

# 综述

## 乙脑流行的数理预测动向

中国医学科学院流行病学微生物学研究所 研究生 曾 光 指导教师 何观清

近年来,随着电子计算机技术的飞速发展,应用数学模型预测疾病流行的报道日益增多。流行性乙型脑炎(以下简称乙脑)是流行因素比较复杂的传染病,研究其预测流行不仅对乙脑防治工作有重要意义,亦可为预测其它疾病流行提供参考。中日两国学者已应用了不同的研究设计和统计方法开展了对这一课题的探索,取得了一定的结果。综述如下:

### 一、传染病数理预测的简要回顾

数学模型方法是流行病学的四种研究方法之一[25]。预测用数学模型大体可分流行模型(epidemic model)和统计算法模型两种。过去人们多应用流行模型来预测疾病流行。应用该种模型,首先根据流行病学特点选择能反映某种疾病流行规律的因子(可泛指影响疾病流行的因素。这里指选择其中的一些因素用作预测方程自变量。如易感者比例,某种致病因素、预防措施因素等)。用作预测方程的自变量(X),然后通过数理分析来表达这些因子和流行指标系指应变量(如发病数、死亡数、流行时间等)——即应变量(y)间的函数关系,并依据调查或实验结果求出方程的参数(指方程的系数,在一定条件下是常数,在不同条件下又是变数)。1908年, Ross首先用此法预测疟疾的流行[26,27]。此后,人们亦用以预测麻疹、天花、水痘、丝虫病、血吸虫等传染病的流行[1,2,31,32]。对流行因素较简单的疾病,应用少许因子及参数即可进行推算预测。例如,1928年Reed-Frost二氏发表了以易感者数和有效接触率(effective contact rate)等为因子和参数的数学模型,以后该模型被用于预测麻疹、天花等病的流行趋势[3,4]。再如,1971年Бароян以易感者数和人群流动量为因子预测了流感的爆发时间[26]。疟疾的流行因素比较复杂, Ross最初预测其流行时,考虑到的因子及参数有人群阳性者比例以及接触率和恢复率(recovery rate)。以后MacDonal Garret-Jones及Grab Dietz等氏把基本繁殖率(basic re production rate)、媒介能量(vectorial capacity)和重复感染(superinfection)等概念,以数学

语言引入了方程[5,6,27];疟疾数学模型趋于复杂化。类此复杂模型的运算必须依靠电子计算机来完成。

近年来,随着电子计算机技术的发展和普及,统计算法的数学模型应用得逐渐增多。应用该种数学模型,系以历年积累的疫情及有关资料为基础,通过统计学处理建立数学方程,并以此方程作出疾病流行的前瞻性推断。包括有各种回归分析、判别分析、聚类分析等在内的多元统计分析法,已被用于预测数学模型研究。如对乙脑、流脑、森林脑炎等疾病的预测[26,33]。这些统计方法还被用于探索非传染病如心血管、恶性肿瘤等病的危险因子。现代预测研究愈来愈依靠流行病学家、数学家和电子计算机工作者的合作才得以完成。

### 二、乙脑流行监察和数理预测的开展

乙脑是流行于亚洲东部地区的自然疫源性疾病。60年代开始,日本首先加强了对乙脑流行的监察工作。如冈山、长崎、爱知等县每年有计划地捕获蚊虫及分离病毒,开展猪血清HI抗体或2-ME敏感抗体的测定,并就乙脑病毒越冬及气象和媒介蚊生态等方面作了大量调查[9~17]。1965年日本厚生省在各地建立了监察系统,重视作乙脑流行的动态分析。70年代以来,我国亦有计划地开展了乙脑流行监察。例如,北方八省区乙脑防治科研协作组作了大量的工作,对三带喙库蚊的生态环境、人群及猪群的抗体水平、病毒分离等方面进行了大量的调查。在乙脑监察中,中日两国学者积累的大量资料和数据,为开展数理预测奠定了基础。同时,飞速发展的电子计算机技术,则为从事复杂的统计运算提供了条件。

首先探索乙脑数理预测的是日本绪方、大崎二氏。1966年,该二氏首先就乙脑在日本流行的“北东进”趋势与各地气温的关系作了相关分析[18]。之后,1967~1973年间,他们先后选用了3~6年的气象、猪血清学、蚊虫密度、经纬度等资料作为待选自变量(即最初参加因子筛选的自变量,应用一定的统计方法筛选其中的一部分组建方程)。观测值,采用

自变量的数目由2个渐增至40个。预测的流行指标有冈山及日本各地的乙脑发病率,乙脑流行中位数日(其英文名称为median date of prevalence of Japanese encephalitis作者规定为:从6月20日开始算起到50%的猪HI抗体阳转日的天数)和乙脑流行众数日(其英文名称为mode date of prevalence of Japanese encephalitis,作者规定为:从6月20日算起到检测出猪HI抗体阳转率最高日的天数)。因其采用自变量观测值的期间多终止于当年7月,而7月已临近流行高峰,故其研究属于近期预测[19,20]。

70年代,我国张可祥氏开始研究预测辽宁省乙脑发病率,同时进行了早期预测和近期预测。早期预测采取上年各季度的气象因子和前两年乙脑发病率的“自由组合”(即将前两年的发病率分别做加、减、乘、除运算,所构成的不同组合)。为待选自变量,年初可算得预测值。近期预测采取当年1~7月份的气象因子为待选自变量,7月末可算得预测值[31]。张可祥还重视研究了乙脑流行与气象因子的相关分析,并以其结果反驳了以往笼统地认为,只要夏季气温高,雨水多乙脑即易流行的模糊认识[32]。

我国开展乙脑数理预测的还有郭存三氏、太原市卫生防疫站、临汾地区卫生防疫站、忻县地区卫生防疫站和屈明元氏等。郭存三氏采取当年1~6月份的气象因子为待选自变量,属于近期预测;其余各家均采用上年1~12月份的气象因子为待选自变量,属于早期预测。我国各地采用资料的年限为6~21年,均长于日本;预测过的应变变量分别为省、地、市的乙脑发病率[33~38]。

无论日本和我国都重视了气象因子作为待选自变量的作用。采用气象因子类别大致相当;计有:平均气温、降水量、相对湿度和日照时间。我国学者多仅采用气象因子为自变量,其原因是缺乏其它方面的积累资料。

### 三、统计方法的应用

业已开展的乙脑数理预测应用的都是统计算法的数学模型,模型的种类同为回归型。

组建回归模型时应用的回归分析法有多种。绪方、大崎二氏在预测研究的初期,取用一个或多个自变量不经统计筛选直接建立和预测指标之间的回归方程,根据这些方程内自变量数目为一个或为一个以上,可分为单元或多元回归方程。该二氏经预测实践认识到,多元回归方程的预测效果优于单元回归

方程。以后随着自变量数目的增加,这种不经统计筛选而建立回归方程的计算方法显得不适用。为此,1972年后该二氏开始研究采用先筛选待选自变量再组建模型的多因素分析方法[21~24]。

多元回归的各种因子筛选法,可以说是简化预测因子的一种理想的数学方法。该法从诸多待选自变量间的相互影响和综合作用的角度,筛选出数个因子用于预测,尤其适用于预测乙脑这样流行因素比较复杂疾病的疫情。绪方和大崎二氏应用较多的是“前进选择法”(forward selection method)“后退选择法”(backward selection method)和“任意选择法”(method of voluntary selection)。前二法为逐步回归算法中的两种。“前进选择法”的特点为:选择自变量时,按其对方程回归贡献的大小顺序,依次引入回归方程;已进入方程的不再考虑剔除。故该算法可概括为“只进不出”。“后退选择法”的特点为:选择自变量时,先将全部待选自变量引入回归方程,然后逐一将作用较小的一个剔除,已剔除方程的不再考虑引入。故该算法可概括为“只出不进”。“任意选择法”是二氏自行设计的统计筛选法,目的在于寻求最优的回归方程。

我国国内乙脑数理预测应用过的因子筛选法有两种。一种是张可祥氏与郭存三氏应用的方法:先作自变量的单相关筛选,再经方差分析以确定预测用的回归模型[31~33]。可以说,这种算法的基础是单因素分析。另一种是山西、保定等地应用过的铁回归分析法[36~38],这是一种非参统计算法的逐步回归法,算法较简单,但利用信息似不充分。

### 四、预测结果与衡量标准

各研究者表达预测结果的方式可分为点估值和可信限区间预测值(为应用回归模型表示预测结果的两种方式。前者指预测值以一个确定的值来表示,后者表示以一定的概率信度,计算出预测值的范围)两种。目前,多数研究者仅采用点估值来表达预测结果。一般认为,点估值和实际发病率值越接近,预测的准确性越高。按照这种标准衡量,1973年绪方、大崎二氏和1978~1979年屈明元氏的点估值预测结果似较理想。我国张可祥氏亦以点估值来表达预测结果,他在评价其1976~1978年辽宁省乙脑发病率的预测结果时,提出了“量级”衡量的方法。即以辽宁省“常年发病水平”15/10万(为该省29年乙脑发病率的均值加上一倍标准差之和)作为衡量标准[32]。如点估值和实际发病率值同时超过或同时低于15/10万,无

表1 乙脑流行强度数理预测研究概况

研究者	应变量	待选自变量	资料年限	统计方法	预测早晚	预测值的表达及衡量标准	预测年数
绪方、大崎 二氏 (1973)	冈山及日本各县乙脑发病率	上年10月至当年7月的日照时间, 平均气温、雨量、经纬度, 50%猪的HI抗体阳转日; 7月末猪的HI抗体阳性率, 以及蚊虫密度对数。	6年	前进法 后退法 任意选择法。	8月初	①点估值 ②近似的可信限区间值	1年 (1973)
张可祥	辽宁省乙脑发病率	早期预测: 前二年发病率自由组合; 上年度各季平均气温、降水量、日照时间。 近期预测: 当年1~7月平均气温、降水量、日照时间。	13年	相关分析和多元回归	早期预测: 年初。 近期预测: 8月初。	点估值	3年 (1976~1978)
郭存三	吉林省乙脑发病率	当年1~6月的平均气温、降水量、日照时间。	21年	同上	7月初	①点估值 ②不详	3年 (1977~1979)
太原市卫生防疫站	太原市乙脑发病率	上年1~12月雨量、日照、时间和平均气温	14年	同上	年初	点估值	2年 (1977~1978)
临汾地区卫生防疫站	临汾地区乙脑发病率	同上	6年	秩回归分析	同上	同上	2年 (1978~1979)
忻县地区卫生防疫站	忻县地区乙脑发病率	上年1~12月平均气温、降水量、相对湿度、日照时间。	16年	同上	同上	同上	1年 (1978)
屈明元	保定地区乙脑发病率	上年1~12月平均气温、日照时间、降水量。	15年	同上	同上	①点估值 ②点估值加减1倍标准估计误差	2年 (1978~1979)

表2 乙脑流行强度数理预测结果一览表

作者	预测年份	预测指标	方程标准估计误差	实际发病率(1/10万)	预测点估值(1/10万)	预测区间	备注
绪方、大崎	1969	冈山县乙脑发病率		0.5	单回归(平均): 4.43 多元回归(平均): 3.02		
大崎二氏	1973	冈山县县南乙脑发病率	方程1: 1.53 方程2: 1.35	0	方程1: 0.9602 方程2: 0.8430	-0.4598~2.3802 -0.5070~2.1930	为近似的
大崎二氏	1973	冈山县县北乙脑发病率	同上	0	方程1: -0.7948 方程2: -1.3581	-2.2148~0.6252 -2.7081~0.0081	可信限区间
张可祥	1976	辽宁省乙脑发病率		1.87	早期预测 12.12 近期预测 9.72	早期预测 11.68 近期预测 8.62	按“量级”衡量
张可祥	1977	辽宁省乙脑发病率		2.03			预测符合
张可祥	1978	辽宁省乙脑发病率		1.60			预测符合
郭存三	1977	吉林省乙脑发病率		0.32		3.24	
郭存三	1978	吉林省乙脑发病率		0.22		4.77	
郭存三	1979	吉林省乙脑发病率		0.07		2.78	
屈明元	1978	保定地区乙脑发病率		1.48		2.49	
屈明元	1979	保定地区乙脑发病率	3.32	5.56		8.59	
太原市卫生防疫站	1977	太原市乙脑发病率		2.25		4.60	
太原市卫生防疫站	1978	太原市乙脑发病率		0.75		14.36	

论二值相差多大, 都认为预测符合; 反之, 如点估值和实际发病率值中, 一值高于而另一值低于15/10万,

无论二值相差多小,也认为预测不符合。仅以此标准衡量,其预测结果满意。1970年后,绪方二氏应用了可信限区间值作为衡量预测准确性的标准(确切讲,该二氏应用的为近似的可信限预测区间)。如实际发病率在预测区间之内,则认为预测符合;否则,认为预测不符合。由于此标准可大致反映出预测的概率和相应区间,似比较单用点估值或应用“量级”衡量的标准客观、可靠。1973年该二氏对冈山县南部和北部地区发病率的预测,实际值均在预测区间之内〔23〕,可视为一次较成功的预测。

鉴于时间、地点、疫情及衡量标准等方面的不同,将各家的预测结果作严格比较是困难的。但可明显看出,多数预测值和实际值间的误差尚较大,方程的标准估计误差(为考察回归方程精度的一项指标。该值越小,回归方程的精度越高;反之亦然。计算公式为:  $Sy = \sqrt{\frac{Q}{n-m-1}}$  式中:  $Sy$  为标准估计误差,  $Q$  为拟合残差,  $n$  为样本数,  $m$  为自变量的个数)。亦偏大,预测的准确性还有待于进一步提高。

### 五、结束语

乙脑数理预测要达到的目标是具有准确性的早期预测,唯此可用于实际防治工作。从目前的预测水平到用于乙脑防治工作,还需要经历艰巨的探索过程。随着我国乙脑防治工作和监察工作的开展,也必然会为开展数理预测提出新的课题。笔者相信,只要在探索的道路上继续努力,不断地积累资料并合理地革新预测方法,是有可能使得预测水平取得突破性的提高的。

### 参 考 文 献

1. Sartwell P: Am J Epidemiol, 103: 138, 1976.
2. Dietz K: ALTA UTAH July. 8, 1974.
3. Fine PE: Am J Epidemiol, 106: 87, 1977.
4. Frost WH: Am J Epidemiol, 103: 141, 1976.
5. Molineaux L et al: Lectures on epidemiological mathematic models, 1~18, (Unpublished data) Shanghai, 1980.
6. Dietz K et al: Bulletin of the world Health Organization, 50: 347, 1974.
7. Becker N: The use of epidemic models, Biometrics, Vol 35, 295, (1979a).
8. B. CVJETA NOVIC et al: Bull WHO, (Supplement 1), 56: 1, 1978.
9. Inatomi S et al: Okayama-Igakkai-Zasshi, (supplement), 78(1): 75, 1966.
10. Inatomi S et al: Okayama-Igakkai-Zasshi, (Supplement), 78(1): 84, 1966.

11. Inatomi S et al: Okayama-Igakkai-Zasshi, (Supplement), 79(1,2): 1~2, 14, 1967.
12. Wada Y et al: Trop Med, 12(2): 79, June, 1970.
13. Hayashi K et al: Trop Med, 15(4): 214, December, 1973.
14. Wada Y et al: Trop Med, 11(1): 37, 1969.
15. Makiya K et al: Jap J Sanitary Zoology, 22(3): 186, 1971.
16. Wada Y: Trop Med, 14(1): 41, 1972.
17. Wada Y: Trop Med, 16(4): 171-199, March, 1975.
18. Ogata M et al: Okayama-Igakkai-Zasshi 80 Supplement 1, 1968.
19. Ogata M et al: Okayama-Igakkai-Zasshi 81 Supplement 11, 23, 1969.
20. Ogata M et al: Okayama-Igakkai-Zasshi 81 Supplement 12, 19, 1970.
21. Ogata M et al: Okayama-Igakkai-Zasshi 83 Supplement 13, 1, 1971.
22. Ogata M et al: Okayama-Igakkai-Zasshi 84 Supplement 14, 11, 1972.
23. Ogata M et al: Okayama-Igakkai-Zasshi 86 Supplement 16, 11, 1974.
24. Ogata M et al: Acta Med Okayama, 28: 125, 1974.
25. 何观清: 流行病学基础、进展及用途, 内部资料, 1981.
26. 耿贯一主编: 流行病学, 上册, 209~221, 人民卫生出版社, 1980.
27. 上海第一医学院编译: 流行病学数学模型参考资料, WHO 流行病学数学模型讨论班, 1~170, 上海第一医学院, 1980, 10.
28. 徐大麟: 国外医学流行病学传染病学分册, 3: 114, 1979.
29. 张可祥: 流行病防治研究, (4): 247, 1978.
30. 宁夏回族自治区卫生防疫站: 流行病防治研究, (2): 99, 1976.
31. 李守正: 流行病防治研究, (1): 10, 1974.
32. 张可祥: 流行病学杂志, 1(2): 72, 1980.
33. 郭存三等: 传染病预测法研究, 1~21页, 1980, (吉林省参加第一次全国流行病学学术会议专著文献)。
34. 山西省卫生防疫站: 关于应用秩回归分析方法预测乙脑疫情介绍, 北方八省区乙脑防治科研协作组第四次会议资料, 1979.
35. 太原市卫生防疫站: 关于应用回归分析法预测乙型脑炎强度的探讨, 北方八省区乙脑防治科研协作组第四次会议资料, 1979.
36. 临汾地区卫生防疫站: 应用秩回归分析预测乙型脑炎流行的趋势, 北方八省区乙脑防治科研协作组第四次会议资料, 1979.
37. 忻县地区卫生防疫站: 应用秩回归方法预测流行性乙型脑炎北方八省区乙脑防治科研协作组第四次会议资料, 1979.
38. 屈明元: 乙脑流行病学特点及应用秩回归早期预测, 医药卫生交流(河北省参加全国流行病学学术会议论文选编), 1980.