

原发性肝癌理论流行病学研究

南京医学院 叶本法

理论流行病学在传染病和肿瘤流行病学中都已有了专著和文献报导^[1~4]。在原发性肝癌(以下简称肝癌)病因研究中,尚未见到数学模型的运用。

自1972年以来,江苏省启东县建立了比较系统完整的肝癌发病和死亡登记报告制度,这就给予肝癌理论流行病学研究,奠定了可靠数据的基础。为了阐明识别该地区肝癌病因是属于外因为主抑或内因为主,我们先后使用1973、74、75、76年564名和1,020名确诊为肝癌的病例,进行了普哇松(Poisson)分布和勃奴里(Bernulli)即二项分布的模型研究。

研究方法

一、普哇松分布模式:

1、设计:根据普哇松分布的概率模式要求,将启东县境内新成陆冲积平原上的灌溉渠道几等距离和百万居民住宅基本上均匀分布其间的特点,取1:20万比例的县图,选择较为合适的2×2平方公里为单位面积,画成网格。再将两年内确诊肝癌死亡个例564名,准确地标记在各网格的相应位置上,绘成肝癌流行病学点图。然后统计每单位面积内的病例数。

2、模式:
$$p(x) = \frac{\mu^x \cdot e^{-\mu}}{x!} \dots\dots(1)$$

$$np(x) = n \frac{\mu^x \cdot e^{-\mu}}{x!} \dots\dots(2)$$

$$np(x+1) = np(x) \frac{\mu}{x+1} \dots\dots(3)$$

式中, $x=0,1,2,\dots$ 即单位面积内阳性事件(病例)数;

$p(x)$ 为出现阳性事件数为 x 的理论概率;

$x!$ 为 x 的阶乘, $0!=1$; e 为自然对数底=2.71828;

μ 为在 n 个样本单位内出现阳性事件总

数的均数。

二、二项分布模式:

1、设计:选取1,020确诊肝癌病例,在全县范围内进行家系调查。调查对象为二系:父系、母系;三代:上、同、下三代;三堂:堂兄弟、堂姊妹;三表:姑表、舅表、姨表。回顾20年(即1958~77)以来肝癌发生例数,并按家系记录之。同期,在一公社还调查了21户99人中,对某病的患病情况有否家族集聚性问题,以资对照。

2、模式:
$$(p+q)^k = p^k + kp^{k-1}q + \dots\dots + kpq^{k-1}q^k$$

式中 $p+q=1$ 或 $p=1-q$, $k=1,2,\dots$

设 $p=q=1/2$, $k \rightarrow \infty$ 时,则二项分布呈常态曲线分布;

如 $p \neq q$ 时,则二项分布呈偏态分布。

若某病为家系遗传倾向性的疾病,则该病有家族集聚性。其集聚现象,可以超越二项分布 $(p+q)^k$ 的概率范围。

本文中,设 P 为家族成员的肝癌出现平均概率; P' 为全县居民肝癌出现平均概率。

$P = P_1 P_2 P_3$; $P' = P'_1 P'_2 P'_3$ 。 P_1, P'_1 ; P_2, P'_2 为相应人群中患肝癌显示率和免疫水平反映率,则 $P_1 = P'_1$; $P_2 = P'_2$ 。 P_3 和 P'_3 为相应人群中对肝癌有遗传倾向性的易感率,如家系成员对肝癌有遗传倾向性,则 $P_3 > P'_3$,由此,理论上应: $P_1 \cdot P_2 \cdot P_3 = P > P' = P'_1 \cdot P'_2 \cdot P'_3$

若家族遗传倾向性愈大,则家族出现肝癌集聚性也愈大,概率 P 也愈高,与全县肝癌平均概率 P' 比较,则愈有显著差异。

计算结果

一、普哇松分布:

4平方公里肝癌例数: 0 1 2 3

4 5 6 7 8⁺计
4平方公里(单位)数: 43 84 74 48
25 11 2 3 0 290

均数μ为: $564 \div 290 = 1.9103$

e^{-u} 的对数为: $\mu \log e = (1.9103)(0.434295)$

代入式(2)演化式,得各项理论数(表1)。

结果提示:该地区居民与癌因分布的接触机会似为随机趋势。说明外环境中致癌因素远大于内环境肌体中遗传倾向性的因素。

三、二项分布:

1.调查21户共99人。按家庭人口多少分析患病的人数得表2,括号中数字系按二项分布式展开求得理论数。

现在,实际数与理论数的 X^2 检验见表3,其结果为:

$$1.0538 < X^2_{(1)} 0.30 = 1.074, P > 0.30.$$

表1 肝癌病例在各小区内观察频数
与理论频数比较

小区内肝癌 病例数(X)	观察频数 (A)	理论频数 (T)	(A-T)	$\frac{(A-T)^2}{T}$
0	43	41	2	0.036
1	84	81	3	0.031
2	74	78	-4	0.236
3	48	51	-3	0.051
4	25	25	0	0.083
5	11	10	1	0.444
6	2	3	-1	0.214
7	3	1	2	5.600
总计	290	290	0	$X^2 = 6.63$

自由度为 $6 - 2 = 4$ 查表: $0.90 > P > 0.75$ 无显著差异

说明该公社社员家属患某病的分布情况与二项分布的概率没有显著差别。我们不能根据上述资料认为某病有家族集聚性;病因还得从外环境中探索。

表2 研究患病家族集聚性计算表[7]

病例数	家 庭 人 口									计
	1	2	3	4	5	6	7	8	9	
0	0	1(0.72)	0(0.86)	0(0.65)	0(0.16)	1(0.28)	0	0(0.02)	0(0.01)	2(2.69)
1	0	1(0.96)	3(1.73)	0(1.73)	0(0.52)	1(1.12)	0	0(0.09)	0(0.06)	8(6.20)
2		0(0.32)	0(1.15)	2(1.73)	2(0.69)	2(1.87)	0	0(0.21)	0(0.06)	4(6.13)
3			1(0.26)	2(0.77)	0(0.46)	1(1.66)	0	0(0.28)	0(0.25)	5(3.67)
4				1(0.13)	0(0.15)	0(0.83)	0	0(0.23)	0(0.25)	1(1.59)
5					0(0.02)	1(0.22)	0	0(0.12)	0(0.17)	1(0.53)
6						0(0.02)	0	0(0.04)	0(0.07)	0(0.14)
7								0(0.04)	0(0.02)	0(0.03)
8								0(0.00)	0(0.00)	0(0.00)
9									0(0.00)	0(0.00)
家数	0	2	4	5	2	6	0	1	1	21
人口数	0	4	12	20	10	36	0	8	9	99

病例数为5及以上的理论数共计(0.70647808)

表3 二项分布配合适度 X^2 检验计算表

病例数	实际数	理论数	O-T	$(O-T)^2$	$\frac{(O-T)^2}{T}$
0-1	10	8.90	1.10	1.21	0.1360
2	4	6.13	-2.13	4.5369	0.7401
3 ⁺	7	5.97	1.03	1.0609	0.1777
合计	21	21.00			1.0538

自由度 = $3 - 2 = 1$ $X^2 < X^2_{(1)} 0.30 = 1.074 \therefore P > 0.30$

2.1977年启东肝癌家族史调查研究中,发

现家族中肝癌例数分布如表4。

表4 家族中肝癌例数分布比较

	K							合计
	1	2	3	4	5	6	7	
家族数	587	272	89	38	21	8	5	1020
%	57.55	26.67	8.37	3.73	2.06	0.78	0.46	100.00

根据启东肝癌防治研究所近8年来的比较可信的肝癌登记报告资料统计,该县居民肝癌

发病率接近于50/10万波动, 现设其平均发病率 $P' = 50/10$ 万, 即 $P' = 0.05\%$ 。

据该县组织几次肝癌家族史死亡回顾调查, 其回顾时间(t)达到1958年, 即20年时间里家族史肝癌发生情况, 则 $t = 20$ 。同时, 每个家族调查其两系、三代、三堂、三表等成员, 则其家族人数可达百名左右, 现取其均数 \bar{X} 为100, 则启东一个家族平均肝癌人年发生数为 $P \cdot \bar{X} \cdot t = 0.05\% \times 100 \times 20 = 1$ (例)。即家族成员肝癌发病率 $P = 0.05\%$ 。查 Poisson 变数期望值可信限表〔6〕其P值范围可在0.037%与0.066%之间。则1020家族数出现肝癌例, 按二项分布 $(p+q)^k$ 式展开, 出现1, 2, 3, 4, 5, 6, 7+例肝癌数概率的家族数理论数应于587, 272, 89, 38, 21, 和5的上下范围内波动。因为家族肝癌发病率均值 $P = 0.05\%$ 与启东县居民多年来肝癌平均发病率 P' 几乎相一致, 即 $P \approx P'$, 无统计学差异。这似表明启东肝癌家族集聚现象是在概率分布的许可范围内出现, 非本质差别。致癌因素, 仍须在环境中探索。

讨论与小结

理论流行病学是以数学语言或以数学模型来表达疾病流行过程中各因素之间的定量关系〔1〕。其中概率论更为反映流行过程的客观规律性。17、18世纪概率论学家勃奴里(1737)和普哇松(1873)首创二项分布和普哇松分布的数学模型〔5〕, 这对描述特定时间、空间和人间中发生概率的离散型随机事件, 是一种很好的模型。

普哇松分布是一种离散随机变量分布概型, 是当母体很大, 而概率很小, 即 $P \leq 0.01$ 的分布概型, 是二项分布的一种极限形式。在肿瘤流行病学应用中, 常用来描述时间和空间的特定分布, 以及用来推理流行因素的离散和随

机程度。本课题中主要运用普哇松分布来对肝癌初期防治研究中, 判断外环境致癌因素的分布情况, 籍此作为防治研究的定向参考。

应用二项分布的展开式, 可计算概率的大小, 估计总体率的可信限, 以及研究某病的家族或家庭集聚性等机遇法则。应用二项分布的条件是①各事件是相互独立的, 是相互排斥的。②概率 $P \leq 1\%$ 或 $>99\%$ 时分布非常偏态, 不宜适用。因此, 一般发病率低的疾病, 往往运用普哇松分布公式。本课题如有阴性(病例)家族与分布数据, 使用普哇松分布公式来检验其有无家族集聚性, 比用二项分布式的概型检验更为理想〔7, 8〕。

不论二项分布或普哇松分布, 都是严格要求准确和完整的数据为基础。而且按照它们的特性设计, 从总本中取足够样本或单位来建立模型进行统计处理, 以阐明疾病因素间的定量关系和性质。本课题中两个概型, 均反映了江苏启东肝癌高发区的外环境中, 有一定强度的致癌因素在起着损害该地区居民健康的作用, 各有关方面极须重视, 并应及时采取合理的有效的对策措施。

参 考 文 献

1. Muengh H: Catalytic Models in Epidemiology, pl~95, 1.ed Harvand Univ Press, 1959.
2. Lilienfeld AM: Foundations to Epidemiology, p233~244, 4ed, Oxford Univ Press, 1979.
3. MacDonaed D: Public Health Reports, 76(9): 752, 1961.
4. Radkovsky J: JHEMI, p142~155, 1977.
5. England JM: Medical Reserch a Statistical and Epidemiological Approach, p10~30, 1ed, Churchill Lirrngstone, 1975.
6. 上海第一医学院流行病学教研组: 肿瘤流行病学, 2~23页, 1978.
7. 上海第一医学院统计教研组: 医学统计方法, 149~159页, 第一版, 上海科技, 1979.
8. 薛仲三: 医学统计方法和原理, 157~175页, 第一版, 人卫, 1978.