

· 论著 ·

时间序列分析预测洞庭湖区退田还湖试点 血吸虫病疫情变化趋势^{*}

赛晓勇¹, 闫永平¹, 张治英¹, 蔡凯平², 李岳生², 周晓农³, 徐德忠^{1**}

(1. 第四军医大学预防医学系流行病学教研室, 陕西西安 710033; 2. 湖南省血吸虫病防治研究所, 湖南岳阳 414000; 3. 中国疾病预防控制中心寄生虫病预防控制所, 上海 200032)

【摘要】 目的 预测洞庭湖区退田还湖试点退田还湖前后血吸虫病疫情的变化趋势, 为血吸虫病防治策略的制定提供依据。方法 应用时间序列中指数平滑法、移动平均法、自相关分析及自回归综合移动平均法 (ARIMA) 对洞庭湖区退田还湖试点血吸虫病疫情建模、分析, 并比较 4 种方法的预测效果。结果 集成垵试点和濠口试点 ARIMA 模型粪检阳性率误差平方和最小, 分别为 100.16 和 22.54。其 ARIMA 预测模型分别为 $\hat{y}_t = 1.2877Y_{t-1} + 0.4246Y_{t-2} - 0.7123Y_{t-3} - 0.4305e_{t-1} + 0.9981e_{t-2} - 0.0544$ 和 $\hat{y}_t = 0.6487Y_{t-1} + 0.3513Y_{t-2} + 0.8671e_{t-1} + 0.0223$; 退田还湖后的患病率预测值高于退田还湖前。结论 退田还湖后血吸虫病粪检阳性率预测值有升高的趋势, 应加强当地的血吸虫病防治工作。

【关键词】 统计预测; 时间序列分析; 血吸虫病; 退田还湖

【中图分类号】 R383.24

【文献标识码】 A

【文章编号】 1001-6627(2004)06-0353-03

自 1998 年我国实行退田还湖的治水策略后, 原来的耕地、居住地等在汛期被洪水淹没, 从而使血吸虫的孳生环境发生了一些变化。为预测退田还湖后疫情的变化趋势, 本研究探索性地应用时间序列方法对洞庭湖区退田还湖试点的疫情资料进行建模分析, 以便为血吸虫病防治策略的制定提供依据。

*** 材料和方法

1 试点概况

洞庭湖区华容县集成垵试点为双退点 (退人、退田), 1998 年后完全废弃用于泄洪, 退田还湖后滞留人口 2 600 人, 面积为 2 200 万 m^2 ; 濠口试点为单退点 (退人不退田), 洪水期人转移, 洪水后返回种田, 常住人口 1 176 人, 面积为 297 万 m^2 。两试点均为湖南省血吸虫病重疫区。

2 粪检资料来源

试点人群 1990~2002 年血吸虫病粪检资料由湖南省血吸虫病防治所及华容县洪山头镇血吸虫病防治站和澧县小渡口血吸虫病防治站提供。

3 方法

运用时间序列分析法建立研究试点的预测模型, 以 1998 年开始退田还湖为界, 计算两试点退田还湖前后连续 5~6 年的粪检阳性率预测值。统计方法采用方差分析、 t 检验, 数据以 $\bar{x} \pm s$ 表示。

3.1 移动平均法 利用一组观察值的均值作为下一期的预测值, 设时间序列为 $x_1, x_2, x_3 \dots$, 可以表示为 $F_{t+1} = \frac{1}{N} \sum_{i=t-N+1}^t x_i$, 式中 x_t 为最新观察值; F_{t+1} 为下一期的预测值, N 为一组观察值的个数。 q 阶移动平均模

型的公式为: $Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q}$, 用自相关系数识别, 它的自相关系数为:

$$r_k = \begin{cases} \frac{-\theta_k + \theta_1 \theta_{k+1} + \dots + \theta_{q-k} \theta_q}{1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2} & 1 \leq k \leq q \\ 0 & k > q \end{cases}$$

时间序列相差 k 个时期两项数据序列之间的依赖程度可用自相关系数 r_k 表示为:

$$\sum_{t=k+1}^n (Y_t - \bar{Y})(Y_{t-k} - \bar{Y}) / \sum_{t=1}^n (Y_t - \bar{Y})^2$$

式中 n 是时间序列 Y_t 的数据的个数; Y_{t-k} 是其滞

后 k 期数据形成的序列。 $\bar{Y} = \frac{1}{n} \sum_{t=1}^n Y_t$, 是时间序列的平均值。 r_k 取值范围为 -1 到 1, $|r_k|$ 与 1 越接近, 说明时间序列的自相关程度越高。

3.2 指数平滑法 用序列过去值的加权均数来预测将来的值, 并给近期的序列值更大的权数, 远期的给以较小的权数。表达式为 $\hat{z}_{t+1} = \alpha z_t + (1-\alpha) \hat{z}_t$, α 为平滑指数, \hat{z}_{t+1} 为下一年预测值, z_t 为当年真实值, \hat{z}_t 为当年预测值。到时期 t 时, 只需知道实际数值和本期预测值两个数据就可预测下一个时间的数值。

3.3 自回归分析 对时间序列求其本期与不同滞后期的一系列自相关系数和偏自相关系数以识别其特性, 用偏自相关系数来判定模型的阶数。 P 阶自回归

* 【基金项目】 国家“十五”科技攻关课题 (No. 2001BA705B08)。

【作者简介】 赛晓勇 (1974—), 男 (回族), 山东济南, 2004 年毕业于第四军医大学预防医学系流行病学与卫生统计学专业, 硕士, 现为在读博士, 硕士期间主要从事遥感流行病学研究。

E-mail: saixiaoyong@163.com

** 【通讯作者】 徐德忠, Tel: 029-83374955

AR(P) 模型的公式为: $Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + e_t$, 它的偏自相关系数满足:

$$\varphi_{ki} = \begin{cases} \varphi_i & 1 \leq i \leq p \\ 0 & p+1 \leq i \leq k \end{cases}$$

偏自相关是时间序列 Y_t 在给定了 $Y_{t-1}, Y_{t-2}, \dots, Y_{t-k+1}$ 的条件下, Y_t 与滞后 k 期时间序列之间的条件相关。它用来度量在滞后 1, 2, 3, $\dots, k-1$ 期时间序列作用已知的条件下 Y_t 与 Y_{t-k} 之间的相关程度, 用 Φ_{kk} 度量。 $\Phi_{kk} = (r_k - \sum_{i=1}^{k-1} \Phi_{k-1,i} \times r_{k-i}) / (1 - \sum_{i=1}^{k-1} \Phi_{k-1,i} \times r_i)$ 。 $k=2, 3, \dots$ 。 式中: $\Phi_{k,i} = \Phi_{k-1,i} - \Phi_{kk} \times \Phi_{k-1,k-i}$, $i=1, 2, \dots, k-1$ 。 r_k 和 r_i 表示 k 期和 i 期的自相关系数。

3.4 自回归综合移动平均法 (ARIMA) 首先判定数据有无随机性、平稳性、季节性, 然后要在预测之前实现最优拟合、建模, 最后进行预测及评价。模型为 ARIMA(p, d, q), 它将移动平均、自回归分析及差分结合起来。确定 3 个参数, 即自回归阶数(p)、差分次数(d)、移动平均阶数(q), 通过差分把时间序列的季节性消除之后(达到数据平稳), 然后建模, 最后估计参数。对非季节数据, 一般求一阶差分即可。若时间序列的季节性的变动周期为 T , 时间序列 Y_t 的一阶季节差分序列 $\nabla_T Y_t$ 为 $\nabla_T Y_t = Y_t - Y_{t-T} (t \geq T)$ 。自相关分析图将自相关系数和偏自相关系数绘制成图, 并标出了置信区间, 利用它可分析时间序列的随机性、平稳性和季节性。随机性是指时间序列各项之间没有相关关系的特性。判定准则: 自相关系数基本上落在置信区间内。平稳性是指时间序列的统计特征不随时间推移而变化。判定准则: 自相关系数 r_k 在 $k > 3$ 时都落入置信区间内并逐渐趋于 0。季节性是指在某一固定时间间隔上重复出现的某种特性。判定准则: 某一时间序列在 $k=2$ 或 3 以后的自相关系数 r_k 值存在着周期性的显著不为 0 的值, 则有季节性^[1]。

4 预测效果评价

比较 4 种方法计算的试点人群历年血吸虫病粪检阳性率预测值与真实值的误差平方和, 较小者为优。

结 果

1 预测模型的建立

集成垵试点和濠口试点 ARIMA 预测模型分别为 $\hat{y}_t = 1.2877Y_{t-1} + 0.4246Y_{t-2} - 0.7123Y_{t-3} - 0.4305e_{t-1} + 0.9981e_{t-2} - 0.0544$ 和 $\hat{y}_t = 0.6487Y_{t-1} + 0.3513Y_{t-2} + 0.8671e_{t-1} + 0.0223$; \hat{y}_t : 预测当年的值, $Y_{t-1}, Y_{t-2}, Y_{t-3}$ 分别为上一年的实测值, 余类推。 e_{t-1} 为上一年预测值的误差, 余类推。其余 3 种方法从略。

2 预测效果比较

4 种方法计算的粪检阳性率预测值与真实值的误

差平方和见表 1。由表 1 可知集成垵试点和濠口试点 ARIMA 模型粪检阳性率误差平方和最小, 分别为 100.16 和 22.54, 应将 ARIMA 模型作为主要的分析方法。

表 1 1994~2002 年不同方法预测效果比较
Table 1 Predictive effect of different methods from 1994 to 2002

方法 Method	误差平方和 Sum of square error	
	濠口试点 Haikou	集成垵试点 Jichengwan
指数平滑法 Exponential smoothing	39.40	190.42
移动平均法 Moving average	39.86	233.86
自相关分析 Autoregressive model	26.63	187.29
自回归综合移动平均法 ARIMA	22.54	100.16

3 退田还湖前后发病率预测值比较

濠口试点退田还湖前后发病率预测值经方差分析显示方差齐($F=0.429, P=0.514$), t 检验结果见表 2, 显示退田还湖后发病预测值高于退田还湖前($P<0.01$)。图 1 为发病趋势预测图, 其发病预测值变化趋势不明显。集成垵试点退田还湖前后发病率预测值经方差分析显示方差齐($F=0.429, P=0.514$), t 检验结果见表 2, 显示退田还湖后发病预测值高于退田还湖前($P<0.01$)。图 2 为发病趋势预测图, 其发病预测值变化有升高的趋势。

表 2 退田还湖前后血吸虫病发病预测值(%) 比较
Table 2 Comparison of prevalence before and after BDOSW *

试点 Village	退田还湖前 Before BDOSW		退田还湖后 After BDOSW	
	年数 Case	预测值 Predicted value	年数 Case	预测值 Predicted value
濠口 Haikou	6	7.41±1.52 ^a	6	8.47±1.10
集成 Jicheng	7	11.75±2.84 ^a	5	27.36±6.38

a 与退田还湖后比较 (vs. after BDOSW), $P<0.01$
BDOSW: Breaking dikes or opening sluice for water store.

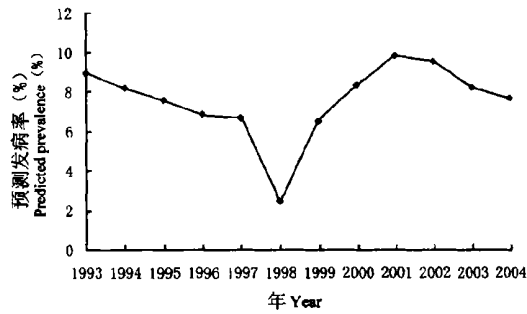


图 1 濠口试点 ARIMA 模型发病趋势预测图

Fig 1 Predicted prevalence by ARIMA model in Haikou

讨 论

传染病疫情的快速、准确预报可以为政府和决策部门合理地分配有限的医疗资源提供重要依据。时间序列分析方法曾被广泛应用于重大传染病的预警预

报。李仲来等^[2~4]应用时间序列方法进行趋势外推,再利用回归模型对 2 种黄鼠鼠疫疫源地进行了预警。程颖恺等^[5]运用时间序列外推模型预测 1975~1995 年和 1985~1995 年两个时间序列上的麻疹疫情,符合率(65.5%、37.0%)差异有显著性($\chi^2=4.54$, $P<0.05$),表现出随时间序列的延长预测符合率增高的趋势,得出了可从资料来源、时间序列长度等方面来判断资料的预测价值的结论。

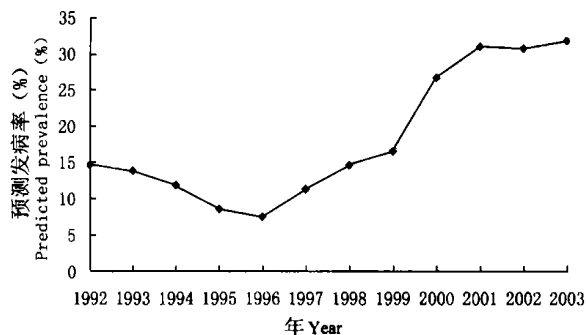


图 2 集成垵试点 ARIMA 模型发病趋势预测图
Fig. 2 Predicted prevalence by ARIMA model in Jichengwan

血吸虫病流行环节复杂,快速监测对血吸虫病防治有重要意义。本研究采用 4 种时间序列方法预测单

退和双退两个试点退田还湖前后血吸虫病疫情,结果显示,ARIMA 模型的预测效果较好,退田还湖前后发病预测值的比较与退田还湖前后的实际发病情况的比较结论相同,即退田还湖后的患病率高于退田还湖前,在一定程度上验证了退田还湖前后血吸虫病患病率变化的趋势。

本研究分析了退田还湖前后血吸虫病发病变化趋势,样本量较小,系统评价退田还湖对血吸虫病流行的影响缺少对照点研究资料,统计预测方法本身也存在着一定的误差,有待进一步研究。

【参考文献】

- [1] 张文彤. SPSS11 统计分析教程(高级篇)[M]. 北京:北京希望电子出版社, 2002. 284.
- [2] 李仲来, 李书宝, 张万荣, 等. 北方 3 种动物鼠疫预警系统研究[J]. 中国地方病防治杂志, 1999, 14(4): 193-196.
- [3] 秦长育, 李仲来. 阿拉善黄鼠疫源地动物鼠疫预报的数学模型[J]. 宁夏医学院学报, 1995, 17(2): 115-117.
- [4] 李仲来, 张万荣. 鄂尔多斯动物鼠疫动态预报[J]. 中国地方病学杂志, 1998, 17(4): 223-225.
- [5] 程颖恺, 曾光. 麻疹、猩红热流行中存在“Z-D 现象”的进一步论证及其预测价值研究[J]. 中华流行病学杂志, 1999, 20(4): 200-203.

【收稿日期】 2004-06-24 【修回日期】 2004-11-12

FORECAST OF SCHISTOSOMIASIS ENDEMIC SITUATION IN THE AREAS OF “BREAKING DIKES OR OPENING SLUICE FOR WATER STORE” IN DONGTING LAKE USING THE TIME SERIES ANALYSIS METHODS

SAI Xiao-yong, YAN Yong-ping, ZHANG Zhi-ying, CAI Kai-ping, LI Yue-sheng, ZHOU Xiao-nong, XU De-zhong

(Department of Epidemiology, School of Preventive Medicine, Fourth Military Medical University, Xi'an 710033, China)

【Abstract】 Objective To forecast the schistosomiasis endemic situation in the areas of “breaking dikes or opening sluice for water store” in Dongting Lake and provide evidences for working-out the control countermeasure of schistosomiasis. **Methods** Time series analysis, including Moving Average(MA), Exponential Smoothing(ES), Autoregressive model(AR) and Autoregressive integrated moving average model (ARIMA model) were applied to forecast the schistosomiasis endemic situation in the areas as above. At one time, the predicted effect of four methods of time series analysis was compared. **Results** The error sum of square of ARIMA model in Jichengwan and Haokou were the smallest(100.16 and 22.54 respectively); The predicted models were $\hat{y}_t = 1.2877Y_{t-1} + 0.4246Y_{t-2} - 0.7123Y_{t-3} - 0.4305e_{t-1} + 0.9981e_{t-2} + 0.0544$ and $\hat{y}_t = 0.6487Y_{t-1} + 0.3513Y_{t-2} + 0.8671e_{t-1} + 0.0223$ respectively; The predicted prevalence after 1998 was higher than that before 1998. **Conclusion** After breaking dikes or opening sluice for water store, the predicted prevalence was increased. Local schistosomiasis preventive work should be reinforced.

【Key words】 Statistical prediction; time series analysis; schistosomiasis; breaking dikes or opening sluice for water store

(上接 345 页)

3 讨论

武进区 1992 年基本消灭疟疾以来,采取以监测为先导,通过疟疾疫情预警系统提供信息资料,指导全面防治。同时实施以加强传染源管理为主的综合性防治措施,效果显著,疟疾疫情在较低水平(0.004‰~0.015‰)波动,但近几年呈小幅回升态势,主要为输入性病例增多所致。因此,应重点加强流动人口管理和疟疾监测,发现病人及时治疗。

监测点的监测结果显示,本地疟疾媒介密度高,叮人率高,容易引起疟疾传播和疫情扩散,应加大防蚊灭蚊力度。

“三热病人”血检是及时发现疟疾传染源的重要措施,基层一线医生要不断提高疟疾防治知识水平和技能,提高检出率,防止漏检、漏报。疟疾病例实行首诊负责制,对流动人口疟疾病例实行属地管理制度,建病案管理卡,使病人能得到正规治疗,防止疫情传播。

【收稿日期】 2004-02-17 【修回日期】 2004-07-23