

胃癌高发区流行因素分析

徐州医学院 李学信 周卫平 韩友圻 王少林 陈平

胃癌在江苏省癌死因中居第二位,粗率(CMR)30.33/10万,标化率(SRD)25.85/10万。为探讨胃癌与软食种类及习惯、水源、经济生活和社会心理等诸因素的关系,服务于防治实践的需要,我室于1983年5~9月深入我省胃癌高发的扬中、靖江两县进行现场流行病学调查,所得资料用PC-1500计算机处理,现将结果报告如下。

材料及方法

全部资料均来自以上两县,选用“病例-对照”研究方法,各400例。病例来自两县1980~82年登记报告的确诊胃癌者;对照来自与病例住址相近、年龄相差在5岁以内,性别、职业、文化水平相同的健康人。均由我室专业人员按同一表格调查。

单因素分析选用“Mantel-Haenszel”方法。多因素分析,共50个因素变量,为解决PC-1500机内存不足,先用信息量分析(熵函数)筛选出重要的变量,而后用分段小程序连续输入数据的方法进行逐步回归分析和模糊聚类分析。自变量(即流行因素X)分别赋值直接运算,定性指标取0,1;半定量指标一般取0,1,2,3从小到大排列,均选用与因变量(Y)呈正相关联系的取值方法。因变量胃癌者取值为1,对照取0,以上计算在PC-1500机上进行。

结果

一、资料均衡性检验:病例与对照各400例,经均衡性检验分析,两组资料无显著性差异(P>0.05)。

二、50种流行因素单因素分析:发现有16

个因素与胃癌发病呈明显正相关(表1,2)。

三、用熵函数筛选主要流行因素:在逐步回归分析前,我们用熵函数分析方法从50个因素中筛选主要流行因素^[1](程序由下列公式编出 $H_P = - \sum_{k=1}^n P_k \text{LOG}_2 P_k$),结果熵函数值取0.88382~0.41057范围,共筛选出21个因素,并参与逐步回归分析。

表1 胃癌流行因素单因素分析结果 (M-H方法)

因素	RR	RR95%CL	χ^2	P
饮用污水	27.19	40.07~18.41	276.70	<0.01
口粮<400斤*	5.44	7.53~3.93	104.56	<0.01
烫食	3.82	5.82~2.79	70.69	<0.01
家族胃癌史	3.71	5.72~2.41	35.49	<0.01
常生闷气	3.41	4.88~2.39	45.04	<0.01
精神创伤	2.73	3.68~2.03	43.57	<0.01
支出>收入*	2.64	3.52~1.97	43.14	<0.01
支出>收入*	2.57	4.08~1.62	16.18	<0.01
性格固执	2.54	3.38~1.01	40.69	<0.01
饭前饥饿	2.43	3.26~1.61	35.25	<0.01
紧张过劳	2.12	2.82~1.59	26.25	<0.01
饮用死水	1.99	2.65~1.49	21.73	<0.01
暴饮暴食	1.85	2.49~1.38	16.49	<0.01
食酱制品	1.83	2.43~1.38	17.40	<0.01
常吃不饱	1.55	2.17~1.11	6.53	<0.05
进食快	1.37	1.84~1.02	4.39	<0.05

注: * 发病5年前 • 发病15年前

四、21个流行因素逐步回归分析结果:本资料病例和对照各400例,每例为21个自变量和1个因变量,用PC-1500机采用分段小程序接转运算方法进行逐步回归分析。当界限值等于17.2时,有7个自变量入选,依作用大小分别列入表3。

复相关系数(R)为0.65627,常数项(Bo)0.18061, P<0.05~0.01,见表3。

表 2 胃癌保护性因素单因素分析结果 (M-H方法)

因 素	RR	RR95%CL	χ^2	P
食用辣椒	0.22	0.15~0.31	65.66	<0.01
人均收入 >120元*	0.24	0.17~0.33	68.41	<0.01
食用莴苣	0.28	0.10~0.40	45.42	<0.01
大 蒜	0.37	0.27~0.51	36.06	<0.01
腌 菜	0.49	0.30~0.90	8.13	<0.01
罗 卜	0.52	0.39~0.69	19.98	<0.01
黄 瓜	0.55	0.39~0.76	12.31	<0.01
人均收入 >120元*	0.63	0.47~0.84	10.19	<0.01
食用蕃茄	0.65	0.47~0.89	7.14	<0.01
水 果	0.71	0.52~0.97	4.64	<0.05

注: **同表1

表 3 逐步回归分析结果

入选变量	V 值	回归系数
饮后沟污染水	0.09834	0.36135
人均口粮水平低*	0.06922	0.28206
不吃辣椒	0.0429	0.25207
生闷气	0.03464	0.22496
家族胃癌史	0.01929	0.1587
经济支出>收入*	0.0156	0.09586
人均年现金收入少*	0.0124	0.07759

注: * 同表1

五、聚类分析结果: 为了分析以上21个流行因素的各自性质及其相互关系, 我们选用模糊聚类分析^[2](表4)。

讨 论

一、通过对江苏省胃癌高发区的流行因素研究发现16个因素与胃癌发病呈正相关关系。尤其与饮用房屋后沟水联系最为密切, 这是我省较为突出的问题, 其机理有待深入研究。另外还明显看出与食用新鲜蔬菜、瓜果呈负相关关系, 是可能的胃癌保护性因素。鉴于胃癌多病因学说, 为控制偏倚和混杂的影响, 分析中采用多元分析技术处理, 逐步回归分析结果(当界限值17.2时), 有7个变量入选, $R = 0.65627$, 而其他大部分变量的作用可为以上入选变量所代替, 若回归方程包括21个变量, 相应其R值会有所提高, 但提高很小(如取界

表 4 模糊等价矩阵R^{**}结果

λ值	分 类 情 况	合计类数
0.91	X ₅ , X ₁₃	20
0.90	X ₇ , X ₉ , X ₁₃	19
0.89	X ₅ , X ₆ , X ₇ , X ₉ , X ₁₃	18
0.88	X ₄ , X ₇ , X ₉ , X ₁₃ , X ₁₅ , X ₆	17
0.87	X ₄ ~X ₇ , X ₉ , X ₁₃ , X ₁₅	15
0.85	X ₄ ~X ₉ , X ₁₃ , X ₁₅	14
0.79	X ₄ ~X ₉ , X ₁₃ , X ₁₅ , X ₂₀	13
0.78	X ₂ , X ₄ ~X ₉ , X ₁₃ , X ₁₅ , X ₂₀	12
0.75	X ₁ , X ₂ , X ₄ ~X ₉ , X ₁₃ , X ₁₇ , X ₁₉ , X ₂₀	10
0.73	X ₁ , X ₂ , X ₄ ~X ₉ , X ₁₃ , X ₁₅ , X ₁₇ , X ₁₉ , X ₂₀	9
0.63	X ₁ , X ₂ , X ₄ ~X ₉ , X ₁₂ ~X ₂₅ , X ₁₇ , X ₁₉ , X ₂₀	7
0.60	X ₁ ~X ₉ , X ₁₂ ~X ₁₇ , X ₁₉ , X ₂₀	5
0.55	X ₁ ~X ₁₀ , X ₁₂ ~X ₁₇ , X ₁₉ , X ₂₀	4
0.53	X ₁ ~X ₁₀ , X ₁₂ ~X ₁₇ , X ₁₉ , X ₂₀ , X ₂₁	3
0.48	X ₁ ~X ₁₇ , X ₁₉ ~X ₂₁	2
0.35	X ₁ ~X ₂₁	1

注: * 没有写入的变量各自成为一类

限值为1.5, 则入选变量为18个, $R = 0.70003$)。当然, 由于 $R = 0.65627$, 故对其他未知因素应进一步深入探讨。以上基本和国内外现阶段的研究相一致^[3,4]。

二、在多因素分析中, 我们首先引进熵函数来筛选主要流行因素, 熵函数值愈小则该因素作用越大, 从50个因素中筛选出21个主要因素, 和用“M-H”方法的分析结果基本一致。该信息量分析方法较之“M-H”方法更为简单、直观, 在多因素分析中, 是一个有效的初筛方法。另外在分析中, 我们还对21个变量进行模糊聚类分析, 以看各变量的性质和相互作用关系, 从表4可看出逐步回归所入选的7个主要流行因素, 基本上都是自成一类。但是当入值为0.89时, 食用辣椒和大蒜两个因素同属一类, 作用相同; 当入值为0.79时, 发病5年前口粮低(X₂₀), 发病15年前经济收支情况(X₄)、食用辣椒(X₅)、大蒜(X₆)、莴苣(X₇)、烫食(X₈)、稀饭(X₉)、腌菜(X₁₃)的情况, 发病15年前现金收入情况(X₁₉)以及发病5年前口粮多少(X₂₀)等8个因素同属一类, 它们在性质上及作用上是一致的。所以模糊聚类分析有助于我们对多因素的分析, 弄清它们综合的

与各自在胃癌发病中的作用。

三、以上数据分析均在PC-1500计算机上(8k模块)完成,为解决该机内存较小的缺点,我们将原始数据编制小程序,用CHAIN语句结合录音机,使分段程序接转(该方法可处理任意组数据),从而完成大数据多变量的分析,若将该机换用16k模块或使用该机扩充卡,则处理的变量更多,在计算时还要方便。以上全部数据处理该机运转1天即可完成。

摘 要

本文介绍用“病例-对照”研究,对江苏省胃癌高发区流行因素进行分析。因素包括饮食种类和习惯、水源、经济生活和社会心理共50种,资料的单因素分析发现有16种因素与胃癌呈正相关联系($P < 0.05 \sim 0.01$);有10种因素与胃癌呈负相关联系($P < 0.05 \sim 0.01$)。多因素分析,首先用信息量分析而后用逐步回归和模糊聚类分析,发现主要入选因素依次为饮用河沟污染水、发病5年前人均口粮低、不食辣椒、生闷气、家族、胃癌史,发病5年前经济亏空及现金收入低等7个因素,复相关系数0.65627,常数项0.18601, $F_{进} = F_{出} = 17.2$ 。全部资料用PC-1500机处理,数据用分段小程序输入解决该机内存不足。

Case-control Study of Gastric Cancer in high-incidence Areas Li Xuexin et al., Xuzhou Medical College

Case-control study was used to analyze the fac-

tors of gastric cancer in high-incidence areas, Yangzhong and Jingjiang counties, Jiangsu Province. A total of 50 factors were examined, including the factors of diet, food habit, water supply, socioeconomic status, psychological factors etc. Through the study of individual factors it was revealed that positive correlation existed in gastric cancer with 16 factors ($P < 0.05 \sim 0.01$) and negative correlation with 10 factors ($P < 0.05 \sim 0.01$). For multiple factor analysis, the data were initially processed with entropy function and then treated by stepwise regression fuzzy cluster. As a result, 7 factors were found to be associated with gastric cancer; they in order were: polluted drinking water, low food provision in the period of 5 years previous to diagnosis of cancer, non-ingestion of capsicum, low spirits, family history of gastric cancer and low income (in cash) or being in debt period of 5 years before onset of the illness (coefficient of multiple correlation = 0.65627, constant = 0.18601, $F_{i} = F_{o} = 17.2$). All of the data were processed with PC-1500 computer, and segmented program was adopted for input data so as to suit the computer's limited capacity of internal storage.

参 考 文 献

1. 蒋大宗,等.数值诊断的统计方法.西安1981:111~256.
2. 严洪范,等.模糊数学 1983:(3)53~56.
3. Gottlieb A, et al. Intern J Epidemiol 1981:10(2):117.
4. 易应南,等.中华流行病学杂志 1981:2(2)123.
(参加该项研究工作的还有李瑛、楚建军、封锦芳、于鸣娟、吴秀娟、杨平镇以及杨中、靖江两县卫生防疫站有关同志)

济宁市区流行性出血热的流行病学调查

济宁市市中区卫生局

周 华

从济宁市区出血热发病流行情况分析,发病具有3~5月、11~12月两个高峰,呈现明显的季节性;褐家鼠为主要传染源,室内为主要感染场所。在市郊交错居住的条件下,工人、农民发病均较高,平房及楼下居住者显著高于楼上居住者;且有一户一例、一

村一例的高度散发特征。发病年龄以青壮年为主,且男多于女。我们认为在疫区不论男女老幼及不同职业均可感染发病,在预防上应根据当地情况,加强领导,开展卫生防病知识宣传,灭鼠灭螨,做好个人防护,加强疫情报告和管理,降低发病率和病死率。