者: 秦江梅, 教授, 100191, 北京市海淀区学院路38号, 卫生部卫生发展研究中心。 qinjm@nhei.cn

电话: 010-82801398 传真: 010-82805512

收稿日期: 2013-12-02 修回日期: 2014-02-26

接受日期: 2014-02-28 在线出版日期: 2014-04-08

### Prediction of risk for esophageal cancer in Kazak people in Xinjiang

Jing Li, Jiang-Mei Qin

Jing Li, Department of Preventive Medicine, Shihezi University School of Medicine, Shihezi 832002, Xinjiang Uygur Autonomous Region, China

Jiang-Mei Qin, Health Development Research Center of the Ministry of Health, Beijing 100191, China

Supported by: National Natural Science Foundation of China, No. 30660161

Correspondence to: Jiang-Mei Qin, Professor, Health Development Research Center of the Ministry of Health, 38 Xueyuan Road, Haidian District, Beijing 100191, China. qinjm@nhei.cn

Received: 2013-12-02 Revised: 2014-02-26

Accepted: 2014-02-28 Published online: 2014-04-08

### Abstract

**AIM:** To develop risk prediction models for esophageal cancer, and provide evidence and information for the establishment of an esophageal cancer monitoring system for high-risk Kazak population in Xinjiang.

**METHODS:** Data of the local population and controls, including environmental, behavioral risk factors, and genetic data, were collected. Risk factors were screened using conditional Logistic regression to establish Logistic discriminant predictive models.

**RESULTS:** Two groups were selected as controls in this study: one was Kazakh population and the other was general population. Compared with the Kazakh control group, values of

odds ratios (ORs) of related risk factors were as follows: alcohol drinking (OR = 2.77), irregular diet (OR = 3.42), frequent binge eating (OR = 4.01), fruit eating (OR = 2.65), gastropathy history (OR = 2.66), eating speed (OR = 1.94), family history of esophageal cancer (OR = 2.06), HLA-DRB1\*0901 (OR = 2.83), TAP2379 (OR = 2.09), and CYP2E1 (OR = 1.60), and 10 factors were selected into the model. As for the population-based control, values of OR were: age (OR = 1.10), alcohol drinking (OR = 6.27), irregular diet (OR = 118.05), frequently eating hot and burning food (OR = 3.02), frequent binge eating (OR = 2.11), fruit eating (OR = 6.80), smoked meat eating (OR = 17.14), gastropathy history (OR = 5.31), and 7 factors were selected into the model. The discriminant accuracy rates of the four models were 74.6%, 70.1%, 76.1% and 96.9%, respectively, with the discriminant model 4 having the best accuracy.

**CONCLUSION:** The models established with environmental and poor behavioral factors are suitable for basic health care facilities. Self-assessment of risk probability could be implemented using these simple and effective risk prediction models.

© 2014 Baishideng Publishing Group Co., Limited. All rights reserved.

**Key Words:** Esophageal cancer; Risk prediction; Logistic discriminant model

Li J, Qin JM. Prediction of risk for esophageal cancer in Kazak people in Xinjiang. *Shijie Huaren Xiaohua Zazhi* 2014; 22(10): 1442-1445 URL: <http://www.wjgnet.com/1009-3079/22/1442.asp> DOI: <http://dx.doi.org/10.11569/wcjd.v22.i10.1442>

### 摘要

**目的:** 利用不同危险因素建立食管癌风险预测模型, 为新疆哈萨克族食管癌高危人群监测体系的建立提供循证依据和基础信息。

**方法:** 利用前期收集对照和当地人群对照环境、行为危险因素和基因检测数据, 采用条件

■同行评议者  
姜春萌, 教授, 大连医科大学附属第二医院消化科

Logistic回归筛选危险因素, 利用Logistic判别建立预测模型。

**结果:** 以哈萨克族对照, 饮酒(OR = 2.77)、饮食不规律(OR = 3.42)、经常暴饮暴食(OR = 4.01)、少吃水果(OR = 2.65)、胃病变史(OR = 2.66)、进食速度快(OR = 1.94)、食管癌家族史(OR = 2.06)、HLA-DRB1\*0901(OR = 2.83)、TAP2379(OR = 2.09)、CYP2E1(OR = 1.60)10个因素进入模型; 以人群对照时, 年龄(OR = 1.10)、饮酒(OR = 6.27)、饮食不规律(OR = 118.05)、经常热烫饮食(OR = 3.02)、经常暴饮暴食(OR = 2.11)、少吃水果(OR = 6.80)、吃熏制肉(OR = 17.14)、胃病变史(OR = 5.31)7个因素进入模型; 从判别效果看4个模型的判别正确率分别为74.6%、70.1%、76.1%和96.9%。模型四的判别效果最好。

**结论:** 以环境、不良行为因素建立的风险预测模型适合基层应用, 风险预测模型可以通过该简单、有效的预测概率模型进行风险自我评估。

© 2014年版权归百世登出版集团有限公司所有。

**关键词:** 食管癌; 风险预测; Logistic判别模型

**核心提示:** 通过病例对照研究, 采用Logistic回归分析与食管癌相关的危险因素, Logistic判别方法建立食管癌判别模型及风险预测模型。

李婧, 秦江梅. 新疆哈萨克族食管癌风险的预测. 世界华人消化杂志 2014; 22(10): 1442-1445 URL: <http://www.wjgnet.com/1009-3079/22/1442.asp> DOI: <http://dx.doi.org/10.11569/wjcd.v22.i10.1442>

## 0 引言

预测模型已经大量应用于疾病一级预防, 通过预测模型确定患者未来发病风险。疾病风险预测模型不仅在发病率较高的疾病高血压、糖尿病等<sup>[1-3]</sup>应用广泛, 在肿瘤发病率相对较低, 乳腺癌、肺癌等<sup>[4-6]</sup>也有应用。食管癌是消化系统肿瘤中可获得较好防治效果的癌种, 人群一级预防措施可行<sup>[7,8]</sup>。前期已在新疆伊犁州新源县食管癌高发区地建立了哈萨克族食管癌监测点, 哈拉布拉乡和那拉提镇已经建立4万人的研究队列。本研究拟在前期研究的基础上, 利用Logistic分析方法建立高发区、高发民族哈萨克族食管癌风险预测模型。

## 1 材料和方法

**1.1 材料** 收集2008-2011年新疆维吾尔自治区人民医院、伊犁州友谊医院、伊犁州奎屯医院、新源县人民医院、阿勒泰地区人民医院和阿勒泰北屯医院六所医院, 经病理活检、X线或内窥镜确诊的哈萨克族新发食管癌166例。同期收集哈萨克族非肿瘤、非消化道疾病的住院病人或正常人群370例; 收集新疆新源县哈拉布拉乡居民4878例(该数据来自于科技支撑计划《新疆维吾尔族、哈萨克族高发防治适宜技术与示范》在哈拉布拉乡基于人群做的健康普查资料)。

**1.2 方法** 采用条件Logistic回归进行食管癌危险因素分析, 计算回归系数估计值、P值、OR值。利用Logistic判别法建立判别函数: Logistic回归中的两类判别, 第一类Y = 0, 第二类Y = 1, 根据指标可以建立logit(p)关于自变量X<sub>1</sub>, X<sub>2</sub>, ..., X<sub>m</sub>的Logistic回归方程:  $\text{logit}(p) = \beta_0 + \beta_1 X_1 + \dots + \beta_m X_m$  得概率估计公式:

$$p = \exp(\beta_0 + \beta_1 X_1 + \dots + \beta_m X_m) / [1 + \exp(\beta_0 + \beta_1 X_1 + \dots + \beta_m X_m)]$$

根据估计概率进行判别归类。计算预测概率P值, 建立风险预测图: 基于病例对照研究的非条件Logistic回归不能直接获得 $\beta_0$ 后的估计值,  $\beta_0$ 可在得到模型估计参数 $\beta_0$ 后, 对常数项进行校正, 获得 $\beta'_0 = \ln(p_0 / (1 - p_0)) - (\beta_1 \bar{X}_1 + \beta_2 \bar{X}_2 + \dots + \beta_k \bar{X}_k)$

其中为 $p_0$ 总体的患病率,  $\bar{X}_1, \bar{X}_2, \dots, \bar{X}_k$ 为X<sub>1</sub>, X<sub>2</sub>, ..., X<sub>k</sub>的总体均数, 由于总体均数不易获得, 可用样本均数代替, 后再用调整后的 $\beta'_0$ 作为Logistic回归方程的常数项就可以计算预测概率P值<sup>[9-11]</sup>。

**统计学处理** 数据核查后整理建库, 用SPSS17.0进行数据处理和分析, 统计学方法采用Logistic回归分析。P < 0.05为差异有统计学意义。

## 2 结果

**2.1 危险因素分析** Logistic回归分析显示: 以哈萨克族对照, 饮酒(OR = 2.77)、饮食不规律(OR = 3.42)、经常暴饮暴食(OR = 4.01)、少吃水果(OR = 2.65)、胃病变史(OR = 2.66)、进食速度快(OR = 1.94)、食管癌家族史(OR = 2.06)、HLA-DRB1\*0901(OR = 2.83)、TAP2379(OR = 2.09)、CYP2E1(OR = 1.60)10个因素进入模型; 以人群对照时, 年龄(OR = 1.10)、饮酒(OR = 6.27)、饮食不规律(OR = 118.05)、经常热烫饮食(OR = 3.02)、经常暴饮暴食(OR = 2.11)、少吃水果(OR

### ■ 研发前沿

在食管癌的预防研究中, 转变预防管理思想, 将疾病管理从单一危险因素管理转变为食管癌的风险预测管理, 从而提高了预防及干预力度。

## ■ 相关报道

孙凤等也利用Logistic回归构建了台湾省慢性肾病个体化风险预测模型,其模型是基于大规模人群长期随访的结果所建立,研究结论更能反映台湾当地居民的真实情况,应用价值很好,而且其简便易行,可直接估计体检者5年内新发疾病风险,其认为风险预测模型具有直观的可视化效果,不仅能够对中高度危险者提出强烈的警示信号,也能够用于那些有良好习惯的低度危险人群,鼓舞他们保持健康生活方式的信心。

表 1 危险因素的多因素分析

因素	对照1(哈萨克族对照)			对照2(全人群对照)		
	$\beta$	P值	OR	$\beta$	P值	OR
X <sub>1</sub> 年龄>50岁(对照≤50岁)	—	—	—	0.096	0.00	1.10
X <sub>2</sub> 饮酒(对照=否)	1.018	0.00	2.77 <sup>1</sup>	1.835	0.00	6.27 <sup>1</sup>
X <sub>3</sub> 饮食不规律(对照=规律)	1.230	0.00	3.42 <sup>1</sup>	4.771	0.00	118.1 <sup>1</sup>
X <sub>4</sub> 经常热烫饮食(对照=否)	—	—	—	1.105	0.00	3.02 <sup>1</sup>
X <sub>5</sub> 经常暴饮暴食(对照=否)	1.390	0.00	4.01 <sup>1</sup>	0.747	0.03	2.11 <sup>1</sup>
X <sub>6</sub> 少吃水果(对照=多吃)	0.975	0.02	2.65 <sup>1</sup>	1.918	0.00	6.80 <sup>1</sup>
X <sub>7</sub> 吃熏制肉(对照=否)	—	—	—	2.842	0.00	17.1 <sup>1</sup>
X <sub>8</sub> 胃病变史(对照=无)	0.975	0.00	2.66 <sup>1</sup>	1.670	0.00	5.31 <sup>1</sup>
X <sub>9</sub> 进食速度快(对照≥10分钟)	0.663	0.04	1.94	—	—	—
X <sub>10</sub> 食管癌家族史(对照=否)	0.721	0.02	2.06 <sup>1</sup>	—	—	—
X <sub>11</sub> HLA-DRB1*0901(对照=阴性)	1.039	0.00	2.83 <sup>1</sup>	—	—	—
X <sub>12</sub> TAP2379(对照=G/G型)	0.739	0.00	2.09 <sup>1</sup>	—	—	—
X <sub>13</sub> CYP2E1(对照=C1/C2或C2/C2型)	0.468	0.03	1.60	—	—	—

<sup>1</sup>OR>2.0.

表 2 模型判别效果的评价

	判别正确率(%)			判别函数
	对照组	病例组	总	
模型一	93.2	33.1	74.6	Logit(p) = -1.657+1.042X <sub>2</sub> +0.787X <sub>9</sub> +1.248X <sub>3</sub> +1.406X <sub>5</sub>
模型二	93.5	18.1	70.1	Logit(p) = -1.498+1.137X <sub>11</sub> +0.769X <sub>12</sub> +0.574X <sub>13</sub>
模型三	92.2	40.4	76.1	Logit(p) = -1.803+1.001X <sub>2</sub> +0.692X <sub>9</sub> +1.219X <sub>3</sub> +1.423X <sub>5</sub> +1.005X <sub>11</sub>
模型四	98.9	57.8	96.9	Logit(p) = -11.487+0.106X <sub>1</sub> +1.622X <sub>2</sub> +4.47X <sub>3</sub> +1.377X <sub>5</sub> +0.835X <sub>4</sub>

风险等级 □ <1% ■ 1%-9% ■ 10%-50% ■ >50%

图 1 模型四食管癌风险预测图.



= 6.80)、吃熏制肉(OR = 17.14)、胃病变史(OR = 5.31)7个因素进入模型,除外暴饮暴食,其他相同因素,以人群为对照的OR值更大(表1)。

**2.2 风险判别模型的建立** 建立包含不同危险因素的风险预测模型,模型一是基于可改变的不良行为危险因素、模型二是基于基因危险因素、模型三以基于基因和行为因素相结合、模型四利用的是人群对照,基于可改变的不良行为危险因素。从判别效果看上述4种模型的判别

正确率分别为74.6%、70.1%、76.1%和96.9%。模型四的正确率(96.9%)较其他3种模型高(表2)。

**2.3 建立风险预测图** 模型四具有较高的判别正确率,对照来自一般人群,符合人群实际,以此制定风险预测图。模型预测图中涉及年龄、饮酒、饮食规律性、热烫饮食、暴饮暴食等5个风险因素。如当个体年龄>50岁、饮酒、饮食不规律、热烫饮食、暴饮暴食未来患食管癌的风险概率>50%(图1)。

### 3 讨论

3.1 食管癌危险因素的影响 依据不同对照, 患病的风险存在很大不同如以人群为对照时, 相同危险因素的OR值更大这是由于以相同民族为对照, 其生活环境和饮食习惯等具有相似性, 部分高危因素在所有哈萨克族人群中普遍存在, 这会降低这些危险因素的风险. 从建立的以哈萨克族对照的3个风险模型的判别正确性看, 行为危险因素和基因危险因素同时存在模型判别正确率为76.1%, 较单独行为危险因素的74.6%和单独基因的70.1%, 增加幅度不明显, 该结果一方面说明行为危险因素与基因危险因素间存在共线性, 另一方面也与基因对食管癌发生的作用的微效性有关. 在哈族聚集地这样的偏远落后地区基因检测技术不完善, 环境、不良行为因素数据更易获得, 风险预测模型四更加适用.

3.2 风险预测模型的应用 WHO利用年龄、性别、吸烟等简单易测的指标建立的心血管疾病风险预测图, 已作为一种干预措施利用于初级卫生保健领域<sup>[12]</sup>. 风险预测图简单, 可对个体进行风险自我评估. 基于当地哈萨克族人群食管癌发病率为20/10万, 本研究认为个体食管癌预测概率高于1%即可作为界点, 高于1%的人群可视为危险人群. 孙凤等<sup>[13-15]</sup>也利用Logistic回归构建了台湾省慢性肾病和2型糖尿病个体化风险预测模型, 直接估计体检者5年内新发疾病风险. 其认为风险预测模型具有直观的可视化效果, 不仅能够对中高度危险者提出强烈的警示信号, 也能够用于那些有良好习惯的低度危险人群, 鼓舞他们保持健康生活方式的信心. 本研究建议基层采用风险预测模型告知食管癌高危危险人群, 让高危人群意识到患癌风险和自觉改善不良行为, 可能会降低食管癌的发病风险达到预防食管癌的目的.

3.3 研究的局限性 本研究具有一定的局限性, 主要包括两个方面, 一是风险模型的建立不是基于前瞻性的队列研究, 二是研究还未进行人群验证研究, 旨在为新疆哈萨克族食管癌高危人群监测体系的建立提供循证依据和基础信息.

### 4 参考文献

- 1 Cardiovascular risk. Geneva, World Health Organization, 2007
- 2 Guilbert JJ. The world health report 2002 - reducing risks, promoting healthy life. *Educ Health* (Abingdon) 2003; 16: 230 [PMID: 14741909]
- 3 周玲, 胡旭, 徐天亮, 刘继永, 徐海蓉. 2型糖尿病高危人群的早期预测及其评价. *中国公共卫生* 2004; 20: 293-294
- 4 Tucker JD, Preston RJ. Chromosome aberrations, micronuclei, aneuploidy, sister chromatid exchanges, and cancer risk assessment. *Mutat Res* 1996; 365: 147-159 [PMID: 8898995]
- 5 林欢, 钟文昭, 杨学宁, 严红虹, 吴一龙. 肺癌家系肿瘤风险度病例对照研究与预测模型. *中国肺癌杂志* 2011; 14: 581-587
- 6 戴俊程. 基于全基相关研究的中国女性乳腺癌风险预测模型研究. *南京医科大学* 2011: 34-38
- 7 魏文强, 乔友林, 邵壮, 王贵齐, 李新庆, 范金虎, 王国清, 董志伟. 食管癌高危人群的预防与控制研究进展. *实用肿瘤杂志* 2001; 16: 371-373
- 8 杨秋兰, 韩璐. 三成以上癌症可以预防. *健康报* 2009; (2): 2
- 9 孙振球, 徐勇勇. *医学统计学*. 北京: 人民卫生出版社, 2010: 278-336
- 10 陈峰. *医用多元统计*. 北京: 高等教育出版社, 2002: 85-91
- 11 张家放. *医用多元统计方法*. 北京: 高等教育出版社, 2004: 128-157
- 12 Package of essential noncommunicable disease interventions for primary health care in low-resource settings. Geneva, World Health Organization, 2010
- 13 孙凤, 陶庆梅, 陶秋山, 杨兴华, 曹纯铿, 詹思延. 中国台湾地区35-74岁健检人群2型糖尿病患病率以及5年发病风险预测模型. *中华疾病控制杂志* 2013; 17: 369-373
- 14 孙凤, 郁凯, 陶庆梅, 陶秋山, 杨兴华, 曹纯铿, 詹思延. 中国台湾省35-74岁健检人群慢性肾病5年发病风险预测模型. *中国预防医学杂志* 2013; 14: 47-52
- 15 孙凤, 陶庆梅, 陶秋山, 杨兴华, 曹纯铿, 詹思延. 中国台湾省35-74岁健检人群肥胖5年发病风险预测模型. *中华流行病学杂志* 2012; 33: 47-52

### ■同行评议

本文统计方法运用恰当合理, 利用多因素条件Logistic回归分析有一定的学术价值.

编辑 田滢 电编 鲁亚静

