

# 上海市气温变化对医院急诊人次的影响

董英 赵耐青 王爱荣 贾洗怡

**【摘要】** 目的 研究上海市日均气温与三级医院医保急诊人次的关系。方法 应用广义相加模型分析时间序列资料,并在模型中加入自回归模型 AR(P)处理时间序列资料自相关的问题,在控制了与时间有关的中长期趋势、星期效应、节假日效应、SARS 事件、大气污染等混杂因素的基础上,拟合二次函数结合导数原理分析上海市 2002—2004 年间日均气温与三级医院医保急诊人次的关系。结果 当气温低于 14.71℃ 时,气温每升高 1℃ 对应急诊人次增加 RR 值的 95% CI 均小于 1;当气温高于 19.59℃ 时,气温每升高 1℃ 对应急诊人次增加 RR 值的 95% CI 均大于 1;而在温度段为 14.71 ~ 19.59℃,气温每升高 1℃ 对应急诊人次增加 RR 值的 95% CI 包含 1,称之为“最适温度段”。结论 目前上海市日均气温偏离最适温度段时,气温变化对三级医院医保急诊人次有影响。

**【关键词】** 气温; 急诊人次; 时间序列; 广义相加模型

**The impact of air temperature variation on the visits to emergency room in Shanghai** DONG Ying<sup>\*</sup>, ZHAO Nai-qing, WANG Ai-rong, JIA Guang-yi. *Department of Preventive Medicine, Shanghai Traditional Chinese Medicine University, Shanghai 201203, China*

**【Abstract】 Objective** To assess the association between air temperature and emergency room visits among patients covered by medical care program from ‘third-grade’ hospitals in Shanghai. **Methods** Generalized additive model (GAM) was used to analyze time series, and AR(P) was used to deal with auto correlation of time series. After controlling factors as both medium-term and long-term trends, day of the week, vocation, typical pneumonia and pollutants, the association between air temperature and emergency room visits in virtue of quadratic curve and differential coefficient principle were estimated. **Results** When air temperature was below 14.71℃, the increase of 95% confidence interval to relative risk in corresponding emergency room visits along with 1℃ increase of air temperature, was less than 1. However, when air temperature was above 19.59℃, the relative risk’s 95% confidence interval was greater than 1. When air temperature varied at the range of 14.71℃–19.59℃, the 95% confidence interval of the relative risk would include 1. Hence, air temperature range between 14.71℃–19.59℃, was called the optimum temperature range. **Conclusion** Our findings indicated that the current air temperature had an acute impact on the number of emergency room visits among patients covered by medical care program visiting those third grade hospitals in Shanghai.

**【Key words】** Air temperature; Emergency room visits; Time series; Generalized additive model

许多研究表明气温过高过低均可直接或间接导致疾病发生和死亡危险增加<sup>[1-5]</sup>。有关日均气温和死亡率关系的多数研究显示,两者的关系曲线大致呈“V”形<sup>[6]</sup>。“V”形曲线最低部的温度称为“最适温度”,离开最适温度,气温升高或降低都会导致居民死亡或发病的增加。日均气温对城市居民健康的影响是随着不同国家、地区、时期而不同,在当前气候条件下研究上海市日均气温对城市居民健康的影响十分必要。本研究以上海市三级医院日均医保急诊人次

作为健康事件,分析气温变化的健康效应。鉴于资料的特点,本研究应用广义相加模型(general additive model, GAM)分析时间序列资料,并加入自回归模型 AR(P)处理时间序列资料自相关的问题,模型中通过样条函数控制一系列混杂因素的影响。

## 资料与方法

1. 资料:上海市三级医院医保急诊人次资料来自于上海市医疗保险信息中心,选择 2002—2004 年间均在医保范围的 20 ~ 85 岁并于三级医院急诊科就诊的对象,从中按年龄分层随机抽取 50 万人进行研究。上海市中心气象台监测数据购自上海市气象局,包括每日 24 个时点的气温、气湿、气压、露点温度以及相应的日平均值。大气污染资料来自上海市

DOI: 10.3760/cma.j.issn.0254-6450.2009.01.010

作者单位:201203 上海中医药大学预防医学教研室(董英);复旦大学公共卫生学院卫生统计教研室(赵耐青);上海市医疗保险信息中心(王爱荣、贾洗怡)

通信作者:赵耐青, Email: nqzhao@shmu.edu.cn

环境监测中心网上数据库,包括我国常规监测污染物,其中可吸入性颗粒物(PM<sub>10</sub>)、二氧化硫(SO<sub>2</sub>)和二氧化氮(NO<sub>2</sub>)的空气污染指数,根据标准换算为浓度值。

2. 统计学分析:以 GAM 的 Poisson 回归为基本统计方法<sup>[7]</sup>。对于可能存在的星期效应,模型中以哑元变量的形式进行控制。对于时间序列的中长期趋势使用三次样条函数来控制。大气污染物 PM<sub>10</sub>、SO<sub>2</sub>和 NO<sub>2</sub>的混杂作用以对数线性拟合。节假日以及突发公共卫生事件如 SARS 疫情等都可能对医院的急诊人次产生影响,在本研究模型中以哑元变量的方法控制假日效应,通过构造局部样条函数控制 SARS 疫情对医院急诊人次的影响。

每日三级医院医保急诊人次资料是一个时间序列的资料,各个观察单位之间可能是非独立的,拟合模型时还需处理自相关性的问题,根据残差散点图和偏自相关函数(partial autocorrelations function, PACF)诊断残差相关性。如果残差散点图呈零均值随机分布,且偏自相关系数绝对值小于 0.1,可以认为对模型参数的统计推断基本没有影响;如果不满足,则采用下列 GAM+AR(P)随机回归模型进行统计推断:

$$\begin{aligned}
 Y_{t|u_t} &\sim \text{Poisson}(\mu_{t|u_t}) \\
 \log(\mu_{t|u_t}) &= \alpha + \sum_{i=1}^q \beta_i X_i + f'(temp) + \sum_{j=1}^m f_j(Z_j) \\
 &\quad + w_i(week) + w_i(va) + SARS1 \\
 &\quad + SARS2 + u_t \\
 u_t &= a_1 u_{t-1} + a_2 u_{t-2} + \dots + a_p u_{t-p} + \varepsilon_t \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2)
 \end{aligned} \tag{1}$$

式中:  $Y_{t|u_t}$  为观察日  $t$  当天三级医院医保急诊人次,  $Y_{t|u_t}$  服从总体均数为  $E(Y_t)$  的 Poisson 分布;  $\alpha$  为截距;  $X$  为对因变量产生对数线性影响的解释变量,大气污染物浓度等;  $\beta$  为回归模型中的解释变量系数;  $f'(temp)$  为刻画气温与居民死亡人数的效应函数;  $f()$  为 3 次样条函数(cubic spline);  $Z$  为对应变量

发生非线性影响的变量,时间变量、气压、相对湿度、露点温度;  $w_i(week)$  为星期哑元变量,处理星期效应问题;  $w_i(va)$  为节假日哑元变量,处理节假日效应 SARS1、SARS2 为通过构造局部样条函数得到的模型中控制 SARS 疫情因素的变量;  $u_t$  为自相关解释部分。

根据惩罚样条结合 GCV (generalized cross validation) 值判断模型拟合优度,选择模型中包含的成分,以及各样条函数的自由度。

根据气温与医保急诊人次的关系特点,刻画二者关系的效应函数  $f'(temp)$  取二次函数,根据导数原理最终求得日均气温每改变一个单位时,每日医保急诊人次增加的 RR 值及其 95%CI。本研究选择研究日当天与前 6 d 的气温滑动平均值作为气温指标(记为 T06),分析在过去的一周里日均气温变化对急诊人次的影响。

所有统计分析在 S-PLUS 6.2 和 R 语言中实现。

### 结 果

1. 大气污染、气象指标及三级医院医保急诊人次的分布:表 1 为 2002—2004 年上海市大气污染物、气象指标及三级医院日均医保急诊人次分布情况。可见,多数大气污染、气象相关指标年平均值在 3 年间变化不大或多呈波动变化,而 SO<sub>2</sub>、日均急诊医保人次在 3 年间似有逐年上升的趋势(但因本研究利用时间序列资料进行分析,所以更多关注的是指标的时序变化)。

2. 三级医院医保急诊人次的时序图:图 1 显示上海市每日三级医院医保急诊人次是呈一定的周期性和随机波动,并且在所研究的 3 年间呈现一个逐渐上升的趋势,急诊人次随时间的波动和上升趋势需通过样条函数拟合时间变量来控制。

3. 日均气温与三级医院医保急诊人次关系:在控制其他因素情况下,以样条函数拟合气温指标

表 1 2002—2004 年上海市大气污染物、气象指标及日均医保急诊人次分布情况

指标	$\bar{x} \pm s$			
	2002 年	2003 年	2004 年	2002—2004 年
PM <sub>10</sub> (μg/m <sup>3</sup> )	109.40 ± 73.45	97.35 ± 52.35	101.24 ± 56.62	102.66 ± 61.63
SO <sub>2</sub> (μg/m <sup>3</sup> )	35.45 ± 19.51	44.36 ± 23.57	55.70 ± 27.13	45.18 ± 25.01
NO <sub>2</sub> (μg/m <sup>3</sup> )	67.21 ± 23.04	65.46 ± 25.63	70.60 ± 26.74	67.76 ± 25.25
日均气温 (°C)	17.87 ± 7.93	17.63 ± 9.13	18.07 ± 8.58	17.85 ± 8.56
相对湿度 (%)	73.77 ± 13.46	73.08 ± 10.86	69.62 ± 11.92	72.15 ± 12.20
露点温度 (°C)	12.66 ± 9.14	12.38 ± 9.39	11.96 ± 9.25	12.33 ± 9.26
日均气压 (hPa)	1015.97 ± 8.61	1016.54 ± 8.94	1016.50 ± 8.73	1016.34 ± 8.76
日均急诊人次	93.47 ± 22.76	112.65 ± 26.07	125.08 ± 29.27	110.41 ± 29.21

注:日均急诊人次是 2002—2004 年均 在 医保 范围 的 20 ~ 85 岁 并 于 三 级 医 院 急 诊 科 就 诊 的 对 象 中 按 年 龄 分 层 随 机 抽 取 50 万 人 进 行 研 究 的 结 果

T06和急诊人次对数关系(图2)。根据图2显示日均气温与三级医院医保急诊人次的关系特点,选择刻画二者关系的效应函数为二次函数。

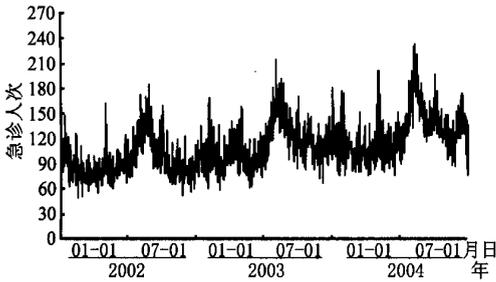


图1 2002—2004年上海市三级医院日均医保急诊人次时序图

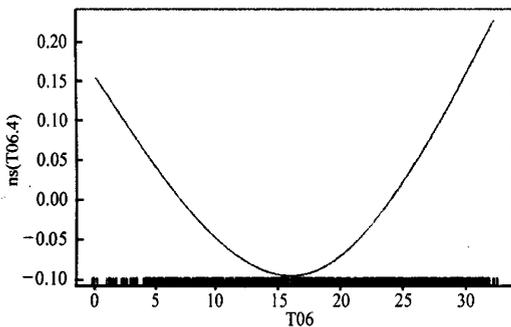


图2 2002—2004年上海市气温与三级医院医保急诊人次对数关系

4.模型评价与修正:首先拟合GAM模型:

$$\log(\mu_i | u_i) = \alpha + \sum_{i=1}^q \beta_i X_i + f^*(temp) + \sum_{j=1}^m f_j(Z_j) + w_i(week) + w_i(va) + SARS_1 + SARS_2 \quad (2)$$

绘制气温效应GAM模型残差散点图和偏自相关图(图3、4)。

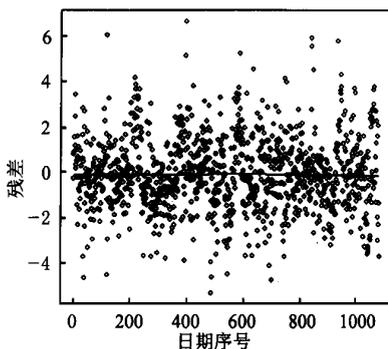


图3 2002—2004年上海市气温对三级医院医保急诊人次的当日效应GAM残差散点图

模型残差散点图提示残差呈零均值随机分布,但偏自相关图显示7阶以内残差偏自相关系数大于0.1,存在自相关性。因此,进一步拟合GAM+AR

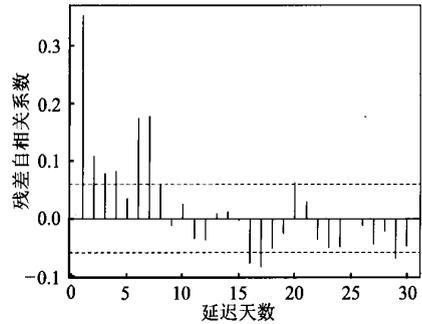


图4 2002—2004年上海市气温对三级医院医保急诊人次的当日效应GAM残差偏自相关图

(P)模型:

$$\log(\mu_i | u_i) = \alpha + \sum_{i=1}^q \beta_i X_i + f^*(temp) + \sum_{j=1}^m f_j(Z_j) + w_i(week) + w_i(va) + SARS_1 + SARS_2 + u_i$$

$$u_i = a_1 u_{i-1} + a_2 u_{i-2} + \dots + a_7 u_{i-7} + \varepsilon_i \quad \varepsilon_i \sim N(0, \sigma_\varepsilon^2)$$

5.定量估计日均气温对三级医院医保急诊人次的影响:在GAM+AR(P)中拟合气温与三级医院医保急诊人次对数的二次函数关系,根据导数原理求得各温度点对应的急诊就诊RR值及其95%CI,结果见表2。

表2 拟合气温效应各温度点对应的日均急诊医保人次RR值及其95%CI

T06	RR值	RR值95%CI
0.20	0.9605	0.9514 ~ 0.9697
0.27	0.9607	0.9516 ~ 0.9698
...	...	...
14.61	0.9940	0.9884 ~ 0.9996
<b>14.71</b>	0.9943	0.9887 ~ 0.9999
14.77	0.9944	0.9888 ~ 1.0000
14.81	0.9945	0.9889 ~ 1.0001
...	...	...
19.51	1.0057	1.0000 ~ 1.0115
19.53	1.0057	1.0000 ~ 1.0115
<b>19.59</b>	1.0059	1.0001 ~ 1.0116
19.63	1.0060	1.0002 ~ 1.0118
...	...	...
32.17	1.0364	1.0269 ~ 1.0461
32.21	1.0366	1.0270 ~ 1.0462

注:结果按气温排序,由于篇幅所限只给出了部分结果,体现“适温段”的两个端点,表中以粗体显示

结果显示,在气温约低于14.71℃时,RR值均小于1(95%CI不包含1);在气温约高于19.59℃时,RR值均大于1(95%CI不包含1)。而在这两个温度范围内气温变化引起急诊人次RR值的95%CI包含1。由此得到对三级医院日均医保急诊人次无显著性影响的温度段为14.71~19.59℃,称之为“适温段”。在此温度范围内RR值的95%CI包含1,不能认为气温在此范围内的改变对急诊人次有影响,随气温升高或降低而偏离此温度段,都引起急诊人次

的增加(图 5)。

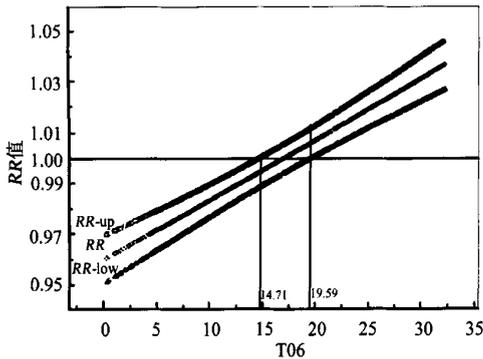


图 5 气温-急诊人次的 RR 值图

## 讨 论

目前我国普通医院门诊就诊行为受医保政策等诸多混杂因素影响较大,而三级医院床位饱和,不能满足居民的住院需求,从而住院率指标也不能很好地反映居民的健康状况。本研究选择受混杂因素影响相对较小的三级医院医保急诊人次作为健康效应指标,基本上能够反映上海市一般经济水平居民的健康需求。以二次函数拟合滑动周日均气温和急诊人次的关系,解决了两者之间关系非线性时的危险度估计问题,根据每个温度点对应的气温每改变一个单位医保急诊人次增加的 RR 值及 95% CI,得到“适温段”为 14.71 ~ 19.59℃,随气温升高或降低而偏离此温度段,三级医院医保急诊人次都增加。

由于三级医院医保急诊人次时间序列存在明显

的自相关,通过拟合 GAM+AR(P)模型进行统计推断校正自相关解释部分,并在模型中控制了众多可能混杂因素的影响。但本研究仍然存在一些不足之处:由于无法得到研究期间逐日流感发病资料,没有对此因素进行控制;另外,研究日当天的天气情况诸如是否降雨、风力的大小等都有可能直接影响居民的就诊行为,故本研究也没有做此方面的控制,但这些因素对急诊人次的影响相对普通门诊要小。在条件允许的情况下,将进一步进行相关内容的深入研究。

## 参 考 文 献

- [1] Demosthenes BP, Christina C, Christos P, et al. Climatological variations in daily hospital admissions for acute coronary syndromes. *Int J Cardiol*, 2004, 94(2-3): 229-233.
- [2] Wilson AM, Wake CP, Kelly T, et al. Air pollution, weather, and respiratory emergency room visits in two northern New England cities: an ecological time-series study. *Environ Res*, 2005, 97(3): 312-321.
- [3] Hajat S, Haines A. Associations of cold temperatures with GP consultations for respiratory and cardiovascular disease Amongst the elderly in London. *Int Epidemiol Assoc*, 2002, 31(4): 825-830.
- [4] Hajat S, Kovats RS, Atkinson RW, et al. Impact of hot temperatures on death in London: a time series approach. *J Epidemiol Community Health*, 2002, 56(5): 367-372.
- [5] Donaldson GC, Keatine WR, Nayha S. Changes in summer temperature and heat-related mortality since 1971 in North Carolina, South Finland, and Southeast England. *Environ Res*, 2003, 91(1): 1-7.
- [6] Curriero FC, Heiner KS, Samet JM, et al. Temperature and mortality in 11 cities of the Eastern United States. *Am J Epidemiol*, 2002, 155(1): 80-87.
- [7] Ahmad RH, Mohammad HF, Masoud Y, et al. Air pollution and hospitalization due to angina pectoris in Tehran, Iran: a time-series study. *Environ Res*, 2005, 99(1): 126-131.

(收稿日期: 2008-06-26)

(本文编辑: 张林东)

## · 读者 · 作者 · 编者 ·

## 本刊对统计学方法的要求 □

**研究设计:** 应告知研究设计的名称和主要方法。如调查设计(分为前瞻性、回顾性还是横断面调查研究), 实验设计(应告知具体的设计类型, 如自身配对设计、成组设计、交叉设计、析因设计、正交设计等), 临床试验设计(应告知属于第几期临床试验, 采用了何种盲法措施等); 主要做法应围绕 4 个基本原则(重复、随机、对照、均衡)概要说明, 尤其要告知如何控制重要非试验因素的干扰和影响。

**资料的表达与描述:** 用  $\bar{x} \pm s$  表达近似服从正态分布的定量资料, 用  $M(Q_n)$  表达呈偏态分布的定量资料, 用统计表时, 要合理安排纵横标目, 并将数据的含义表达清楚; 用统计图时, 所用统计图的类型应与资料性质相匹配, 并使数轴上刻度值的标法符合数学原则; 用相对数时, 分母不宜小于 20, 要注意区分百分率与百分比。□

**统计学分析方法的选择:** 对于定量资料, 应根据所采用的设计类型、资料具备的条件和分析目的, 选用合适的统计学分析方法, 不应盲目套用  $t$  检验和单因素方差分析; 对于定性资料, 应根据所采用的设计类型、定性变量的性质和频数所具备的条件及分析目的, 选用合适的统计学分析方法, 不应盲目套用  $\chi^2$  检验。对于回归分析, 应结合专业知识和散点图, 选用合适的回归类型, 不应盲目套用直线回归分析; 对具有重复实验数据检验回归分析资料, 不应简单化处理; 对于多因素、多指标资料, 要在一元分析的基础上, 尽可能运用多元统计分析方法, 以便对因素之间的交互作用和多指标之间的内在联系做出全面、合理的解释和评价。□

**统计结果的解释和表达:** 当  $P < 0.05$  (或  $P < 0.01$ ) 时, 应对对比组之间的差异具有统计学意义, 而不应说对比组之间具有显著性(或非常显著性)差异; 应写明所用统计分析方法的具体名称(如: 成组设计资料的  $t$  检验、两因素析因设计资料的方差分析、多个均数之间两两比较的  $q$  检验等), 统计量的具体值(如:  $t = 3.45$ ,  $\chi^2 = 4.68$ ,  $F = 6.79$  等); 在用不等式表示  $P$  值的情况下, 一般情况下选用  $P > 0.05$ 、 $P < 0.05$  和  $P < 0.01$  三种表达方式即可满足需要, 无须再细分为  $P < 0.001$  或  $P < 0.0001$ 。当涉及总体参数(如总体均数、总体率等)时, 在给出显著性检验结果的同时, 再给出 95% 可信区间。

本刊编辑部 □