

利用样本数据估计流感相关死亡负担的研究进展

李飒 冯录召 余宏杰

【关键词】 样本数据; 流感; 死亡; 统计模型

Research progress on estimation of influenza associated mortality burden based on sample data Li Sa, Feng Luzhao, Yu Hongjie. Division of Infectious Disease, Key Laboratory of Surveillance and Early-warning on Infectious Disease, Chinese Center for Disease Control and Prevention, Beijing 102206, China

Corresponding author: Yu Hongjie, Email: yuhj@chinacdc.cn

【Key words】 Sample data; Influenza; Deaths; Statistical model

流感季节性流行和大流行均可导致大量门诊就诊、住院、死亡,严重危害人类健康和社会经济^[1]。由于流感诊断需要特异的实验室检测且临床上不开展常规检测,因此住院、死亡负担通常采用间接方法进行评价。据 WHO 估计,季节性流感的年度流行导致全球 300 万~500 万人出现严重疾病,25 万~50 万人死亡^[2]。1918 年、1957 年和 1968 年三次大流行分别导致全球约 2 000 万、400 万和 200 万人死亡^[3],2009—2010 年的甲型 H1N1 流感大流行[A(H1N1)pdm]导致 201 200 人死于流感相关呼吸系统疾病^[4]。上述全球范围的流感死亡负担均是基于部分国家的样本数据利用统计学模型进行估计而得。全球和各国流感相关死亡负担研究有助于增加决策者和公众对流感危害的认识。

我国流感监测包括门诊急症病例和住院病例的哨点监测,参与监测的哨点医院数量有限,且缺乏全国范围的门诊急症、住院登记信息系统以及基于全人群的死因登记系统,因此,直接估计全国和各省流感疾病负担存在较大技术难度。本文通过文献回顾,重点对利用部分国家或局部地区的样本数据估计全球和区域范围流感疾病负担的文献进行综述。以下分别针对流感大流行和季节性流行所导致的疾病负担进行阐述。

1. 流感大流行所致疾病负担:目前,国际上利用样本数据估计流感大流行在全球、区域或全国范围死亡负担的方法有国民总收入-纬度模型(gross national income-latitude method, GNI-latitude 模型)、填补法(imputation methods)、自助法(bootstrap method)、乘数法(multiplier method)及系统综述。

(1)GNI-latitude 模型:该模型由 Murray 等^[5]于 2006 年提

出,基于“西班牙流感”大流行期间 27 个国家和地区的超额死亡率预测了类似严重程度的流感大流行于 2004 年发生时在全球和各国所造成的死亡负担。2009—2010 年 A(H1N1)pdm 大流行期间,WHO 全球大流行死亡负担项目组(The Global Pandemic Mortality Burden Project, GLaMOR)^[6]采用相同模型,利用 20 个高、中收入水平国家的 A(H1N1)pdm 相关超额死亡资料,分别估计了 2009 年全球、WHO 各区域及 197 个国家 A(H1N1)pdm 导致的死亡负担。

该模型基于线性回归原理,其前提是因变量和自变量之间存在线性关系,模型的一般形式为 $\log(Y) = \beta_0 + \beta_1 * \log(X_1) + \beta_2 * X_2 + e$,其中因变量 $\log(Y)$ 为经过对数转换的流感超额死亡率,自变量 $\log(X_1)$ 为经过对数转换的国家人均国民总收入,自变量 X_2 为国家纬度绝对值, β_0 为常数项, β_1 和 β_2 均为回归系数(通常采用最小二乘法估计),分别表示自变量人均国民收入和纬度值对因变量流感超额死亡率的作用大小, e 为随机误差项。模型中各参数值确定后,在自变量的取值范围内可得到因变量的估计值。Murray 等^[5]认为人均国民总收入和纬度是影响流感大流行死亡负担的重要因素,因此,对已有部分国家的流感超额死亡率取对数合并为因变量,取对数后的人均国民总收入和纬度为自变量,建立 GNI-latitude 模型后,代入其他国家的人均国民总收入和纬度值,预测流感大流行的超额死亡率。Murray 等^[5]预测严重程度类似于“西班牙流感”的流感大流行于 2004 年发生时在全球、WHO 区域所导致的超额死亡,全球约 6 200 万人死亡,93% 的死亡发生在发展中国家。GLaMOR 也尝试使用此模型估计各国 2009—2010 年 A(H1N1)pdm 的超额死亡率。

通过对 GNI-latitude 模型总体的假设检验判断模型是否有统计学意义,而决定系数 R^2 判断模型的拟合效果, R^2 越大,表明拟合效果越好^[7]。GNI-latitude 模型在流感超额死亡率及其影响因素之间建立因果关系,适用于影响因素数据容易获得且与因变量关系明确的数据。模型原理简单,变量间关系明确,但仅以人均国民总收入和纬度作为解释变量,解释作用不够充分,可能造成结果低估,可纳入更多与流感死亡密切相关的影响因素以提高结果的精确性。

(2)填补法:GLaMOR 研究组在估计 2009—2010 年 A(H1N1)pdm 流行期间各个国家和地区流感超额死亡水平时,除应用 GNI-latitude 模型之外,还将无流感超额死亡率的国家和地区作为缺失数据,分别采用多重填补法和匹配法(matching method)进行估计^[6]。

填补是处理缺失数据常用的一种统计方法,是指借助一定辅助变量信息,为缺失数据确定合理填补值以得到完整的数据集。根据不同数据缺失模式,可采取不同填补方法,但

DOI: 10.3760/cma.j.issn.0254-6450.2015.02.020

作者单位:102206 北京,中国疾病预防控制中心传染病预防控制处 传染病监测预警重点实验室

通信作者:余宏杰, Email: yuhj@chinacdc.cn

理想的填补值应尽量接近缺失数据的原始值,减小数据缺损产生的偏倚。多重填补法由 Rubin^[8]于 1987 年提出,经过不断完善,现已成为处理缺失数据的主要方法。假设数据集由 m 个国家及其 n 个辅助变量组成的 $m \times n$ 矩阵,多重填补法中,利用流感超额死亡率和其辅助变量之间的关系为每个国家缺失的流感超额死亡率构建概率分布模型,并从中随机抽取若干值作为填补值形成填补数据集,一般要求数据缺失模式为随机缺失^[9]。匹配法则借鉴近邻匹配思想,将国家已知的流感超额死亡率及其辅助变量作为原始数据集,将辅助变量值和原始数据集中同一变量值进行比较,差值绝对值最小者对应的流感超额死亡率值作为填补值^[10]。

考虑到社会经济水平、人口统计学、地区差异等是流感死亡负担的重要影响因素^[6,11],进行填补时,GLaMOR 研究组选择 10 项指标(人均国民总收入、人口密度、纬度、WHO 区域、死亡率年龄组分布、医务人员密度、肥胖率、主要感染性疾病、农村人口比例、人口年龄构成比)作为辅助变量,以 20 个已知 A(H1N1)pdm 超额死亡率的高、中收入国家的数据作为原始数据集,其他国家和地区的流感超额死亡率以缺失变量处理^[6]。由于采用多项辅助变量进行比较,实际上每个国家也生成多个填补值(图 1)。同一 WHO 区域内国家的辅助变量间存在相似性,估计得到国家的 A(H1N1)pdm 超额死亡率值之间的差异不仅包括自身特征差异,也包括 WHO 区域间的差异,因此以 WHO 区域和国家为不同层级,采用分层线性回归模型得到 197 个国家、WHO 各区域和全球范围 A(H1N1)pdm 流行的超额死亡水平。GLaMOR 研究组采用多重填补法估计全球 A(H1N1)pdm 超额死亡率为 2.77/10 万,62%~85% 发生在 <65 岁年龄组^[6]。我国的流感超额死亡率为 2.13/10 万,64% 发生在 <65 岁年龄组,与我国直接估计结果类似^[12]。

GLaMOR 通过可靠性系数、外部校准比较两种方法得到结果的精确性。可靠性系数取值范围为 0~1,当取值 >0.8 时认为方法可靠。外部校准是将两种方法估计的结果与部分国家采用其他方法估计的结果进行比较,结果越接近时认为此方法越好。填补法利用流感超额死亡率和辅助变量间

关系,通过构建模型产生合理的填补值,需要大量辅助变量信息。模型精确性较高,但原理复杂计算过程耗时,填补值易受到辅助变量信息的影响。

(3)自助法:该法是以原始数据为基础样本进行重复随机抽样的一类统计方法^[13],基本思想:假定原始样本数据能够较好地反映总体分布,对其进行有放回的重复抽样(如蒙特卡罗抽样)以扩大样本量并模拟总体分布,对获得的样本分布特征进行统计学推断。GLaMOR 研究组假定已知的 20 个高、中等收入国家和地区 2009—2010 年 A(H1N1)pdm 超额死亡率均能代表全球流感超额死亡率的平均水平,将其作为原始样本采用自助法估计 A(H1N1)pdm 超额死亡率的全球水平^[6],结果显示,全球流感超额死亡率的平均水平为 2.0/10 万,地区间无差异。

自助法在原始样本内随机重复抽样,适合代表性好且差异无统计学意义的小样本数据。其计算特性也使得抽样产生的随机样本局限于原始样本范围内,估计结果精度较低。

(4)乘法法:除上述模型外,美国疾病预防控制中心 Dawood 等^[4]在估计 2009—2010 年 A(H1N1)pdm 的死亡负担时使用了乘法法。该方法的核心是乘数应既与流感相关的呼吸系统疾病死亡有关,又能调整死亡的地区差异。Dawood 等^[4]首先利用 12 个高、中、低收入水平国家的 A(H1N1)pdm 确诊病例的发病率与 5 个高收入水平国家的 A(H1N1)pdm 确诊病例的病死率相乘估计 A(H1N1)pdm 的粗死亡率,然后借助 WHO 发布的 2008 年各国下呼吸道感染死亡率和 WHO 成员国死亡分层标准设置乘数,估计各国 A(H1N1)pdm 导致的死亡负担。Dawood 等^[4]估计 A(H1N1)pdm 导致 20 万人死于呼吸系统疾病和 8 万人死于循环系统疾病,其中 13%~20% 发生在 ≥65 岁年龄组,A(H1N1)pdm 相关的全球呼吸系统疾病死亡率为 2.9/10 万,此结果与 GLaMOR 研究组采用多重填补法估计结果类似^[6]。

乘法法借助乘数体现地区差异,适合因数代表性好且乘数意义明确的小样本数据。乘法法原理简单易懂,因数意义明确,但乘数的设置基于与流感死亡负担相关的同一指标,当指标不易获得时方法难以实施。

已知甲型 H1N1 流感超额死亡率的 国家	辅助指标(举例)			甲型 H1N1 流感超额死亡率 (/10 万)
	呼吸系统疾病死亡率 (/10 万)	糖尿病患病率 (%)	肥胖率 (% , BMI > 30)	
1	32	6.6	19.5	20
2	15	9.3	25.5	12
3	8	4.7	4.5	15
4	17	6.9	16.2	30

无甲型 H1N1 流感超额死亡率的 国家	辅助指标			匹配数据			
	呼吸系统疾病死亡率 (/10 万)	糖尿病患病率 (%)	肥胖率 (% , BMI > 30)	甲型 H1N1 流感超额死亡率 (/10 万)	辅助指标	国家	
国家 Y	9	6.9	24.9	匹配 1	15	呼吸系统疾病死亡率	3
				匹配 2	30	糖尿病患病率	4
				匹配 3	12	肥胖率	2

图 1 基于辅助指标进行匹配填补 A(H1N1)pdm 超额死亡率的示意图^[6]

(5)系统综述: Van Kerkhove 等^[14]收集2004年1月至2010年6月19个高、中、低收入水平国家27篇文献的血清学调查研究结果,采用系统综述估计了全球各年龄组和特定年龄组人群在A(H1N1)pdm大流行前期、流行期和后期的感染水平,结果显示,大流行期间全球A(H1N1)pdm感染率为24%,其中5~19岁组最高(46%),≥65岁组最低(11%),年龄组及WHO区域间差异有统计学意义。

系统综述属于二次研究,需要大量符合标准的文献。我国关于流感疾病负担研究的文献数量较少且局限在部分地区,采用系统综述很难具有代表性和一般性。

2. 季节性流感所致疾病负担:利用样本数据估计季节性流感所致疾病负担的估计方法主要有系统综述和多重填补法。

(1)系统综述:Nair 等^[15]在估计2008年季节性流感引起<5岁儿童呼吸系统感染负担时采用系统综述方法,对1995年1月至2010年10月期间的43篇文献结果进行Meta分析,首先估计季节性流感引起的<5岁儿童流感发病率,急性下呼吸道感染(ALRI)的发病率、病死率以及严重ALRI的发病率;然后使用2008年全球儿童人口数,分别估计2008年全球、WHO区域及发达国家和发展中国家的流感新发病例数、流感相关的ALRI病例数和病死数、严重ALRI病例数。季节性流感流行平均每年可在全球0~5岁儿童中导致9000万流感新发病例、2000万ALRI、100万的严重ALRI。其中,2万~10万ALRI死亡,99%的死亡病例发生在发展中国家,发展中国家的病死率是发达国家的17倍。

(2)多重填补法:GLaMOR研究组在利用多重填补法估计A(H1N1)pdm大流行相关超额死亡水平的同时,也估计了2005年1月至2009年4月WHO各区域和国家季节性流感相关超额死亡率^[6]。结果显示,季节性流感每年导致15万~25万的呼吸系统疾病超额死亡,其中81%发生在≥65岁年龄组。

我国人口众多,地理、气候特征复杂,不同区域的经济水平呈多样化,流感流行的季节模式存在区域性差异^[16],精确估计全国和各省流感疾病负担存在一定难度和挑战。目前,国内流感疾病负担研究刚刚起步,多局限在局部地区的样本数据^[17],无法体现全国总体负担及地区差异。

参 考 文 献

[1] Simonsen L. The global impact of influenza on morbidity and mortality[J]. *Vaccine*, 1999, 17 Suppl 1: S3-10.
 [2] WHO. Influenza (Seasonal) [EB/OL]. (2014-05)[2014-09-25]. <http://www.who.int/mediacentre/factsheets/fs211/en/>.
 [3] Hayward AC, Fragaszy EB, Bermingham A, et al. Comparative community burden and severity of seasonal and pandemic influenza: results of the Flu Watch cohort study [J]. *Lancet Respir Med*, 2014, 2(6): 445-454.

[4] Dawood FS, Iuliano AD, Reed C, et al. Estimated global mortality associated with the first 12 months of 2009 pandemic influenza A H1N1 virus circulation: a modelling study [J]. *Lancet Infect Dis*, 2012, 12(9): 687-695.
 [5] Murray CJ, Lopez AD, Chin B, et al. Estimation of potential global pandemic influenza mortality on the basis of vital registry data from the 1918-20 pandemic: a quantitative analysis [J]. *Lancet*, 2007, 368(9554): 2211-2218.
 [6] Simonsen L, Spreuwenberg P, Lustig R, et al. Global mortality estimates for the 2009 influenza pandemic from the GLaMOR Project: a modeling study [J]. *PLoS Med*, 2013, 10(11): e1001558.
 [7] Harrell FE. *Regression Modeling Strategies: with Applications to Linear Models, Logistic Regression, and Survival Analysis* [M]. New York: Springer, 2001: 101-107.
 [8] Rubin DB. *Multiple Imputation for Nonresponse in Surveys* [M]. New York: John Wiley & Sons, 2009: 202-243.
 [9] Schafer JL, Olsen MK. Multiple imputation for multivariate missing-data problems: a data analyst's perspective [J]. *Multivar Behavioral Res*, 1998, 33(4): 545-571.
 [10] Elliott P, Hawthorne G. Imputing missing repeated measures data: how should we proceed? [J]. *Aust N Z J Psychiatry*, 2005, 39(7): 575-582.
 [11] Yang L, Ma S, Chen PY, et al. Influenza associated mortality in the subtropics and tropics: results from three Asian cities [J]. *Vaccine*, 2011, 29(48): 8909-8914.
 [12] Yu H, Feng L, Viboud CG, et al. Regional variation in mortality impact of the 2009 A (H1N1) influenza pandemic in China [J]. *Influenza Other Respir Viruses*, 2013, 7(6): 1350-1360.
 [13] Efron B. Bootstrap methods: another look at the jackknife [J]. *Ann Statist*, 1979, 7(1): 1-26.
 [14] Van Kerkhove MD, Hirve S, Koukounari A, et al. Estimating age-specific cumulative incidence for the 2009 influenza pandemic: a meta-analysis of A (H1N1) pdm09 serological studies from 19 countries [J]. *Influenza Other Respir Viruses*, 2013, 7(5): 872-886.
 [15] Nair H, Brooks WA, Katz M, et al. Global burden of respiratory infections due to seasonal influenza in young children: a systematic review and meta-analysis [J]. *Lancet*, 2011, 378(9807): 1917-1930.
 [16] Yu H, Alonso WJ, Feng L, et al. Characterization of regional influenza seasonality patterns in China and implications for vaccination strategies: spatio-temporal modeling of surveillance data [J]. *PLoS Med*, 2013, 10(11): e1001552.
 [17] Feng L, Shay DK, Jiang Y, et al. Influenza-associated mortality in temperate and subtropical Chinese cities, 2003-2008 [J]. *Bull World Health Organ*, 2012, 90(4): 279-288B.

(收稿日期:2014-09-28)

(本文编辑:万玉立)