

- 基础理论与方法 -

启东肝癌发病趋势的年龄-时期-队列模型分析

沈其君 张雪峰 陈建国 李文广 姚红玉

【摘要】 目的 反映启东地区肝癌发病的时间变化趋势、出生队列效应和年龄的作用方式。方法 应用年龄-时期-队列模型对启东地区肝癌发病监测资料进行统计分析。结果 时间趋势变化主要呈队列趋势变化,表现出男性1913~1917年和1958~1962年以后的出生队列的发病危险性有减缓的趋势。结论 可以推断1958~1962年以后的出生队列的发病危险性有下降的趋势,是综合防治措施的实施结果。

【关键词】 肿瘤,肝细胞;模型,年龄-时期-队列;时间变化趋势

An age-period-cohort modelling study on primary liver cancer incidence rate in Qidong SHEN Qi-jun*, ZHANG Xue-feng, CHEN Jian-guo, LI Wen-guang, YAO Hong-yu. *School of Medicine Ningbo University, Ningbo 315211, China

【Abstract】 Objective To analyze the incidence trends of primary liver cancer (PLC) in Qidong. **Methods** Data of PLC incidence from 1975 to 1999 in Qidong were analyzed to delineate temporal trends and birth cohort patterns, using age-period-cohort models. **Results** Significant moderation or decreasing trends were began to notice in incidence rates on cohorts born in 1913-1917 and 1958-1962. **Conclusion** Results showed that the incidence risk of the birth cohorts after 1958-1962 started to decline. The changes were possibly associated with the implementation of some practical measures on prevention.

【Key words】 Carcinoma hepatocellular; Model age-period-cohort; Time trends

对肝癌等慢性病发病率的时间变化趋势,仅计算不同时期的年龄别粗发病率和调整发病率具有一定的局限性。年龄-时期-队列模型改进了传统的分析方法,以 Poisson 分布为基础,可同时调整年龄、时期和队列等因素的条件下,反映一定人群的疾病发病或死亡的危险性及变化趋势。我们通过应用年龄-时期-队列模型对启东地区1975~1999年肝癌发病监测资料进行统计分析,观察经过多方面的干预措施以后的肝癌时间变化趋势,以及出生队列和年龄的作用方式,以便为今后肝癌病因学研究和防治评价提供科学依据。

基本原理

年龄-时期-队列模型实际上是 Poisson 回归模型,其基本形式为 $\ln[E(r_{ijk})] = \ln(\theta_{ijk}/N_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ijk}$, 其中 μ 表示年龄、时期、队列作用的疾病危险性参照水平; α_i 表示第 i 年龄组效应

($i = 1, 2, 3, \dots, m$); β_j 为第 j 时期组作用($j = 1, 2, 3, \dots, m$); γ_k 是指与第 i 年龄组和第 j 个时期参数有关的第 k 组出生队列的效应^[1], θ_{ijk} 、 N_{ijk} 分别表示第 i 年龄组、第 j 个时期和第 k 个队列组的发病(死亡)数和人口数(人年数)。对实际资料分析时应分别拟合年龄模型、年龄-趋势模型、年龄-时期和年龄-队列两因素模型以及年龄-时期-队列三因素模型,即因素逐步递增的梯度模型^[2]。当模型中年龄、时期、队列作用同时存在时,三因素间存在完全线性依赖关系,需施加限制才能获得惟一的参数估计值。我们对国外学者提出的惩罚函数法、个体资料法、估计函数法和非参数检验法等参数估计方法进行比较研究,并结合国外学者的研究成果^[1-3],选用估计函数法,通过偏差(线性作用外的剩余部分)或曲率(三因素斜率的二阶改变)来反映三因素作用趋势变化。

应用 Tarone 法检验队列作用的变化趋势^[3],其统计量为 C_{yz}^m , 其中 y, z 为前面队列区段和后面队列区段的首个队列, m 为每个区段包含的队列数。 C_{yz}^m 为零时不存在非线性作用; C_{yz}^m 为负值时曲线的斜率下降,意味着年青队列的发病危险度上升减缓或下降加快;当 C_{yz}^m 为正值时,曲线的斜率增加,表

作者单位 315211 宁波大学医学院(沈其君);江苏省疾病预防控制中心(张雪峰);江苏省启东肝癌防治研究所(陈建国、李文广、姚红玉)

达的意义与负值时相反。

模型拟合度评价统计量为 Pearson χ^2 值和偏差统计量(G^2)^[4], Pearson χ^2 和偏差统计量(G^2)服从自由度为资料格子数减参数个数的 χ^2 分布。当模型拟合不理想时,通过回归诊断和残差分析,若为过度离散可通过引入由 Breslow^[4] 提出的 Extra-Poisson 变量,对极大似然拟合进行改进,拟建 Extra-Poisson 回归模型,此时模型参数检验应用 F 检验。

实例分析

1. 资料来源及整理:收集启东地区 1975~1999 年肝癌发病资料,每例包括发病时间、年龄、性别和居住地等因素。肝癌诊断分类按国际疾病分类 ICD-9,所有肝癌病例都按照全国肝癌协作组制定的诊断标准进行诊断,并于 1982、1994 和 2000 年进行了回顾复核与随访,保证了肝癌病例的诊断质量。在数据库建库过程中,采用双机输入,以保证数据的准确性。人口资料以 1982 年和 1990 年两次全国性人口普查以及 1976 年与 1999 年当地政府组织大规模人口抽样调查的年龄组构成为基础,对相邻两次普(抽)查年间各年度年龄别人口数进行内插。年龄组 5 岁为一组,因小年龄组与大年龄组的例数较少,去除 0~19 岁的发病例数,80 岁以上为一年龄组,即 20~25~...80~,共分为 13 个年龄组;发病时期以 5 年为一组,即 1975~1979 年...1995~1999 年,共分成 5 个时期,将资料按年龄和时期的分组列成二维列联表。出生队列按时期减去年龄来推算,两队列间的重叠通过对出生队列取中值而加以忽略^[1]。分析中分别探讨了不同年龄和时期分组的模型拟合。

2. 模型拟合结果:表 1 为启东市男性与女性肝癌发病资料拟合年龄-时期-队列梯度模型及拟合优度检验。表 1 中所有模型的优度检验 P 值均小于 0.05,模型拟合不理想。通过残差分析没有显示出明显的异常点和显著性的变化趋势。模型拟合度较差的可能是由于模型的过度离散造成的。为此,对资料拟合 Extra-Poisson 回归模型,结果见表 2。偏差统计量接近自由度,模型拟合良好($P > 0.05$)。应用 F 检验对模型中年龄、时期、队列进行检验, P 值均小于 0.01,均有统计学意义。由于三因素间存在着线性依赖性,因此通过估计函数法的曲率来具体地反映三因素作用的变化趋势(表 3),以及应用

Tarone 法检验队列作用的作用方式。

表1 启东市 1975~1999 年男女性肝癌发病资料 年龄-时期-队列梯度模型及拟合优度检验

模型	男 性			女 性		
	偏差 (G^2)	自由 度	P 值	偏差 (G^2)	自由 度	P 值
年龄	340.99	52	0.000	174.50	52	0.0000
年龄-趋势	294.53	51	0.000	174.19	51	0.0000
年龄-时期	275.01	48	0.000	154.52	48	0.0000
年龄-队列	109.45	36	0.000	86.91	36	0.0000
年龄-时期-队列	66.88	33	0.001	57.33	33	0.0054

表2 启东市男女性肝癌发病资料 Extra-Poisson 变量 年龄-时期-队列模型拟合及检验

模型	男 性				女 性			
	偏差 (G^2)	自由 度	F 值	P 值	偏差 (G^2)	自由 度	F 值	P 值
时期-队列	590.050	44	50.610	0.00	183.040	44	13.580	0.00
年龄-时期	156.530	48	8.212	0.00	101.740	48	4.540	0.00
年龄-队列	56.443	36	7.699	0.00	52.171	36	6.178	0.00
年龄-时期-队列	33.346	33			33.637	33		

从表 3 可以看出,男性年龄作用的斜率二阶改变显著的年龄组为 25~岁和 40~岁年龄组。25~岁年龄组的曲率为 -0.45,其意义是 30~岁年龄组对于 25~岁年龄组的发病相对危险度是 25~岁年龄组对于 20~岁年龄组相对危险度的 64% [$\exp(-0.45)$],局部曲线的斜率下降。其他年龄组的斜率二阶改变不明显,呈线性形式,说明 25~29 岁以后的年龄作用上升趋势减弱,25~29 岁年龄以前是发病上升最快的时期,40~44 岁以后进一步减缓。表 3 中,女性的年龄作用的二阶改变主要发生在 35~岁年龄组和 75~岁年龄组,可以认为 35~岁和 75~岁年龄组以后的年龄作用上升趋势减弱,35~39 岁以前是发病上升较快的时期。

男性与女性的各个时期作用中曲率作用均不明显,说明时期作用呈完全线性趋势,不存在非线性波动。男性队列作用的斜率二阶改变主要发生在 1913~1917 年和 1958~1962 年队列组,其曲率分别为 -0.339、-0.266,其他队列组没有发生显著性改变。以 1913~1917 年、1958~1962 年的出生队列为转折点,应用 Tarone 法分别检验这两个出生队列前面与后面的出生队列的总体趋势的变化: $C^4_{1900,1920} = (3\gamma_{1935} + \gamma_{1930} - \gamma_{1925} - 3\gamma_{1920}) - (3\gamma_{1915} + \gamma_{1910} - \gamma_{1905} - 3\gamma_{1900}) = -2.586, s_{\bar{x}} = 0.637; C^4_{1940,1960}$ 值为 -3.906, $s_{\bar{x}} = 1.056$,由此可以认为男性的 1913~1917 年和 1958~1962 年以后的出生队列曲线斜率下降,发病危险性减缓。女性的队列作用的斜率二

阶改变主要发生在 1958~1962 年队列组,其曲率为 -0.576,队列趋势检验 $C^2_{1940,1960}$ 值为 -4.623, $s_{\bar{x}} = 2.28$,女性的 1958~1962 年以后的出生队列的曲线斜率下降,发病危险度有减缓的趋势。

表3 启东市 1975~1999 年男女性肝癌发病资料
年龄-时期-队列曲率估计

研究因素	男 性		女 性	
	曲率	$s_{\bar{x}}$	曲率	$s_{\bar{x}}$
年龄组(岁)				
20~	-	-	-	-
25~	-0.445	0.164	-0.238	0.291
30~	-0.205	0.127	0.090	0.222
35~	-0.156	0.115	-0.411	0.186
40~	-0.188	0.081	-0.036	0.173
45~	-0.165	0.112	-0.220	0.173
50~	0.043	0.115	0.189	0.178
55~	-0.009	0.118	-0.077	0.178
60~	-0.070	0.124	-0.075	0.183
65~	0.082	0.135	0.097	0.195
70~	-0.007	0.154	0.003	0.211
75~	-0.253	0.191	-0.550	0.241
80~	-	-	-	-
时期(年)				
1975~1979	-	-	-	-
1980~1984	0.058	0.083	0.045	0.128
1985~1989	0.101	0.081	0.183	0.126
1990~1994	0.090	0.080	0.089	0.121
1995~1999	-	-	-	-
队列组(年)				
1893~1897	-	-	-	-
1898~1902	0.3231	0.474	0.679	0.521
1903~1907	-0.287	0.291	-0.223	0.362
1908~1912	0.210	0.204	0.197	0.271
1913~1917	-0.339	0.154	-0.153	0.215
1918~1922	0.0202	0.135	-0.074	0.193
1923~1927	0.004	0.125	0.007	0.184
1928~1932	-0.012	0.119	-0.110	0.177
1933~1937	-0.028	0.115	0.115	0.177
1938~1942	-0.074	0.114	-0.205	0.178
1943~1947	0.157	0.115	0.199	0.184
1948~1952	-0.056	0.116	-0.064	0.187
1953~1957	-0.191	0.126	-0.136	0.204
1958~1962	-0.266	0.127	-0.576	0.273
1963~1967	0.090	0.234	0.467	0.455
1968~1972	-0.016	0.481	-0.309	0.751
1973~1977	-	-	-	-

注:表内数值是通过调整 Extra-Poisson 变量得出的结果

讨 论

年龄-时期-队列模型可同时调整年龄、时期和队列等多因素的条件下,估计一定人群的疾病发病或死亡的危险性,描述疾病在时期或队列和年龄上的变化趋势,可以克服单因素分析中时期、队列和年龄的相互影响。但当年龄、时期、队列三因素同时存在时,由于因素间的线性依赖性,只有施加限制才能得出确切的参数估计值。各种任意限制和其他一些参数估计方法对估计结果都造成一定的影响,只有

各因素的非线性部分[曲率(斜率的二阶改变)和偏差]是可识别的,因此可通过它们来反映年龄、时期、队列等因素作用趋势变化。

从年龄作用的曲率估计结果可以看出,男性25~岁和40~岁的年龄组以及女性35~岁和75~岁年龄组以后的年龄作用上升趋势进一步减缓,男性25~岁和女性35~岁以前年龄组是率值上升最快时期。

通过男性与女性的时期和队列的曲率结果,表明时间趋势变化主要呈队列趋势变化,表现出男性1913~1917年和1958~1962年以后的出生队列的发病危险性有减缓的趋势。对于肝癌高发可能的危险因素,20世纪70年代启东地区推广用荧光检测仪去除霉变玉米,80年代以后改变膳食构成;从1972年起,逐步推广饮用地下水等措施^[5]。根据肝癌的致癌作用主要发生在幼年时期和青少年时期的特点^[6],通过以上的干预措施和卫生条件、生活水平的改善,使得在这个时期的幼年和青少年队列以及以后的出生队列所暴露的危险性发生改变。可以推断1958~1962年以后的出生队列的发病危险性有下降的趋势,是综合防治措施的实施结果。

年龄-时期-队列模型也存在着一些缺陷。首先两端的队列有不稳定性:在两端队列中观察年龄组较少且发病频数较小,通过极大似然法获得的参数估计稳定性较差,尤其对于两端队列发病例数较少的资料,就很难评价两端队列的作用。其次致病因素的不确定性:时期、队列作用本身并不是致病因素,对于时期和队列作用的变化,不能具体到是由哪个因素作用引起的,因此拟合年龄-时期-队列模型对参数进行估计,只是为病因学探索性分析、防治措施评价提供线索,具体到每一个致病因素和防治措施的作用还需要进一步进行评价。

参 考 文 献

- Robertson C, Boyle P. Age-period-cohort analysis of chronic disease rates: modeling approach. *Stat Med*, 1998, 17: 1302-1323.
- Clayton D, Schiffrers E. Models for temporal variation in cancer rates: Age-period-cohort model. *Stat Med*, 1987, 6: 469-481.
- Tarone RE, Chu KC. Evaluation of birth cohort patterns in population disease rates. *Am J Epidemiol*, 1996, 143: 85-91.
- Breslow NE. Extra Poisson variation in log linear models. *Applied Statistics*, 1984, 39: 38-44.
- 陈建国, 朱荣源. 肝癌的流行病学与预防战略-高发区的研究及效果初评. *启东肝癌研究*, 1995, 11: 45-47.
- 俞顺章. 原发性肝癌流行病学研究进展. *实用肿瘤杂志*, 1998, 13: 130-133.

(收稿日期 2004-01-07)
(本文编辑 张林东)