

宫颈癌危险因素的条件Logistic 回归分析和通径分析

山东医科大学流行病学教研室

张桂宁 徐爱强

提要 本文对125例宫颈鳞癌病人,按1:1配对进行了病例对照研究。单因素分析,从39个研究因素中初筛出14个可疑危险因素。条件Logistic回归分析,筛选出5个危险因素:宫颈糜烂、家庭经济状况、经期卫生评分、性因素和初婚年龄。通径分析结果表明,宫颈糜烂、不常洗外阴、平时就诊医院和初婚年龄4因素对宫颈癌的发生有直接作用;初婚年龄、月经垫不洁、配偶包皮过长和家庭经济状况4因素通过各自的中间环节起间接作用。

关键词 宫颈癌 条件Logistic 回归分析 通径分析

宫颈癌的发生是多种因素综合作用的结果,但其病因和致病机理迄今仍不十分明了。为了探索宫颈癌的主要危险因素、各因素致病作用的大小及其相互关系,我们应用配对病例对照研究方法,对宫颈癌的各种可能危险因素进行了单因素分析、条件Logistic回归分析及通径分析,结果如下:

资料与方法

一、病例和对照的来源及选择:病例和对照同选自山东医科大学附属医院、山东省肿瘤医院及山东省千佛山医院的妇科门诊和病房病人。选择经病理组织学检查确诊的宫颈鳞癌病人为病例。按1:1配对,从每个病例的同一医院的门诊和病房中,选择同民族、同职业、年龄相差不超过3岁,阴道脱落细胞学检查或/和宫颈组织病理学检查未见有病理变化的子宫肌瘤、卵巢肿瘤病人及健康检查的妇女为对照。

二、调查方法和内容:所有研究对象均用设计的同一调查表进行调查,由一人完成。于1987年10月至1988年11月共调查125对。调查内容包括月经史、经期卫生情况、婚姻史、生育史、避孕方法、婚外性生活史、宫颈疾病

(妇科检查由妇科医师协助完成)、吸烟史、经济文化卫生水平、平时就诊医院的级别、家族肿瘤史、配偶的文化程度、包皮过长否及结婚次数等,共39个因素。

三、资料分析方法:

1.单因素分析:按1:1配对资料分析,计算 χ^2 和OR。三个等级1:1配对计数资料的 χ^2 检验,按下列公式计算:

$$\chi^2 = \frac{\bar{n}_{23} \cdot d_1^2 + \bar{n}_{13} \cdot d_2^2 + \bar{n}_{12} \cdot d_3^2}{2(\bar{n}_{12} \cdot \bar{n}_{13} + \bar{n}_{12} \cdot \bar{n}_{23} + \bar{n}_{13} \cdot \bar{n}_{23})}$$

$$df = k - 1 = 2$$

式中: $\bar{n}_{ij} = (n_{ij} + n_{ji})/2$, $d_k = n_{ko} - n_{ok}$

2.条件Logistic回归分析:用条件Logistic回归模型对资料进行配合,公式为:

$$\ln L^* = - \sum_{i=1}^{125} \ln [1 + \exp (\sum_{k=1}^p \beta_k \cdot d_{ik})]$$

用最大似然函数法和Newton-Raphson迭代法求式中参数 β_k 及其方差的估计值。各危险因素的显著性检验用统计量 $G = 2(\ln L_{t+1}^* - \ln L_t^*)$,若 $G > \chi^2_{0.10}(df)$ 则将该危险因素纳入回归模型。

3.通径分析:

(1)设计通径分析模式:确定通径的原

则为：①两现象发生时间的先后；②这种联系符合逻辑和专业知识；③对个别由①、②不能确定其通径的因素，提出通径假说。

(2) 建立各种数学模型(多元方程)。

(3) 解多元方程求通径系数(P_{yi})：用通径系数 P_{yi} 表示因素对宫颈癌发生的直接影响；用某因素对宫颈癌发生影响通径中各通径系数的乘积表示该因素对宫颈癌发生的间接影响。

上述运算均在Apple-Ⅱ计算机上完成。

结 果

一、单因素分析：定在 $\alpha=0.05$ 水平上有显著性意义或/和 $OR>2$ 者为可疑危险因素。

经分析，从39个因素中初筛出14个可疑因素：月经垫不洁、经期不洗外阴、经期卫生评分①、初婚年龄、结婚次数 ≥ 2 、宫颈糜烂、家庭经济状况、平时就诊医院、配偶结婚次数 ≥ 2 、配偶②包皮过长、配偶性紊乱史③、本人性紊乱史、综合性紊乱史④及性因素⑤。

二、条件Logistic回归分析：以单因素分析筛选出的14个因素进行条件Logistic回归模型配合，定 $\alpha=0.10$ ，运算结果最终进入回归方程的有5个因素，按比值比估计值的大小依次为：宫颈糜烂(X_{23})、家庭经济状况(X_{29})、经期卫生评分(X_{10})、性因素(X_{39})及初婚年龄(X_{12}) (表1)。

各危险因素在不同暴露水平 X_K^* 和 X_K 下的

表1 选入模型的5因素及其有关参数估计值

危险因素 (X_K)	回归系数 $\hat{\beta}_K$	方差 $Var(\hat{\beta}_K)$	标准误 $SE(\hat{\beta}_K)$	标准化回归系数 $STD(\hat{\beta}_K)$	比值比 (OR)
X_{23} ：宫颈糜烂	1.1017	0.1157	0.3401	3.2393	3.01
X_{29} ：家庭经济状况	-1.0549	0.0831	0.2883	-3.6590	2.89
X_{10} ：经期卫生评分	0.9576	0.2215	0.4706	2.0348	2.61
X_{39} ：性因素	0.9276	0.3014	0.5490	1.6900	2.53
X_{12} ：初婚年龄	-0.3884	0.0524	0.2289	-1.6968	1.47

综合比值比近似值为：

$OR=exp [1.1017(X_{23}^*-X_{23})-1.0549(X_{29}^*-X_{29})+0.9576(X_{10}^*-X_{10})+0.9276(X_{39}^*-X_{39})-0.3884 (X_{12}^*-X_{12})]$

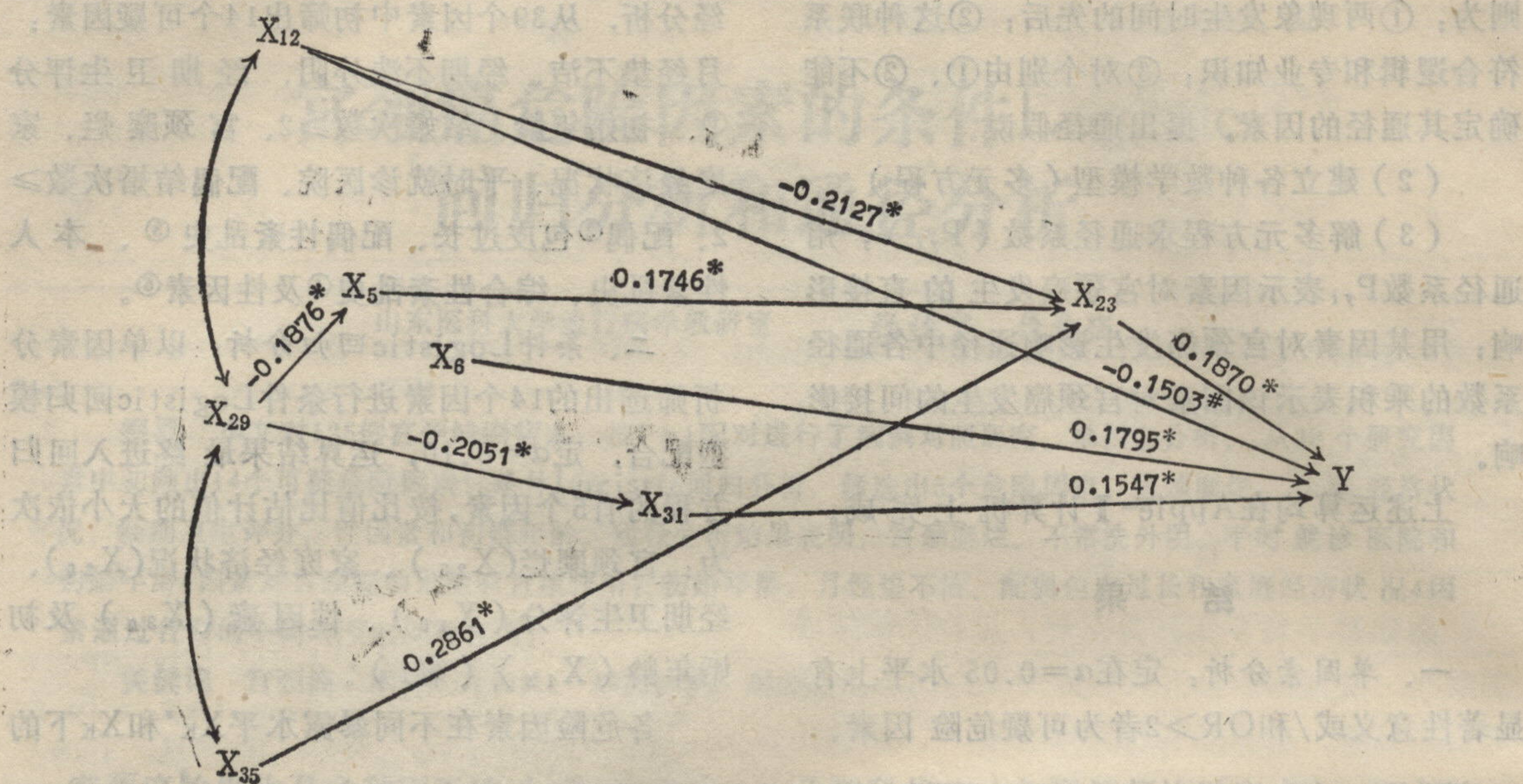
三、通径分析：根据专业知识、结合单因素分析和条件Logistic回归分析的结果，选取月经垫不洁(X_5)、经期不洗外阴(X_6)、经期卫生评分(X_{10})、初婚年龄(X_{12})、宫颈糜烂(X_{23})、家庭经济状况(X_{29})、平时就诊医院(X_{31})、配偶包皮过长(X_{35})及性因素(X_{39})9个因素建立宫颈癌的病因结构模式，并进行通径分析。由通径模式得下列方程：

$Y=P_{y.5}X_5+P_{y.6}X_6+P_{y.10}X_{10}+P_{y.12}X_{12}+P_{y.23}X_{23}+P_{y.29}X_{29}+P_{y.31}X_{31}+P_{y.35}X_{35}+P_{y.39}X_{39}+U_y$

$X_{23}=P_{23.5}X_5+P_{23.6}X_6+P_{23.10}X_{10}+P_{23.12}X_{12}+P_{23.29}X_{29}+P_{23.31}X_{31}+P_{23.35}X_{35}+P_{23.39}X_{39}+U_{23}$
 $X_{10}=P_{10.5}X_5+P_{10.6}X_6+P_{10.29}X_{29}+U_{10}$
 $X_5=P_{5.29}X_{29}+U_5$
 $X_6=P_{6.29}X_{29}+U_6$
 $X_{31}=P_{31.29}X_{29}+U_{31}$
 $X_{39}=P_{39.29}X_{29}+U_{39}$

按上述各方程式计算的结果绘制出宫颈癌危险因素的通径分析模式图(附图)。

①经期卫生评分为月经垫料、月经垫不洁、经期不洗外阴、经期常从事体力劳动、以破布堵经血及经期同房史(每因素均以有为1，无为0)6项之和，其和 ≤ 1 分为0，2~3分为1，4~6分为2)。②配偶：丈夫，若结婚 ≥ 2 次，则包括其前夫。③性紊乱史系指有婚外性伙伴。④综合性紊乱史：本人或/和丈夫有婚外性伙伴；1或2方有为1，2方均无为0。⑤性因素：本人结婚 ≥ 3 次、本人及丈夫均有性紊乱史。



附图 宫颈癌危险因素的通径分析模式

$P < 0.05$

* $P < 0.01$

讨 论

自Domenico (1842) [1]提出结婚与否和宫颈癌的发生有关,一个半世纪以来,人们从常规资料分析、临床观察、实验研究及横断面研究中,先后发现宫颈癌与初婚年龄、性生活开始年龄、结婚次数、性生活紊乱、经期卫生、宫颈糜烂、吸烟、经济文化医疗水平、配偶包皮过长以及多种生物因子(特别是单纯疱疹Ⅱ型病毒和人乳头瘤病毒)等有联系[2~10]。但究竟哪些是宫颈癌的主要危险因素,各家报道不一, Terris等[2]认为,结婚年龄过早、结婚和生育次数、文化水平与宫颈癌及宫颈间变有显著性联系。Reeves等[3]报道,性交年龄过早、多个性伙伴、口服避孕药和HSV—2感染为宫颈癌主要危险因素。Baram等[4]指出,多个性伙伴、性交年龄过早、生育及民族习俗等与宫颈癌有关。陈贻训等[5]报道,宫颈糜烂、性紊乱和经期卫生差是宫颈癌的三大危险因素。张桂宁等[6]对山东省116例宫颈癌病人癌组织HPV 16 DNA相关序列检测的阳性率为50.86%,而正常宫颈组织阳性率仅为

5.56%, $OR=17.60$, $P < 0.001$, HPV 16型感染与宫颈癌之间有很强联系。显然,不同国家或地区、不同民族、不同宗教、不同社会阶层妇女人群的经济文化卫生水平不同,医疗保健服务不同,传统习俗不同,人们对性的态度和观念不同。因此,不同妇女人群宫颈癌的主要危险因素,各因素作用的大小及其相互关系不尽相同。以往有关宫颈癌病因的研究,多数为描述性研究、实验研究,近年来虽有病例对照研究,但多只作了单因素分析,这样,其结论就可能片面性或其中有错的可能性。

本次研究在设计上采取了限制(restriction)、配比等以控制偏倚;在单因素分析的基础上进一步进行条件Logistic回归分析和通径分析。条件Logistic回归分析揭示,宫颈糜烂、家庭经济状况、经期卫生、性因素及初婚年龄过早是宫颈癌的5个高危因素。未被选入方程的9个因素均为已选入方程的有关因素性质相同、但暴露强度较弱者。如本人、配偶的结婚次数和性紊乱史、综合性紊乱史均包涵于性因素而暴露强度弱于性因素;同样,月经垫不洁、经期不洗外阴均系经期卫生评分的一个

组成部分。这突出地显示出条件Logistic回归分析,既反应了各因素的主效应,又消除了潜在的混杂因素的影响作用,只选入其主导的因素,从而更真实、客观地揭示了这些因素与宫颈癌的内在联系。

通径分析表明:①宫颈糜烂(X_{23})、经期不洗外阴(X_6)、平均就诊医院(X_{31})、初婚年龄(X_{12})4个因素对宫颈癌的发生有直接作用;②初婚年龄除有直接作用外,还与宫颈糜烂密切有关,并通过宫颈糜烂起间接作用;③月经垫不洁(X_5)、配偶包皮过长(X_{35})都是通过宫颈糜烂起间接作用的;④家庭经济状况(X_{20})是经由平时就诊医院和月经垫不洁→宫颈糜烂两条通径起间接作用,即家庭经济状况与宫颈癌只是间接联系。附图反映了山东省农村妇女人群宫颈癌的病因网。

李瑛等^[7]在578例宫颈糜烂的活检组织中,发现有癌前病变者64例、早期浸润癌42例、原位癌14例,而宫颈糜烂的发病年龄先于早期癌。Gagnon等^[8]对系统治疗的3000余例慢性宫颈炎追踪观察,没有一例发生宫颈癌。本研究结果,宫颈糜烂者患宫颈癌的OR,单因素分析为3.69(2.09~6.54),条件Logistic回归分析为3.01(1.55~5.86),通径分析验证宫颈糜烂对宫颈癌有直接作用,表明宫颈糜烂是宫颈癌的一个重要高危因素。宫颈糜烂是由多种病原体感染引起的。近年来研究证明,人乳头瘤病毒(HPV)感染与宫颈癌有很强的联系^[6, 9, 10]。性交是病原体侵入宫颈组织的主要途径。因此,性紊乱、不洁性交、经期卫生差及初婚年龄过早等均可致宫颈糜烂、HPV感染机会增加,从而增加宫颈癌发病的危险性。在经济、医疗保健落后的地区,宫颈癌的发病率和死亡率都比较高。本研究条件Logistic回归分析表明,家庭经济状况差是宫颈癌危险因素之一。通径分析阐明,这并不是家庭贫困本身的结果,而是因为家庭贫困,则经期卫生差、平时就诊医院级别低、患宫颈炎或糜烂不能及早发现并系统治疗等密切相关。

目前,宫颈癌防治规划的重点是Ⅱ级预防。本研究运用条件Logistic回归分析和通径分析,首次绘制了宫颈癌的病因网,为本病Ⅰ级预防提供了线索和依据。

Conditional Logistic Regression Analysis and Path Analysis of Risk Factors of Cervical Cancer Zhang Guining, Xu Ai-qiang. Dept. of Epidemiology, Shandong Medical University, Jinan

A case-control study of 125 pairs was carried out in Jinan during 1987~1988. 39 variables were analysed with the single factor analysis, 14 factors were confirmed as the suspicious risk factors. Conditional Logistic regression analysis of the above 14 factors was used. Results showed that the following five risk factors were associated with cervical cancer, i. e., cervical erosion, low family income, poor personal hygiene during menstrual period, sexual factors and early age at marriage. The causal model with 9 risk factors of cervical cancer was set up and studied with path analysis. The result indicated that 4 risk factors had direct and indirect influence on the occurrence of cervical cancer respectively.

Key words Cervical cancer Conditional Logistic regression analysis Path analysis

参考文献

1. Rotkin D. A comparison review of key epidemiological studies in cervical cancer related to current searches for transmissible agents. *Cancer Research* 1973; 33: 1353.
2. Terris M, et al. Comparative epidemiology of invasive carcinoma of the cervix, carcinoma in situ, and cervical dysplasia. *Am J Epidemiol*. 1980; 112(2): 253.
3. Reeves WC, et al. Case control study of cervical cancer in Herrera Province, Republic Panama. *Int J Cancer* 1985; 36: 55.
4. Baram A, et al. Premalignant lesions and microinvasive carcinoma of the cervix in Jewish women: an epidemiological study. *Br J Obstet Gynecol* 1985; 92: 4.
5. 江西省妇女保健院,等.靖安县宫颈癌普查早治12年效果

及其流行病学、病因学研究结果鉴定资料汇编, 1986; 49~66.

6. 张桂宁, 等. 人类乳头瘤病毒与宫颈癌关系的研究. 中华流行病学杂志 1990; 11(1): 31.
7. 李瑛, 等. 宫颈糜烂与子宫颈癌前病变及糜烂型癌的关系. 中华妇产科杂志 1965; 8(2): 221.
8. Gagnon F. Contribution to the study of the etiological and prevention of cancer of the cervix of the uterus. Am J Obstet Gynecol 1950;

60; 516.

9. Durst M, et al. A papillomavirus DNA from a cervical carcinoma and its prevalence in cancer biopsy samples from different geographic region. Proc Natl Acad Sci USA 1983; 80: 3812.
10. Shigeru Y, et al. Human papillomavirus type 16 DNA - Induced malignant transformation of NIH 3T3 cells. J Virol 1986; 57: 572.

(1989年6月15日收稿, 1989年8月28日修回)

家鼠型出血热疫区人群隐性感染及显性感染动态观察

杨占奎¹ 梅志强² 李林森¹ 李继善¹ 米尔英² (指导者)

各地对野鼠型、混合型及家鼠型出血热疫区隐性感染一次性横断面有不少调查与报道, 本文报告1987年1月至1988年8月家鼠型出血热疫区人群隐性感染动态观察及显性感染病后3~4年血清EHFV IgG抗体变化情况。

一、材料与方法:

1. 观察对象和时间: 隐性感染调查是在疫区随机抽取无EHF感染史的村民414人, 1月、8月为观察时间。显性感染选自疫区病人经县以上医院确诊并血清学EHFV IgG抗体阳性者。

2. 采血方法: 常规耳垂采血沾于相同大小的定性滤纸片上, 干燥后4℃保存待检。

3. EHF抗体测定: 将滤纸片剪碎放入小试管中, 处理成1:10的血清, 采用间接免疫荧光法测IgG抗体, 抗原为本省分离的EHFV Rn9株感染Vero-E₆细胞点片。

二、结果:

1. 隐性感染两年4次共检血清1357份, 抗体阳性者24人, 感染率5.80%, 年龄别及性别间感染率均无显著差异 ($\chi^2_{\text{年龄}}=9.66$, $\chi^2_{\text{性别}}=0.08$; $P>0.05$)。

2. 隐性感染达到4次抗体测定的282人中, 4次累计阳性率分别为5.67% (16/282)、6.03% (17/282)、7.45% (21/282) 和8.16% (23/282)。累计阴转率分别为56.25 (9/16)、25% (2/7) 和60% (6/21), 实际阳性率分别为5.67% (16/282)、2.84% (8/282)、3.55% (10/282) 和2.13% (6/282)。

3. 隐性与显性感染抗体滴度的比较, 56例显性感

染者急性期GMT为1:840.31, 病后3~4年GMT下降到1:96.03, 其中45例 (80.36%) 呈下降趋势 (2例完全阴转)、6例 (10.71%) 保持原滴度, 5例 (8.93%) 呈现增高现象。24例隐性感染的抗体滴度均在1:20~1:80之间, 其中4次检测均阳性者3例, 前2~3次的抗体滴度1:80, 第4次下降到1:20, 而没有发现1:20持续4次阳性者, 但可见到阴性阳性交替出现的现象, 大部分仅为一次检测阳性。

三、讨论:

1. 本次调查结果表明, 隐性感染男女间无显著性差异。说明家鼠与人的生活密切程度和家鼠型EHF对每个人的感染机率差别不大。

2. 显性感染和隐性感染除了个体免疫力差异、侵入病原体的毒力大小、数量多少不同而表现出症状有无外, 都有同样的直接或间接接触史。因此, 用带病毒率、带病毒指数等流行病学指标估计人群隐性感染率也是可能的。

3. 调查发现隐性感染具有抗体滴度低, 抗体维持时间短, 可以重复感染的特点, 显性感染抗体效价下降速度快, 且有完全阴转之病例, 对于病后能否获得终身免疫, 隐性感染能否有效防止显性感染, 尚需进一步探讨。

(参加本项工作的还有: 郭崇诚、杨卫国、陈凤叶、陈华犁、黄恩敏、李芳、王顺合、杨根全、史晋明、秦志雄、高秀英、薛云平、李长根等, 一并致谢)

1 山西省临汾地区卫生防疫站

2 山西省卫生防疫站