

# 我国流行性乙型脑炎发病的聚类分析 及其未来发病趋势预测

刘洁, 何钦成

中国医科大学公共卫生学院卫生统计教研室, 辽宁 沈阳 110001

**摘要:** **目的** 利用层次聚类分析法, 对我国各省(直辖市、自治区)流行性乙型脑炎(乙脑)发病进行综合评判分类, 并对全国未来发病趋势进行预测, 为降低乙脑发病提供理论依据。 **方法** 收集全国各省 1992—2011 年乙脑发病率数据; 采用 Q 型聚类分析对我国各省乙脑发病率进行综合评判分类, 应用 GM(1, 1) 模型对全国未来发病趋势进行预测。 **结果** 聚类分析结果显示全国乙脑高发省份为贵州、重庆、云南、河南、四川、安徽、陕西等省(直辖市), 其余省聚为一类, 发病率较低; 我国乙脑发病近年呈下降趋势, GM(1, 1) 模型预测得到乙脑发病的趋势与实际值基本一致, 进一步预测 2012—2014 年乙脑的发病率分别为 0.24/10 万、0.22/10 万、0.20/10 万。 **结论** 我国乙脑的未来发病仍呈下降趋势, 乙脑发病存在明显的高发区, 应对该地区进一步加强防控。

**关键词:** 聚类分析; GM(1, 1) 模型; 发病率; 流行性乙型脑炎

中图分类号: R512.32 文献标志码: A 文章编号: 1003-4692(2013)02-0152-04

## Hierarchical cluster analysis and incidence prediction of Japanese encephalitis in China

LIU Jie, HE Qin-cheng

Department of Statistics, School of Public Health, China Medical University, Shenyang 110001, Liaoning Province, China

Corresponding author: HE Qin-cheng, Email: qche@mail.cmu.edu.cn

**Abstract: Objective** To classify the incidence rates of Japanese encephalitis (JE) in all the provincial-level regions of China by hierarchical cluster analysis and predict the future incidence of JE nationwide, and to provide a theoretical basis for reducing the incidence of JE. **Methods** The data on the incidence rates of JE in all the provincial-level regions of China from 1992 to 2011 were collected. Q cluster analysis was applied to classify the incidence rates of JE in all the provincial-level regions of China. The GM (1, 1) model was used to predict the future incidence of JE nationwide. **Results** The Q cluster analysis showed that Guizhou, Chongqing, Yunnan, Henan, Sichuan, Anhui, and Shaanxi were the regions with high incidence of JE, and other regions, which had relatively low incidence of JE, were classified as a group. In China, the incidence of JE decreased in recent years, and the incidence of JE predicted by the GM (1, 1) model was approximately in accordance with the actual incidence. According to the prediction results by the GM (1, 1) model, the incidence rates of JE would be 0.24/10<sup>5</sup>, 0.22/10<sup>5</sup>, and 0.20/10<sup>5</sup> from 2012 to 2014. **Conclusion** There will be a declining trend in the incidence of JE in China. Control measures should be enhanced in the regions with significantly higher incidence of JE.

**Key words:** Hierarchical cluster analysis; GM (1, 1) model; Incidence rate; Japanese encephalitis

流行性乙型脑炎 (Japanese encephalitis, 乙脑) 是由乙脑病毒引起, 通过蚊媒传播, 以中枢神经系统损害为主的急性传染病, 是我国传染病防治法中规定的乙类传染病。乙脑起病急, 其症状轻重不一, 重型病例幸存者常留有明显的后遗症。近年来随着乙脑疫苗的广泛使用, 我国乙脑病例逐年减少, 但时有局部暴发或流行<sup>[1]</sup>。因此深入分析乙脑流行病学特点, 为了解该疾病的流行规律提供素材, 为进一步有针对性地提出预

防控制措施提供依据具有重要意义。

聚类分析是研究事物分类的一种多元统计方法, 是在没有先验知识的情况下, 根据多个观测指标, 找出能够度量样本或变量之间相似程度的统计量, 并以此为依据, 采取某种聚类方法, 将所有样本或变量分别聚合到不同的类中。其中将多个指标归类的方法又称 R 型聚类, 将多个样品聚类的方法又称 Q 型聚类<sup>[2-4]</sup>。本研究利用我国 2009—2010 年乙脑发病率资料, 分析其变化趋势, 采用样品聚类分析法, 对其进行综合评判并进行分类, 应用 GM(1, 1) 模型对全国未来发病趋势进行预测, 为有关部门制定相关卫生政策提供理论依据。

**作者简介:** 刘洁 (1977-), 女, 博士, 副教授, 从事流行病与卫生统计学研究。Email: liujie1818@yahoo.cn

**通讯作者:** 何钦成, Email: qche@mail.cmu.edu.cn

### 1 材料与方法

1.1 资料来源 收集整理 1992—2011 年《中华人民共和国卫生部公报》转载的我国各省(直辖市、自治区)乙脑发病相关资料;人口资料来自《中国统计年鉴》<sup>[5-7]</sup>。

1.2 聚类分析方法 本研究采用的层次聚类(hierarchical clustering analysis, 又称系统聚类)是实际工作中应用最多的一种将相似的样品或变量归类的聚类方法。层次的含义:开始时将每个样本(或变量)独自视为一类,将距离最近的两类合并;重新计算新类与其他类的距离,再将距离最近的两类合并;重复以上步骤,直至所有的样本(或变量)被并为一类。类与类之间的距离有各种不同的定义方法,常用的方法有类间平均法、类内平均法、最近距离法、最远距离法、离差平方和法等。本研究将每个省视为一个样品,将 2009—2010 年乙脑发病率作为聚类指标,进行 Q 型聚类。类间距离采用常用的类间平均法。

1.3 GM(1,1)模型 GM(1,1)模型的基本原理是对原始数据进行累加,生成具有一定规律的序列,然后建立相应的微分方程模型,用来预测系统的动态发展过程。具体步骤<sup>[8-9]</sup>:

设原始数据序列  $x^{(0)} = \{x^{(0)}(t), t=1, 2, \dots, n\}$ 。

(1) 一阶累加,生成数据序列  $x^{(1)}$ :

$$x^{(1)} = \{x^{(1)}(t), t=1, 2, \dots, n\}, \text{其中 } x^{(1)}(t) = \sum_{i=1}^t x^{(0)}(i) \quad (1)$$

(2) 对累加数列  $x^{(1)}(k)$  计算均值:

$$z(t) = [x^{(1)}(t) + x^{(1)}(t-1)]/2 \quad (2)$$

(3) 建立  $x^{(1)}(k)$  的一阶线性微分方程:

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = u \quad (3)$$

$a, u$  为待定系数。以  $-z(t)$  为自变量,  $x^{(0)}(t)$  为因变量,拟合其直线回归方程,方程的回归系数即为  $a$ , 截距即为  $u$ 。

(4) 预测模型:上述一阶线性微分方程的解即为累加数据的估计值:

$$\hat{x}^{(1)}(t+1) = [x^{(0)}(1) - \frac{u}{a}]e^{-at} + \frac{u}{a}, t=1, 2, \dots, n \quad (4)$$

将式(4)所得数列作递减还原,得原始数列  $x^{(0)}(t)$  的估计值:

$$\hat{x}^{(0)}(t) = \hat{x}^{(1)}(t) - \hat{x}^{(1)}(t-1) \quad (5)$$

(5) 预测模型的拟合精度检验:模型拟合精度高,方可用于外推预测。灰色预测模型 GM(1,1) 的拟合精度检验指标主要有后验差比值  $C$  和小误差概率  $P$ , 其计算公式:

$$C = S_e/S_s$$

$$P = P(|e(t) - \bar{e}| < 0.6745S_e)$$

其中,  $S_e$  为拟合误差的标准差,  $S_s$  为各年实测值的标准差。按后验差比值  $C$  和小误差概率  $P$  综合评定模型拟合精度的标准见表 1。模型拟合精度高,方可用于外推预测。

表 1 后验差比值  $C$  和小误差概率  $P$  综合评定 GM(1,1) 拟合精度标准

拟合精度等级	C 值	P 值
I 级	<0.35	>0.95
II 级	<0.50	>0.80
III 级	<0.65	>0.70
IV 级	≥0.65	≤0.70

1.4 统计分析 采用 SPSS 16.0 软件对数据进行统计分析。

### 2 结果

2.1 2009—2010 年我国 31 个省(自治区、直辖市)乙脑疫情的一般情况 表 1 数据显示,2009 年全国乙脑发病率居前几位的省份(直辖市)为贵州、重庆、云南、河南、四川、安徽、陕西;2010 年全国乙脑发病率居前几位的省份(直辖市)为贵州、云南、河南、重庆、四川、陕西、安徽。各地区具体发病情况见表 2。

表 2 2009—2010 年全国各地乙脑报告发病率(/10 万)

地区	2009 年	2010 年	地区	2009 年	2010 年
北京	0.00	0.00	湖北	0.20	0.22
天津	0.00	0.00	湖南	0.20	0.27
河北	0.02	0.02	广东	0.05	0.08
山西	0.18	0.16	广西	0.11	0.25
内蒙古	0.00	0.00	海南	0.16	0.13
辽宁	0.00	0.00	重庆	0.58	0.75
吉林	0.00	0.00	四川	0.37	0.43
黑龙江	0.01	0.01	贵州	0.61	1.67
上海	0.03	0.07	云南	0.55	1.16
江苏	0.07	0.08	西藏	0.00	0.03
浙江	0.13	0.21	陕西	0.28	0.43
安徽	0.31	0.35	甘肃	0.13	0.21
福建	0.08	0.12	青海	0.02	0.00
江西	0.07	0.12	宁夏	0.03	0.00
山东	0.21	0.14	新疆	0.00	0.01
河南	0.46	0.81			

2.2 2009—2010 年我国 31 个省(自治区、直辖市)乙脑疫情的聚类分析 以乙脑发病率值为聚类指标,对全国 31 个省(自治区、直辖市)进行层次聚类,结果见图 1、2。从图 1 可以看出,2009 年我国各省乙脑发病地理分布情况为:贵州发病率较高,单独作为一组;其次为云南、重庆、河南聚为一组;继续细分组,浙江、甘肃、湖北、湖南、广西、四川、陕西和安徽又被聚为一组。其

余各省(自治区、直辖市)报告发病率均不太突出。图 2 结果显示,2010 年我国各省(自治区、直辖市)乙脑地理分布情况为:贵州、云南、河南和重庆发病率较高,聚为一组;其次为四川、陕西及安徽,发病率较为接近,聚为一组;继续细分组,湖北、湖南、山东、山西、海南又被聚为一组。其余各省(自治区、直辖市)报告发病率均不太突出。聚类结果提示,全国乙脑发病率高的省份,其发病与一般省份的相似程度较低,易被单独分组,而发病率较低的省份,在分组数较少时,不易被单独分列为一组。

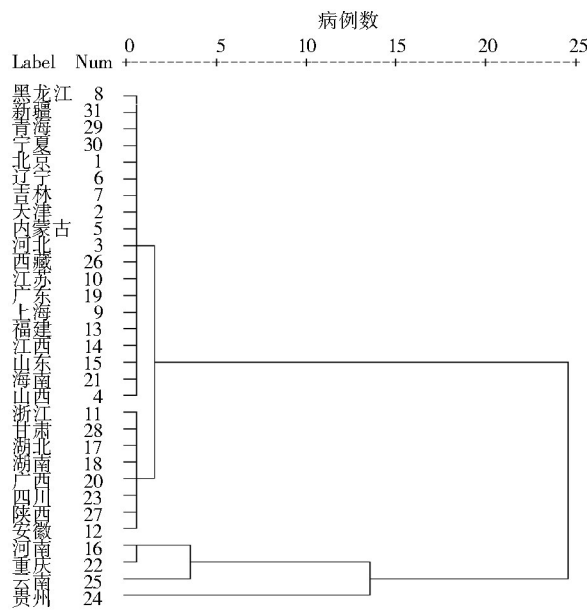


图 1 2009 年我国各地乙脑发病率的聚类分析

Figure 1 Hierarchical cluster analysis of incidence rates of JE across China in 2009

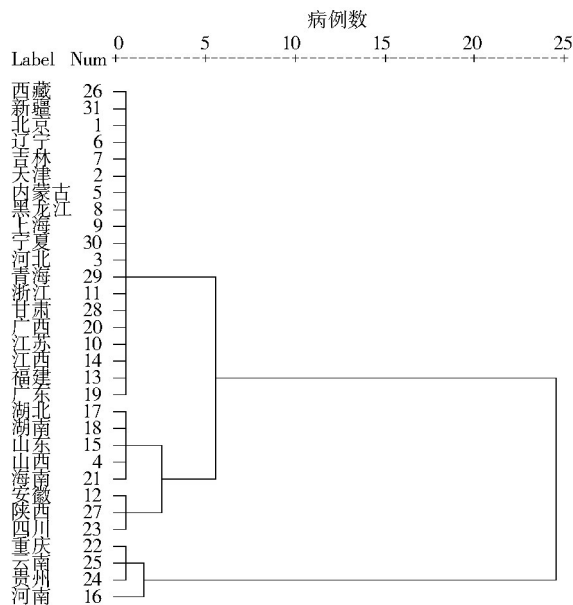


图 2 2010 年我国各地乙脑发病率的聚类分析

Figure 2 Hierarchical cluster analysis of incidence rates of JE across China in 2010

2.3 GM(1,1)预测模型的建立 根据 1992—2011 年全国乙脑发病率,计算累加数列  $x^{(1)}(t)$  和累加均值数列  $z(t)$ ,结果见表 3 第 4、5 列数据。建立以  $x^{(0)}(t)$  为因变量,  $-z(t)$  为自变量的直线回归方程求得:  $a=0.101$ , 截距  $u=1.886$ , 得到预测模型:  $\hat{x}^{(1)}(t+1) = -16.94e^{-0.10t} - 18.67$ , 将 1992—2011 年对应的  $t$  值代入模型,可求得 1992—2011 年的模型回代预测值,结果见表 3 第 6 列数据。进一步计算 1992—2011 年乙脑发病率的标准差  $S_x=0.4495$ , 误差数列  $e(t)$  的标准差  $S_e=0.1277$ , 后验差比值  $C=0.2841$ 。由表 3 最后一列数据判断,所有  $|e(t)-\bar{e}|$  均  $<0.6745$ ,  $S_e=0.3032$ , 即小误差概率  $P=1$ 。根据拟合精度判断标准,可以认为所建立的模型拟合效果理想,拟合精度 I 级。全国发病率进行拟合情况见图 3 的虚线部分,可见实测值与拟合值基本重合,显示 GM(1,1)模型回代拟合较好。进一步预测 2012—2014 年乙脑的发病率分别为 0.24/10 万、0.22/10 万、0.20/10 万。

表 3 1992—2014 年全国乙脑发病率 GM(1,1)模型计算结果(/10 万)

Table 3 Incidence rates of JE in China from 1992 to 2014 as calculated by the GM (1, 1) model (/10<sup>5</sup>)

年度	$t$	实测值 $x^{(0)}(t)$	累加值 $x^{(1)}(t)$	累加均值 $z(t)$	预测值 $\hat{x}^{(0)}(t)$	误差 $e(t)$	$e(t)-\bar{e}$
1992	1	1.73	1.73	-	-	-	-
1993	2	1.54	3.27	2.50	1.63	-0.09	-0.06
1994	3	1.59	4.86	4.07	1.47	0.12	0.15
1995	4	1.32	6.18	5.52	1.33	-0.01	0.02
1996	5	0.87	7.05	6.62	1.20	-0.33	-0.30
1997	6	0.83	7.88	7.47	1.09	-0.26	-0.23
1998	7	1.00	8.88	8.38	0.98	0.02	0.05
1999	8	0.69	9.57	9.23	0.89	-0.20	-0.17
2000	9	0.95	10.52	10.05	0.80	0.15	0.18
2001	10	0.77	11.29	10.91	0.73	0.04	0.08
2002	11	0.65	11.94	11.62	0.66	-0.01	0.02
2003	12	0.58	12.52	12.23	0.59	-0.01	0.02
2004	13	0.42	12.94	12.73	0.54	-0.12	-0.09
2005	14	0.39	13.33	13.14	0.48	-0.09	-0.06
2006	15	0.58	13.91	13.62	0.44	0.14	0.17
2007	16	0.33	14.24	14.08	0.40	-0.07	-0.04
2008	17	0.23	14.47	14.36	0.36	-0.13	-0.10
2009	18	0.29	14.76	14.62	0.32	-0.03	0.00
2010	19	0.19	14.95	14.86	0.29	-0.10	-0.07
2011	20	0.12	15.07	15.01	0.26	-0.14	-0.11
2012	21	-	-	-	0.24	-0.24	-
2013	22	-	-	-	0.22	-	-
2014	23	-	-	-	0.20	-	-

### 3 讨论

本研究利用全国不同地区乙脑的发病率,应用层次聚类分析法,分析疾病的地区分布,为进一步做好乙

(下转第 160 页)

♀ ♀,采集人、时间、地点均同正模。

新种与黑前足茸芒蝇 [*A. ateripraepeda* (He, Huang *et* Feng)]近缘<sup>[7]</sup>,但有各自不同特征可资鉴别:①雄肛尾叶后面观亚后缘两侧各具 1 个耳状突,后者则为 2 个;②雄两侧尾叶后面观,端部内倾呈抱握状,后者则不如是;③雄第 3 背板大部分黄色,后者则为全黑;④雌受精囊椭圆形,后者为圆形。

词源:新种名以汉语拼音“qingchuan”(青川)与拉丁词“ica”组合而成,以示本新种模式种原产地为四川省青川县。

志谢 中国疾病预防控制中心(CDC)传染病预防控制所鲁亮研究员,绵阳市 CDC 刘平,广元市 CDC 刘玉萍、宋晓明,青川县 CDC 袁伟贤,泸州市 CDC 李劲松,德阳市 CDC 周礼晶在蝇类标本采集工作中给予大力支持,特此志谢

参考文献

[1] Pont AC. A review of the Oriental species of *Atherigona* Rondani (Diptera: Muscidae) of economic importance [J]. Control Sorghum Shoot Fly, 1973, 2: 27-104.  
 [2] Pont AC. Some new Oriental shoot-flies (Diptera: Muscidae, genus

*Atherigona*) of actual or suspected economic importance [J]. Bull Entomol Res, 1981, 71: 371-393.  
 [3] Pont AC. Studies on the Australian Muscidae (Diptera), VII the genus *Atherigona* Rondani [J]. Aust J Zool Ser, 1986, 120 Suppl: S1-90.  
 [4] Skidmore P. The biology of the Muscidae of the World. Series Entomology [M]. 29<sup>th</sup>. Hague: Junk Published Dordrecht, 1985: 550.  
 [5] Pont AC, Magpayo FR. Muscid shoot-flies of the Philippine Islands (Diptera: Muscidae, genus *Atherigona* Rondani) [J]. Bull Entomol Res, 1995, 3 Suppl: S111, 123.  
 [6] 范滋德,冯炎,陈之梓,等. 中国动物志. 昆虫纲. 第 49 卷. 双翅目: 蝇科(一) [M]. 北京: 科学出版社, 2008: 869-910.  
 [7] 何建邴,黄步治,窦丰满,等. 中国四川芒蝇亚科芒蝇属一新种 [J]. 寄生虫与医学昆虫学报, 2007, 14(4): 241-243.  
 [8] 方建明,范滋德. 双翅目: 蝇科 [M]//黄复生. 西藏南迦巴瓦峰昆虫. 北京: 科学出版社, 1988: 509-512.  
 [9] 冯炎. 四川蝇科的调查 [J]. 四川动物, 1993, 12(4): 24-26.  
 [10] 刘力. 芒蝇亚科. 芒蝇属 [M]//薛万琦,赵建铭. 中国蝇类. 沈阳: 辽宁科学技术出版社, 1999(1996): 846-864.  
 [11] Shimonaga S. A monograph of the Muscidae of Japan [M]. Tokyo: Tokai University Press, 2003: 112-118.

收稿日期: 2012-11-22

(上接第 154 页)

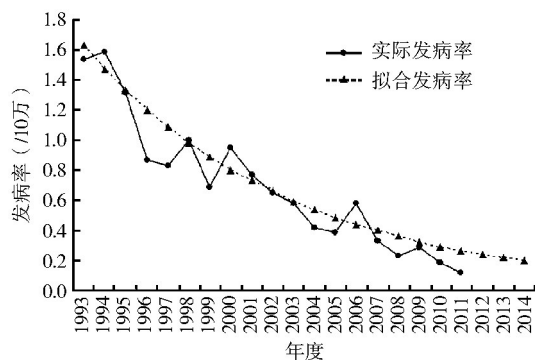


图 3 1993—2014 年全国乙脑发病率预测  
 Figure 3 Predicted incidence rates of JE in China from 1993 to 2014

脑的防控工作提供参考依据。聚类结果提示全国乙脑高发省份为贵州、重庆、云南、河南、四川、安徽、陕西等省(直辖市)。发病率较低的省份,发病率与一般省份的相似程度较高,在分组数较少时,个别乙脑发病率较低的省份未被单独分列为一组。本次研究中聚类分析结果与全国各地乙脑的实际发病特点一致。

1992—2011 年乙脑发病的数据显示,我国乙脑发病率虽然略有波动,但整体呈下降趋势。本研究结果显示 GM(1, 1)模型对乙脑实际值进行了很好地模拟,预测得到乙脑发病的趋势与实际值基本一致,后验差比值 C 和小误差概率 P 均达到拟合标准的最好水平,进一步预测 2012—2014 年乙脑的发病率分别为

0.24/10 万、0.22/10 万和 0.20/10 万。鉴于乙脑对社会的严重危害,应进一步加强综合防控措施,尤其是高发地区的防控措施,努力降低疾病的发生,减少其对社会的危害。

参考文献

[1] 李艺星,尹遵栋,李军,等. 中国 2004—2006 年流行性乙型脑炎流行病学特征分析 [J]. 中国计划免疫, 2007, 13(6): 528-532.  
 [2] 陈建设,陈文培. 聚类分析结合 logistic 回归分析在中医证候诊断量化研究中的应用探讨 [J]. 中国卫生统计, 2009, 26(4): 379-382.  
 [3] 李晓松. 医学统计学 [M]. 北京: 高等教育出版社, 2008: 12.  
 [4] 宋忙华,隋虹,汪婷婷,等. 聚类分析在膳食模式研究中的应用 [J]. 中国卫生统计, 2009, 26(2): 198-199.  
 [5] 中华人民共和国卫生部. 2010 中国卫生统计年鉴 [EB/OL]. (2010-08-31) [2012-10-08]. <http://www.moh.gov.cn/publicfiles/business/htmlfiles/zwgkzt/ptjnj/year2010/index2010.html>.  
 [6] 中华人民共和国卫生部. 2011 中国卫生统计年鉴 [M]. 北京: 中国协和医科大学出版社, 2010: 255-261.  
 [7] 中华人民共和国卫生部. 2012 年中国卫生提要 [EB/OL]. (2012-06-06) [2012-08-12]. <http://www.moh.gov.cn/publicfiles/business/htmlfiles/mohwsbwstjxxzx/s9092/201206/55044.htm>.  
 [8] 金如锋,邱宏,周霞,等. ARIMA 模型和 GM (1, 1)模型预测全国 3 种肠道传染病发病率 [J]. 复旦学报: 医学版, 2008, 35(5): 675-680.  
 [9] 彭志行,鲍昌俊,赵杨,等. ARIMA 乘积季节 GM(1, 1)模型及其在传染病发病预测中的应用 [J]. 数理统计与管理, 2008, 27(2): 362-367.

收稿日期: 2012-12-13