

流感超额死亡率的数学模型研究

胡爱香¹, 余宏杰², 叶冬青¹

【摘要】 流感的季节性流行对人类健康和社会经济造成较大影响,特别是对老年人及慢性基础病患者。而由于流感本身存在种种特殊性,导致很难直接测算流感对死亡的影响。基于流感流行的季节性特征,研究中常使用数学模型的方法,测算流感的超额死亡率,来间接评价流感的影响。本文回顾、分析并比较了既往研究中常用于测算流感超额死亡率的4种数学模型的基本原理、使用方法及应用条件。

【关键词】 流感; 死亡率; 流行病学

【中图分类号】 R511.7

【文献标识码】 A

【文章编号】 1674-3679(2008)04-0372-04

Study on excess mortality of influenza by mathematics models HU Ai-xiang¹, YU Hong-jie², YE Dong-qing¹. 1. Department of Epidemiology and Biostatistics, Anhui Medical University, Hefei 230032, China; 2. Office for Disease Control and Emergency Response, Chinese Center for Disease Control and Prevention, Beijing 100050, China

【Abstract】 Seasonal influenza leads to great impact on human health and economy, especially for old people and people with underlying conditions. It is difficult to calculate the burden caused by influenza directly, as influenza is a kind of special disease. Mathematics models are usually used to calculate the excess mortality of influenza indirectly, based on the seasonal activity characteristic of influenza. This paper reviews, analyses and compares the basic theory, usage and application conditions of 4 kinds of mathematics models.

【Key words】 Influenza; Mortality; Epidemiology

(Chin J Dis Control Prev 2008, 12(4):372-375)

流感是由流感病毒引起的急性呼吸道传染病,感染后可加重原有潜在疾病,引起继发细菌性肺炎、慢性心肺疾患急性发作等。老年人及慢性基础病患者患流感后,易出现严重并发症,死亡率较高^[1,2]。流感可引起季节性流行和全球范围内的大流行,对人类健康和社会经济造成较大影响^[1]。20世纪发生的“西班牙流感”、“亚洲流感”和“香港流感”3次大流行分别导致全球约4 000万、400万和200万人死亡^[3]。每年流感的季节性流行导致全球25万~50万人死亡^[4],其中美国平均每年34 470人死亡(1976~1999年)^[5]、294 128人因呼吸和循环系统疾病而住院(1979~2001年)^[6]。

大流行暴发期间,呼吸道疾病所致死亡多可归因为流感,但在大流行间期,流感季节性流行的影响

远没有大流行严重,并且流感通常是基于临床表现的临床诊断而不是实验室确诊,因此难以直接计算其疾病负担^[7~9]。国外研究多采用数学模型的方法建立死亡率基线,比较流行期与死亡率基线之差,计算流感的超额死亡率而估计其流行导致的疾病负担^[9]。研究流感超额死亡率有助于了解流感流行的严重程度及其疾病负担,掌握流感的时间变化规律及其在不同人群和地区的流行特征,从而指导流感疫苗的应用。本文对流感超额死亡率的相关定义、常用数学模型等方面进行综述。

1 流感超额死亡率的相关定义

流感的超额死亡率即流感流行高峰期的观察死亡率与非流行期季节性死亡率基线之差^[10]。计算时常使用流感流行所致相关疾病的死亡率^[11,12]表示,如流感和肺炎(influenza and pneumonia, P & I)超额死亡率是指观察的P & I死亡率与P & I死亡率基线之差^[13]。

1.1 基线的定义 通过分析往年流感流行期之外的数据,计算其流行导致发病或死亡的一般水平,推算下一年假设不发生流感流行的情况下,期望的发

【作者单位】¹安徽医科大学公共卫生学院流行病与卫生统计学系,安徽合肥 230032

²中国疾病预防控制中心疾病控制与应急处理办公室,北京 100050

【作者简介】胡爱香(1982-),女,安徽怀宁人,硕士。主要研究方向:流行病学。

【通讯作者】叶冬青, E-mail: ydq@ahmu.edu.cn

病率或死亡率水平即为基线^[10,14]。常以 P & I 死亡率基线和全死因 (all cause, AC) 死亡率基线等表示^[10,11,13]。

1.2 流行期及非流行期的定义 流行期的定义有两种:一种定义为实验室病毒学监测中有流感病毒活动的时期^[15~17];第二种定义是将以往流行季节的死因数据代入数学模型,拟合出死亡率基线,基线的 1.645 倍标准差定义为流行阈值,当死亡数连续 2 w 或以上超过流行阈值即为流行期^[10,18]。使用第二种方法定义流行期时,最好同时结合病毒学监测数据对流行期进行验证^[11]。相应地,非流行期则定义为实验室未发现流感病毒活动的时期,或死亡数连续 2 w 或以上超过流行阈值以外的时期。

实际上,在流感流行季节,从出现症状到发生流感相关死亡有一定的时间延迟。虽然有些模型得出的流行期前后几个月的死亡数并没有超过流行阈值,但是此时发生的死亡数已超出死亡数基线且与流行相关,因此有些模型研究常在计算出的流行期前后增加一段时间,作为真正的流行期^[10,11,19]。

1.3 流感超额死亡率的研究指标 流感超额死亡率研究常以医学死亡证明书上列为根本死因的下列疾病作为研究指标,常以国际疾病分类编码 (international classification of disease, ICD) 表示: P & I (ICD-9: 480 ~ 487)^[5,10,11,13,17,19]、AC (所有 ICD-9 编码)^[5,10,11,13,19]、呼吸和循环系统疾病 (cardiovascular and respiratory, C & R, ICD-9: 390 ~ 519)^[5,19]、慢性阻塞性肺部疾患 (chronic obstructive pulmonary disease, COPD, ICD-9: 490 ~ 496)^[19]、缺血性心脏病 (ischemic heart disease, IHD, ICD-9: 410 ~ 414) 等^[19]。P & I、C & R 和 AC 3 个指标拟合死亡率基线,估计流感相关疾病的超额死亡率时,C & R 比 P & I 灵敏度高,比 AC 特异度高^[5]。

2 常用的数学模型

计算流感超额死亡率常用的数学模型有线性回归模型、求和自回归滑动平均模型 (autoregressive integrated moving average, ARIMA)、Poisson 回归模型和率差模型等 4 种。Serfling 早在 1963 年即通过构建线性回归模型,计算流感的超额死亡率^[18],Lui 等于 1987 年对 Serfling 的线性回归模型进行了改进,构建了 Cyclical 线性回归模型^[10]。Thompson 等更进一步改进了 Serfling 模型,构建了 Poisson 回归模型^[5]。Choi 等在 1981 年提出了一种动态预测 P & I 死亡率基线的方法,即 ARIMA 模型^[14]。Barker 等在比较发生率之差的基础上,构建了率差

模型用于计算流感超额死亡率^[18]。

2.1 线性回归模型 常以 $Y = \mu + \epsilon$ 表示 X (自变量) 与 Y (因变量) 之间的线性关系, ϵ 为其他众多未考虑的因素所产生的影响,被看作随机误差。该模型加上关于 ϵ 的 3 个常用假设 (均数为 0, 方差齐性,彼此独立) 构成了最基本的线性回归模型^[20]。季节性流感流行所致相关疾病的死亡数与时间变化存在一定的线性关系,可用线性回归模型进行拟合。但是,季节性流感所致的死亡数与时间之间并非是简单的线性关系,这种线性趋势存在着季节性和周期性。1963 年, Serfling 提出一种数学函数描述 P & I 死亡数的季节性变化,并拟合出了 P & I 死亡率基线,他所构建的线性回归模型如下: $\hat{Y} = \mu + bt + a_1 \cos t + b_1 \sin t$ 其中 μ 是常数项, b 是描述长期趋势的线性项, sine 项和 cosine 项 (傅里叶项) 描述季节性变化, a_1 和 b_1 是调和函数的系数,该模型也称作线性趋势的傅里叶序列^[18]。该模型估计死亡率基线分为 4 步:第 1 步估计长期趋势,第 2 步去除数据的长期趋势,第 3 步估计调整后数据的季节性变化,第 4 步恢复趋势成分^[18]。

为评价 1972 ~ 1985 年,流感流行对美国不同年龄及地区人群死亡的影响,Lui 等在 Serfling 模型的基础上,构建了 Cyclical 回归模型: $0_t = a + b \cdot t + c_1 \cos(2 \cdot t \cdot \div 12) + d_1 \cdot \sin(2 \cdot t \cdot \div 12) + c_2 \cos(4 \cdot t \cdot \div 12) + d_2 \cdot \sin(4 \cdot t \cdot \div 12) + e_t$ ^[10] 其中 t 为过去 5 a 的月份数, a, c_1, c_2 决定 0_t 轴上的截距, e_t 是误差,拟合周期为 12 个月。模型先使用前人研究对流行期的定义拟合出第 1 a P & I 死亡率基线,随后带入前 5 a 非流行期数据建立每年的死亡率基线。Lui 等还对拟合 P & I 死亡率基线的模型进行修正,建立了估计 AC 死亡率基线的模型,并将 AC 死亡数据代入模型计算 AC 死亡率基线^[10] (表 1)。

Simonsen 在估计 1972 ~ 1992 年,美国流感流行对死亡的影响时也使用了 Cyclical 模型: $Y_t = a + b \cdot t + c \cdot t^2 + d \cos(2 \cdot t \cdot \div 52.1667) + f \cdot \sin(2 \cdot t \cdot \div 52.1667) + e_t$ 其中 t 是过去 5 a 的周数, c 是偏离线性趋势的 2 次方系数,拟合周期为 52.1667 w,其余参数与上述模型相同^[11]。

2.2 求和自回归滑动平均模型 (ARIMA) 求和自回归滑动平均模型是 Box 等 1976 年提出的一类随机模型,模型的 d 阶差分是一个静态的混合自回归滑动平均过程^[14]。流感流行导致的每周 P & I 死亡数时间序列不是一个静态的随机过程,适于用 ARIMA 模型进行分析^[14]。Choi 等于 1981 年,基于 ARIMA 模型原理,构建了估计 P & I 死亡数基线或

表 1 流感超额死亡率的部分研究设计一览

Table 1 Some study design of excess mortality of influenza

研究起止时间	所用模型	研究对象	研究指标	数据周期
1972 ~ 1985 ^[10]	cyclical 回归模型	美国全人群	P & I, AC	月数据
1972 ~ 1992 ^[11]	cyclical 回归模型	美国全人群	P & I, AC	周数据
1967 ~ 1978 ^[13]	ARIMA 模型	美国全人群	P & I, AC	周数据 & 月数据
1976 ~ 1999 ^[5]	Poisson 回归模型	美国全人群	P & I, C & R, AC	周数据
1996 ~ 1999 ^[19]	Poisson 回归模型	香港全人群	C & R, P & I, COPD, IHD 和 AC	周数据
1968 ~ 1973 ^[17]	率差模型	美国波特兰市健康计划人口	P & I, 除 P & I 外的其他上呼吸道疾病、急性心衰或炎症不包括 P & I	无

AC 死亡数基线的模型:

$$(1 - \beta - \beta^2 - \beta^3)(1 - \beta^{52})(1 - \beta^{52}) Y_t = a_t \quad (1)$$

$$(1 - \beta - \beta^2 - \beta^3)(1 - \beta^{52})(1 - \beta^{52}) Y_t = (1 - \beta^{52} - \beta^{52}) a_t \quad (2)$$

其中模型(1)是自回归过程,模型(2)是混合自回归滑动平均过程^[14]。该模型在拟合时需要 100 w 以上的数据,第 1 个 2 a 的死亡数基线通过将非流行期数据代入傅里叶序列得到,然后用模型(1)拟合每两年的非流行期数据得到下一年的死亡数基线。为避免剔除流行期数据而低估死亡数,模型(1)使用拟合出的死亡数基线替换流行期的实际死亡数。重复这样操作,直到模型(1)得出所有年份的死亡数基线,然后将所有年份死亡数基线代入模型(2),拟合出下一年的基线^[14]。

2.3 Poisson 回归模型 Poisson 回归模型是广义线性模型的一种,通过一定变量变换,满足或近似满足线性模型分析的要求,从而借助线性模型的优良性质、分析思路,解决或近似解决非线性模型的建模、参数估计和模型评价等问题^[21]。模型因变量的分布是非连续性资料,因变量的函数估计值等于自变量的线性预测。该模型适用于分析以人群为基础的稀有疾病和卫生事件资料^[21]。

季节性流感所致的相关疾病的死亡分布符合 Poisson 回归模型的应用条件,2003 年,为估计美国 1976 ~ 1999 年流感流行所导致的相关死亡,Thompson 应用了 Poisson 回归模型:

$$Y = \exp \left[\begin{matrix} \theta_0 + \theta_1[t] + \theta_2[t^2] + \theta_3[\sin(2 / 52)] + \theta_4[\cos(2 / 52)] \\ + \theta_5[A(H1N1)] + \theta_6[A(H3N2)] + \theta_7[B] + \theta_8[RSV] \end{matrix} \right]$$

其中表示某年龄组某一周死亡数,是偏移项,等于年龄别人口数的对数,t 表示所研究时间序列的周数, θ_0 表示截距, θ_1 表示线性时间趋势, θ_2 表示非线性时间趋势, θ_3 和 θ_4 表示死亡数的季节性变化, $\theta_5 - \theta_8$ 为某一周阳性毒株比例的相关系数^[5]。该模型拟合每周流感病毒阳性株比例与每一种疾病的死亡数,得到每一种疾病与流感相关的死亡数^[5]。有些国家的研究发现,呼吸道合胞病毒(respiratory syncytial virus,RSV)与流感病毒的流行期经常重叠,且 RSV 感染所致的并发症与流感的并发症类似,因此该模型也将 RSV 作为混杂因素引入^[5,16,19]。

Wong 等也使用了上述 poisson 回归模型,同时引入每周平均温度和相对湿度作为协变量,估计香港特区 1996 ~ 1999 年每年流感相关的超额死亡率^[19]。

2.4 率差模型 为估计 1968 ~ 1969 和 1972 ~ 1973 两个 A 型流感季节,14 岁以上成人的超额发病率和超额死亡率,Barker 等构建了率差模型^[17],拟合每年 12 月至次年 3 月的冬季基线率^[17]。Izurieta 也使用率差模型建立了流行间期基线指流感季节流行期以外的时期)和夏季基线指非流感季节流行期以外的时期),测算美国流感季节(10 月 ~ 5 月),流感对儿童呼吸道疾病住院的影响和流感的超额死亡率^[8,16]。

3 四种模型比较

Serfling 模型、Cyclical 模型、ARIMA 及 Poisson 回归模型均是前瞻性预测,利用往年死亡或病毒学数据估计随后一年的死亡率基线,然后将观察死亡率与死亡率基线比较,得出流感超额死亡率。率差模型较前 3 种方法简单,且需要做的假设也较少。

3.1 ARIMA 模型与 Serfling 模型比较 Choi 等使用构建的 ARIMA 模型估计了美国 1967 ~ 1978 年 8 个流行期,流感的超额死亡数,并将结果与 Serfling 模型计算的结果进行比较发现:ARIMA 模型估计每月 P & I 死亡数基线的精确性不如 Serfling 模型好,但能较精确地预测每周 P & I 死亡数^[13]。与 Serfling 模型相比,ARIMA 模型进行预测时,能达到最小均方误差的最佳特性^[14]。Choi 的研究提示,在计算流感超额死亡数时,可以综合使用 Serfling 模型和 ARIMA 模型两种方法^[13]。

3.2 Poisson 模型与 ARIMA 模型及 Cyclical 模型比较 Poisson 回归模型引入了病毒学数据,估计流感各亚型病毒活动与流感相关死亡的关系,较 ARIMA 模型和 Cyclical 模型精确,但是模型的拟合过程也更加复杂^[5,6]。与 ARIMA 模型和 Cyclical 模型相比,Poisson 模型还引入了可能会影响季节性死亡的温度和湿度因素^[5,6]。

3.3 率差模型与 Cyclical 模型及 Poisson 模型比较

率差模型的夏季基线估计超额死亡数比 Cyclical 模型和 Poisson 模型估计值要高,但是,率差模型的流行间期基线估计的结果与 Cyclical 模型和 Poisson 模型的估计结果非常相似^[8]。与其他模型相比,率差模型的使用更加简便,因此其应用范围也更加广泛^[8]。

数学模型在研究流感超额死亡率时存在一些局限性。流感需要通过实验室检测确诊,而临床诊断的死因数据进行超额死亡率的研究结果可能与客观实际不一致^[10]。模型研究结果的精确性也未得到证实,在拟合非流行期死亡率基线时,只能估计流行期以外的超额死亡率,不可能精确估计流行期的超额死亡率^[10]。除此之外,不同的地理区域,人口移动、环境因素以及不同的循环毒株对流感流行也有一定的影响,模型在计算流感超额死亡率时,常常会遇到这些混淆因素的影响,而无法计算出真正的流感导致的死亡负担^[22,23]。流感的疾病负担除超额死亡率之外,还包括发病率、就诊率以及由于流感导致的工厂旷工和学校缺勤^[24],因此,流感的超额死亡率不可能精确解释流感所致的全部疾病负担。

尽管存在一定的局限性,但数学模型是定量测算流感疾病负担的最常用方法。迄今为止,我国尚未开展流感超额死亡率相关研究。通过回顾、整理我国现有的死因报告和实验室病毒学监测数据,借鉴国外研究选择合适模型,尝试拟合我国流感超额死亡率基线,可为我国流感超额死亡率的评价和制定流感疫苗的应用策略提供基础数据。

【参考文献】

- [1] Cox NJ, Subbarao K. Influenza [J]. *Lancet*, 1999, **354**(9186): 1277-1282.
- [2] Nicholson KG, Wood JM, Zambon M. Influenza [J]. *Lancet*, 2003, **362**(9397): 1733-1745.
- [3] Oxford JS. Influenza A pandemics of the 20th century with special reference to 1918: virology, pathology and epidemiology [J]. *Rev Med Virol*, 2000, **10**(2): 119-133.
- [4] WHO. Influenza[OL]. <http://www.who.int/mediacentre/factsheets/fs211/en/index.html>.
- [5] Thompson WW, Shay DK, Weintraub E, et al. Mortality associated with influenza and respiratory syncytial virus in the United States [J]. *JAMA*, 2003, **289**(2): 179-186.
- [6] Thompson WW, Shay DK, Weintraub E, et al. Influenza-associated hospitalizations in the United States [J]. *JAMA*, 2004, **292**(11): 1333-1340.
- [7] Monto AS. Global burden of influenza: what we know and what we need to know [J]. *International Congress Series*, 2004, **1263**: 3-11.
- [8] Thompson WW, Comanor L, Shay DK. Epidemiology of seasonal influenza: use of surveillance data and statistical models to estimate the burden of disease [J]. *J Infect Dis*, 2006, **194**(Suppl 2): S82-S91.
- [9] Monto AS. Influenza: quantifying morbidity and mortality [J]. *Am J Med*, 1987, **82**(6A): 20-25.
- [10] Lui KJ, Kendal AP. Impact of influenza epidemics on mortality in the United States from October 1972 to May 1985 [J]. *Am J Public Health*, 1987, **77**(6): 712-716.
- [11] Simonsen L, Clarke MJ, Williamson GD, et al. The impact of influenza epidemics on mortality: introducing a severity index [J]. *Am J Public Health*, 1997, **87**(12): 1944-1950.
- [12] Simonsen L, Reichert TA, Viboud C, et al. Impact of influenza vaccination on seasonal mortality in the US elderly population [J]. *Arch Intern Med*, 2005, **165**(3): 265-272.
- [13] Choi K, Thacker SB. Mortality during influenza epidemics in the United States, 1967 - 1978 [J]. *Am J Public Health*, 1982, **72**(11): 1280-1283.
- [14] Choi K, Thacker SB. An evaluation of influenza mortality surveillance, 1962 - 1979. I. Time series forecasts of expected pneumonia and influenza deaths [J]. *Am J Epidemiol*, 1981, **113**(3): 215-226.
- [15] Langmuir AD, Henderson DA, Serfling RE. The Epidemiological Basis for the Control of Influenza [J]. *Am J Public Health Nations Health*, 1964, **54**: 563-571.
- [16] Izurieta HS, Thompson WW, Kamarz P, et al. Influenza and the rates of hospitalization for respiratory disease among infants and young children [J]. *N Engl J Med*, 2000, **342**(4): 232-239.
- [17] Barker WH, Mullooly JP. Impact of epidemic type A influenza in a defined adult population [J]. *Am J Epidemiol*, 1980, **112**(6): 798-811.
- [18] Serfling RE. Methods for current statistical analysis of excess pneumonia-influenza deaths [J]. *Public Health Rep*, 1963, **78**: 494-506.
- [19] Wong CM, Chan KP, Hedley AJ, et al. Influenza-associated mortality in Hong Kong [J]. *Clin Infect Dis*, 2004, **39**(11): 1611-1617.
- [20] 王松桂,陈敏,陈立萍. 线性统计模型: 线性回归与方差分析 [M]. 北京: 高等教育出版社, 1999. 1-4.
- [21] 陈峰, 编著, 陈启光. 医用多元统计分析方法 [M]. 北京: 中国统计出版社, 2000. 115-125.
- [22] Dushoff J, Plotkin JB, Viboud C, et al. Mortality due to influenza in the United States—an annualized regression approach using multiple-cause mortality data [J]. *Am J Epidemiol*, 2006, **163**(2): 181-187.
- [23] Rizzo C, Bella A, Viboud C, et al. Trends for influenza-related deaths during pandemic and epidemic seasons, Italy, 1969 - 2001 [J]. *Emerg Infect Dis*, 2007, **13**(5): 694-699.
- [24] Simonsen L. The global impact of influenza on morbidity and mortality [J]. *Vaccine*, 1999, **17**(Suppl 1): S3-S10.

(收稿日期: 2008-05-16)

(李婷婷校)