湖洼地区流行性出血热疫情预测的探讨

安徽医学院流行病学教研组　吴系科　张方振
安徽省颍上县卫生防疫站　郑法扬　樊香亭

安徽省流行性出血热多发于淮河两岸低洼的农垦区。观察证明，本病的流行与疫区黑线姬鼠的密度、年降水量及当年秋季农事活动情况有关。1965年吴系科等报导[1]，以湖洼地为主要疫源地的地区，如参照当地雨量、淮河水位及其邻近湖洼地受淹面积的资料，对预测当年疫情的轻重可能有一定的参考价值。并提出淮河水位差及当地7～9月份的雨量可作为疫情预测的指征之一。1976年安徽省站又提出以黑线姬鼠密度预测本病的看法[2]。为了进一步探讨本病疫情的预测指标和方法，我们对安徽省颍上县王岗公社1964～77年的发病情况及流行因素做了调查，并对它们之间的数量关系作了数理分析。兹报告如下：

材料和方法

选择颍上县王岗公社为观察点。该社位于淮河与颍河之间，海拔仅18.5～25.5米，是典型的湖洼地区。而且，内有唐垛湖与淮河沟通，因此湖地的耕作面积容易受到年降水量及淮河水位差的影响。另外，该社均为农业人口，1964～77年约波动在3.6万～5.3万之间，且秋季以湖内劳动为主。该社是安徽省重点疫区，1958年以来，每年都有病人发生。所以便于从中观察年发病率与有关因素的数量关系。其观察项目有：

一、年发病率：根据县防疫站的疫情报告资料，统计1964～77年各流行年度的实际发病人数，由于该县是老疫区，当地医生对本病诊断的准确性较高，同时基层上报疫情至县站后，一般情况下都经过复核，因此，除个别年份外，均有参考价值。人口资料来自县公安部门。然后，分别计算各流行年度的年发病率(每万人口)。

二、黑线姬鼠密度：以往调查证明，每年9月份的黑线姬鼠密度与该社年发病率有关。故本文鼠情资料系根据县防疫站历年捕鼠记录整理而成，即按夹夜法计算每年9月份的黑线姬鼠密度(%)。

三、淮河水位差：淮河水位高低，直接反映当地年降水量，同时也便于计算定量指标，宜于分析比较，故参照以往报导选用淮河正阳关站7～9月份最高最低水位差(简称“水位差”)，反映年降水量的情况。

四、湖地秋季豆类作物产量(简称作物产量)：王岗公社主要耕地在唐垛湖内，历年来，秋季作物产量的多少与当地居民的农事活动及与湖地接触机会有一定联系，故以此反映农事活动情况。所有资料来自县农业局，我们根据收成大小分成等级，即50万斤以下为1级；50～70万斤为2级；70～90万斤为3级；90～110万斤为4级；110万斤以上为5级。共计5个等级。

分析方法：以该病的年发病率为准则变量，以黑线姬鼠密度、水位差和作物产量等为预测变量，分别进行相关分析，当相关系数的$P<0.05$时，再用来进行多元回归分析。

结　果

一、预测变量与准则变量的相关分析结果如表1。

由表1可见，X_1、X_2、X_3与Y均有一定联系。X_1与Y的指数相关系数略大于直线相关系数。其指数相关系数$r_1=0.6336$。经显著性检验可知，$r_1>r0.05=0.632$，故$P<0.05$。而水位差X_2和作物产量X_3与Y的相关关系，显然均为指数相关。X_2与Y的指数相关系数r_2为-0.8075；X_3与Y的指数相关系数为$r_3=$

0.6559。其相关系数的比较如表2。

表1 预测变量与准则变量的联系

流行年度	准则变量 年发病率（万）y	预测变量 黑线姬鼠密度（%）X_1	水位差（米）X_2	作物产量等级 X_3
1964～	1.63	1.0	3.21	3
65～	9.98	3.2	7.76	5
66～	2.57	（？）	2.25	—
67～	13.57	（？）	3.50	5
68～	0.25	（？）	8.81	1
69～	（？）	（？）	9.12	1
1970～	4.49	1.5	4.74	4
71～	0.82	2.5	5.72	1
72～	0.41	5.4	7.13	1
73～	6.60	21.8	3.97	4
74～	24.61	26.1	3.01	5
75～	0.58	3.2	7.34	1
76～	10.78	7.8	3.60	4
77～	4.05	1.7	5.46	3

表2 预测变量与准则变量指数相关系数的比较

预测变量	指数相关系数	P值
黑线姬鼠密度X_1	$r_1=0.6335$	<0.05
水位差 X_2	$r_2=-0.8075$	<0.01
作物产量 X_3	$r_3=0.9559$	<0.01

二、多元回归分析结果如下：

1.以水位差(X_2)和作物产量(X_3)为预测变量，以1964～77年的年发病率y的对数值(其中缺1966，1969两年)为准则变量，进行三元回归分析。所用回归方程为：

$$\hat{Y}'=\overline{Y}'+b_2(X_2-\overline{X}_2)+b_3(X_3-\overline{X}_3)\cdots\cdots\cdots\cdots(\text{I})$$

$\overline{Y}'$＝年发病率Y的对数值的平均值

$\hat{Y}'$＝预测年发病率的对数值

$b_2=X_2$的回归系数

$b_3=X_3$的回归系数

$\overline{X}_2=X_2$的平均值

$\overline{X}_3=X_3$的平均值

利用二元一次联立方程解出回归系数，即：

$$\begin{cases} b_2\Sigma(X_2-\overline{X}_2)^2+b_3\Sigma(X_2-\overline{X}_2)(X_3-\overline{X}_3) \\ \quad =\Sigma(X_2-\overline{X}_2)(y'-\overline{y}') \\ b_2\Sigma(X_2-\overline{X}_2)(X_3-\overline{X}_3)+b_3\Sigma(X_3-\overline{X}_3)^2 \\ \quad =\Sigma(X_3-\overline{X}_3)(y'-\overline{y}') \end{cases}$$

结果$b_2=-0.0599$； $b_3=0.3297$。故求得经验回归方程式为

$$\hat{Y}'=0.3297X_3-0.0599X_2-0.2237$$

其预测结果如表3。

表3 实际年发病率与预测年发病率的比较

流行年度	实际年发病率（万）	预测年发病率
1964～	1.63	3.74
65～	9.98	9.12
67～	13.57	16.42
68～	0.25	0.38
1970～	4.49	0.47
71～	0.82	0.58
72～	0.41	0.48
73～	6.60	7.19
74～	24.61	17.56
75～	0.58	0.22
76～	10.78	7.57
77～	4.84	2.74

为观察实际年发病率与预测值的相关，我们计算了多元相关系数R，其公式为

$$R=\sqrt{r_2b'_2+r_3b'_3}$$

上式中$r_{2,3}$和$b'_{2,3}$分别代表各自的部分相关系数和标准部分回归系数。运算中先按求解公式将其解出，代入上式即得$R=0.9607>R_{0.01}=0.80$，$\therefore P<0.01$。

2.以$X_1\cdot X_2\cdot X_3$为预测变量，以1964～77年(缺1966，1967，1968，1969)的年发病率Y的对数值为准则变量，进行回归分析。所用方程为

$$\hat{Y}'=\overline{Y}'+b_1(X_1-\overline{X}_1)+b_2(X_2-\overline{X}_2)+b_3(X_3-\overline{X}_3)\cdots\cdots\cdots\cdots(\text{II})$$

利用三元一次联立方程解b_1，b_2，b_3。其结果，$b_1=0.0113$；$b_2=-0.0416$；$b_3=0.3436$。代入方程(II)便得

$$\hat{Y}'=0.0113X_1+0.3436X_3-0.0416X_2-0.421$$

其预测结果如表4。同样以部分相关系数和标准部分回归系数求得两者的多元相关系数，结果$R=0.88>R0.05=0.836$故$P<0.05$。

为了进一步鉴定各预报变量在回归方程中的方差贡献，我们又对$X_1\cdot X_2\cdot X_3$的重要性作了检验，结果如表5。

表4 实际年发病率与预测年发病率的比较

流行年度	实际年发病率	预测年发病率
1964	1.63	3.07
65	9.98	10.24
70	4.49	5.93
71	0.82	0.52
72	0.41	0.49
73	6.60	10.82
74	24.61	29.29
75	0.58	0.45
76	10.78	7.79
77	4.84	2.52

可见作物产量X_3的方差贡献最大，而X_2·X_1次之，但其方差贡献也具有显著意义。

讨　论

流行性出血热的流行环节和流行规律尚未完全明确，本病的疫情能否预测，至今仍有分歧。我们认为，即使本病未知因素尚多，但其流行因素也并非一无所知，特别是鉴于本病高病死率的现状。因此，运用已知流行因素和疫情资料进行预测，是十分必要的，也是值得探索的。因为通过疫情预测，才有可能主动地开展预防工作，减少不必要的盲目性，提高预防效果，最终达到减少发病和控制流行的目的。

表5 X_1·X_2·X_3在回归方程中的重要性检验

变异来源	自由度	估计误差平方和	均方	F	F0.01	$F_{0.05}$	P
四元回归方程	6	0.7313	0.1219				
由X_2X_3估计y'	7	2.0851					
由X_2X_1估计y'	7	3.1260					
由X_3X_1估计y'	7	2.1595					
X_1的方差贡献	1	1.3538	1.3538	11.10	13.74	5.99	<0.05
X_2的方差贡献	1	1.4280	1.4280	11.71	13.74	5.99	<0.05
X_3的方差贡献	1	2.3947	2.3947	19.64	13.74	5.99	<0.01

一、疫情预测的指标：进行疫情预测，首先应确定预测的指标。已经明确疫区野鼠(优势种)的数量变化及其活动情况是决定本病流行的主要因素，因此，野鼠密度是第一个重要指标。就已知资料看，各地所选取的时间不同Потиевский[3]采用春末(5月)调查的黑线姬鼠及东方田鼠的密度为指标，安徽省则采用冬季流行高峰前的野鼠密度为指标。而我们根据观察点的具体情况，推荐采用每年9月份的黑线姬鼠密度作为指标。

鉴于我们所观察的疫源地为湖洼地区，而这些地区的鼠情和农事活动受降水量及洼地附近的淮河水位所制约，因而，我们将淮河涨落的水位差作为疫情预测指标之一。其优点是指标可以数字化，便于作定量分析。

另一预测指标是农事活动，即人群(易感者)与疫源地接触的频率，考虑到本病恒定的流行季节，我们主要取秋季农事活动情况，这项指标从理论上讲应包括诸如野外留宿，兴修水利等活动的人数及时间，但是在实际调查中，无法收集到其准确人数，故以秋季作物产量为指标，间接反映农事活动情况并将其化成等级。以便分析时计算。

流行病学调查证明，影响流行性出血热的因素是极其复杂的，不可能也不必要将有关流行因素的所有数据都纳入回归方程。但是，我们认为最好能包括可反映流行基本环节以及影响本病流行的自然和社会因素的三类指标[4]。因为以单因素预测或仅选用同类影响的几个变量来预测，其效果是不够理想的。而就我们的预测结果来看，不论三元回归分析还是四元回归分析，都具有显著乃至非常显著的统计联系，即预测结果与实际观察结果颇为一致。

二、回归分析的用途和缺点：我们认为，回归分析适用于研究预测变量和准则变量间的相互关系，它在定量化方面较为严密，能对一组预测变量和一个准则变量间的联系作深入的理

论分析，可揭示诸变量间的一些较细致的特点，可用最小二乘法对所列入的变量给以适当权衡。是比较可取的一种预测方法。

回归分析的缺点是其预测的年发病率为一折衷数值，因此不易预测出两极数值。同时，当观察时限较短时，回归方程的稳定性便降低。故要获得较高的预测准确率，应注意下列两点①样本容量应足够大。这样可以减少样本随机波动的影响，降低估计误差。②要有适当数量的预报变量。这里有两重意思，一为数量要适当，二为注意变量间的协同补偿作用。数量太少，供给信息不足。数量太多，不仅计算麻烦，当观察时限短时还会降低预报的稳定性，出现较大偏差。相同作用的变量(如水位差、降水量等)，有时虽与准则变量相关好，但因在回归分析中无相互补偿的作用，同时引入方程，不能改变预测的效果。故应选择易于定量且作用不同的变量，这样即便个别变量相关差，但因有相互补偿作用，预测结果仍可能较为理想。

在这项研究中，我们曾反复使用多种变量进行了回归分析，但以上述三变量最佳。值得注意地是，水位差与准则变量的相关是有一定数值范围的，象水位差低至3.01米以下时，上述那种指数关系，就不复存在。如1966年，全年大旱减产，水位差仅有2.25米，而年发病率却仅为2.57/万，这可能与干旱减产，导致下湖人数减少，野外留宿时间缩短以及鼠密度降低有关。鼠密度仅收集到10年资料，在其相应年份的回归分析中，结果仅次于三元回归分析。这可能与预测变量较多，观察年限较短，因而稳定性下降有关。此外，这项指标常受多种因素的影响，各流行年度的鼠密度可能存在一定误差，也是影响预测的因素之一，今后需要进一步探讨。

结　语

为了探讨湖洼地区流行性出血热的疫情预测，我们采用定点观察和相关与回归分析的方法进行研究，研究结果表明：

一、多元回归分析方法适用于本病多指标的疫情预测。

二、在湖洼型疫源地，选用黑线姬鼠密度、水位差及农作物产量作为预测指标比单个指标更符合本病的客观规律。

三、重要性检验结果证明，在上述三个指标中湖地农作物产量的作用更大。

（郑世鸿、陈亚丽、纪竟雄参加本文资料的收集及整理，特此致谢）

参 考 文 献

1.吴系科等：湖洼地区流行性出血热流行因素的观察，内部资料，1965。

2.安徽省卫生防疫站：根据黑线姬鼠密度预测流行性出血热发病趋势的探讨，内部资料，1976。

3.Потиевскйп ЭГ и др: ЖМЭИ，（7）：112，1975.

4.耿贯一主编：流行病学，下册，159～87，人卫，1979。