

# 甘肃省 2009—2015 年发热呼吸道症候群主要病原的 Bayes 判别分析法分类研究

李学朝 李娟生 孟蕾 白亚娜 于德山 刘小宁 刘新风 蒋小娟 任晓卫  
杨筱婷 申希平 张继巍

730000 兰州大学公共卫生学院流行病与卫生统计学研究所(李学朝、李娟生、白亚娜、刘小宁、任晓卫、申希平、张继巍); 730000 兰州, 甘肃省疾病预防控制中心(孟蕾、于德山、刘新风、蒋小娟、杨筱婷)

通信作者: 李娟生, Email: lijsh16@163.com; 孟蕾, Email: mleicdc@163.com

DOI: 10.3760/cma.j.issn.0254-6450.2017.08.019

**【摘要】** 目的 了解甘肃省发热呼吸道症候群主要病原体, 构建区分主要病原感染病例的 Bayes 判别函数。方法 以 2009—2015 年甘肃省各哨点医院监测的发热呼吸道症候群病例为样本, 描述其病原谱构成, 确定主要病原; 并在样本病例中筛选有意义的临床指标进行 Bayes 判别分析。结果 在发热呼吸道症候群病毒检测中, 流感病毒和鼻病毒检测阳性率较高(13.79% 和 8.63%), 分别占病毒总阳性例数的 54.38% 和 13.73%; 细菌中以肺炎链球菌和流感嗜血杆菌检测阳性率较高(44.41% 和 18.07%), 分别占细菌总阳性例数的 66.21% 和 24.55%。运用筛选后的 11 个临床指标建立判别函数, 其初始验证正确率为 73.1%、交叉验证为 70.6%。结论 甘肃省发热呼吸道症候群主要病原体为流感病毒、鼻病毒、肺炎链球菌及流感嗜血杆菌; Bayes 判别分析在主要病原感染的分类诊断中有较高的正确率, 具有一定的应用价值。

**【关键词】** 发热呼吸道症候群; Bayes 判别分析; 病原体

**基金项目:** 国家科技重大专项(2012ZX10004-208)

**Study on the classification of dominant pathogens related to febrile respiratory syndrome, based on the method of Bayes discriminant analysis** Li Xuechao, Li Juansheng, Meng Lei, Bai Yana, Yu Deshan, Liu Xiaoning, Liu Xinfeng, Jiang Xiaojuan, Ren Xiaowei, Yang Xiaoting, Shen Xiping, Zhang Jiwei

Department of Epidemiology and Health Statistics, School of Public Health, Lanzhou University, Lanzhou 730000, China (Li XC, Li JS, Bai YN, Liu XN, Ren XW, Shen XP, Zhang JW); Gansu Provincial Center for Disease Control and Prevention, Lanzhou 730000, China (Meng L, Yu DS, Liu XF, Jiang XJ, Yang XT)  
Corresponding authors: Li Juansheng, Email: lijsh16@163.com; Meng Lei, Email: mleicdc@163.com

**【Abstract】 Objective** To understand the dominant pathogens of febrile respiratory syndrome (FRS) patients in Gansu province and to establish the Bayes discriminant function in order to identify the patients infected with the dominant pathogens. **Methods** FRS patients were collected in various sentinel hospitals of Gansu province from 2009 to 2015 and the dominant pathogens were determined by describing the composition of pathogenic profile. Significant clinical variables were selected by stepwise discriminant analysis to establish the Bayes discriminant function. **Results** In the detection of pathogens for FRS, both influenza virus and rhinovirus showed higher positive rates than those caused by other viruses (13.79%, 8.63%), that accounting for 54.38%, 13.73% of total viral positive patients. Most frequently detected bacteria would include *Streptococcus pneumoniae*, and haemophilus influenza (44.41%, 18.07%) that accounting for 66.21% and 24.55% among the bacterial positive patients. The original-validated rate of discriminant function, established by 11 clinical variables, was 73.1%, with the cross-validated rate as 70.6%. **Conclusion** Influenza virus, Rhinovirus, *Streptococcus pneumoniae* and Haemophilus influenzae were the dominant pathogens of FRS in Gansu province. Results from the Bayes discriminant analysis showed both higher accuracy in the classification of dominant pathogens, and applicative value for FRS.

**【Key words】** Febrile respiratory syndrome; Bayes discriminate analysis; Pathogeny

**Fund program:** National Science and Technology Major Project (2012ZX10004-208)

发热呼吸道症候群(febrile respiratory syndrome, FRS)是呼吸系统相关器官病变或功能紊乱时出现的一群症状,即由若干疾病或病因引起的一组症状的集合<sup>[1]</sup>,其中呼吸道感染是导致 FRS 的主要原因。FRS 可由多种病毒和细菌感染所致,且近年来呈现发病率升高和病原不断变化的特点<sup>[2-3]</sup>。因此快速明确感染病原体的类型对 FRS 早诊和早治有重要意义。本文拟通过建立 Bayes 判别函数,对 FRS 病原感染类型诊断进行初步分类,以期指导其早期诊断及临床选择用药;并为暴发及流行时病原确定提供帮助。

### 资料与方法

1. 资料来源:国家科技重大专项“传染病监测技术平台”信息管理系统 2009—2015 年甘肃省各哨点医院收集 FRS 为 7 107 例,男性 4 471 例、女性 2 636 例,37.7% 的病例年龄 < 5 岁,13.3% > 65 岁,年龄中位数为 14 岁(四分位数间距为 3~14 岁)。其中 749 例为检测了所有病原病例。采用统一的“呼吸道症候群病例信息调查表”收集病例人口学信息、临床症状和体征、血常规、生化检测、影像学检查等信息;采集病例痰液、鼻咽拭子、鼻咽抽吸物、支气管肺泡灌洗液等标本进行病原检测。病例资料和标本收集获得知情同意。

2. 病例定义:具备急性感染表现(发热、寒战、体温降低、白细胞升高或降低,或白细胞分布异常之一者)、呼吸道临床表现(咳嗽、咳痰、气短、胸痛、呼吸急促以及听诊呼吸音异常之一者)及胸部 X 线片提示肺部炎性改变者作为纳入标准。

3. 病原检测:采集的标本进行流感病毒、呼吸道合胞病毒、人副流感病毒、人腺病毒、人偏肺病毒、人冠状病毒、人博卡病毒、鼻病毒和肺炎链球菌、金黄

色葡萄球菌、肺炎克雷伯菌、铜绿假单胞菌、A 组链球菌、流感嗜血杆菌、军团菌以及肺炎支原体、肺炎衣原体检测。DNA 病毒与肺炎支原体、肺炎衣原体采用聚合酶链式反应(PCR)进行核酸检测,RNA 病毒采用反转录-聚合酶链式反应(RT-PCR);标本经过细菌培养,根据菌落形态和革兰染色镜检结果鉴定除军团菌外的 6 种细菌,细菌培养阴性者采用 PCR 检测上述 6 种细菌及军团菌。

4. 统计学分析:描述 FRS 病例病原谱构成,确定主要病原体;采用逐步判别分析法筛选样本病例有意义的临床指标(流涕、胸痛、乏力、胸腔积液、咽喉痛、淋巴细胞计数百分比、肺部啰音、体温、发热、血小板计数、白细胞计数)构建 Bayes 判别函数。统计软件采用 SPSS 20.0,逐步判别分析变量选入的概率为 0.05,移出的概率为 0.1。

### 结 果

1. FRS 主要病原分布:7 107 例 FRS 中检出病毒阳性病例 1 668 例,其中流感病毒阳性率最高(13.79%),其次为鼻病毒(8.63%),分别占病毒总阳性数的 54.38%、13.73%;953 例细菌检测阳性,肺炎链球菌和流感嗜血杆菌阳性率(44.41%、18.07%)最高,占细菌总阳性数的 66.21%、24.55%(表 1)。

2. 主要检出病原的 Bayes 判别分析:

(1)临床指标的赋值及变量筛选:对流感病毒、鼻病毒、肺炎链球菌、流感嗜血杆菌主要 4 种病原感染病例进行 Bayes 判别分析。为保证判别分析的效果,选取其中 749 例检测了所有病原病例(8 种病毒和 7 种细菌以及肺炎支原体、肺炎衣原体)构建判别函数(749 例检测的主要病原同 7 107 例一致),对表 2 中 4 种主要病原阳性病例剔除缺失数据和混合感染后,得人流感病毒阳性 124 例、鼻病毒阳性 98 例、肺

表 1 2009—2015 年甘肃省 FRS 病例病原检测情况

病 原	检测例数	阳性		构成比 (%)	病 原	检测例数	阳性		构成比 (%)
		例数	率(%)				例数	率(%)	
人流感病毒	6 577	907	13.79	54.38	肺炎链球菌	1 421	631	44.41	66.21
鼻病毒	2 655	229	8.63	13.73	流感嗜血杆菌	1 295	234	18.07	24.55
呼吸道合胞病毒	5 630	159	2.82	9.53	肺炎克雷伯菌	1 921	24	2.47	2.52
人冠状病毒	5 627	124	2.20	7.43	肺炎支原体	970	35	1.82	3.67
人副流感病毒	5 628	119	2.11	7.13	铜绿假单胞菌	1 921	11	1.14	1.15
人腺病毒	5 632	64	1.14	3.84	肺炎衣原体	966	13	0.68	1.36
人偏肺病毒	5 627	46	0.82	2.76	金黄色葡萄球菌	987	3	0.30	0.31
人博卡病毒	2 857	20	0.70	1.20	军团菌	1 211	2	0.17	0.21
合 计	-	1 668	23.47	100.00	合 计	-	953	13.41	100.00

注:未检出 A 组链球菌

炎链球菌阳性 144 例以及流感嗜血杆菌阳性 69 例, 分别用  $Y_1$ 、 $Y_2$ 、 $Y_3$ 、 $Y_4$  代表 4 种阳性病例, 将其临床症状发热 ( $X_1$ )、咳嗽 ( $X_2$ )、流涕 ( $X_3$ )、咽喉痛 ( $X_4$ )、咳痰 ( $X_5$ )、胸痛 ( $X_6$ )、呼吸急促 ( $X_7$ )、呼吸困难 ( $X_8$ )、头痛 ( $X_9$ )、乏力 ( $X_{10}$ )、腹痛 ( $X_{11}$ )、腹泻 (每天稀便  $\geq 3$  次) ( $X_{12}$ )、肺部啰音 ( $X_{13}$ )、肺部有无渗出 ( $X_{14}$ )、胸腔积液 ( $X_{15}$ ) 进行赋值 (是 = 1、否 = 0); 体温 ( $X_{16}$ )、白细胞计数 ( $X_{17}$ )、淋巴细胞计数 ( $X_{18}$ )、中性粒细胞百分比 ( $X_{19}$ )、淋巴细胞百分比 ( $X_{20}$ )、血红蛋白 ( $X_{21}$ )、血小板计数 ( $X_{22}$ ) 均采用实测值。通过逐步判别分析筛选后获得 11 个有统计学意义的临床指标, 其变量筛选步骤及统计学检验见表 3。

(2) Bayes 判别函数的建立及其效果评价: 筛选的 11 个变量建立 Bayes 判别函数为

$$Y_1 = -716.725 - 30.048 X_1 - 5.223 X_3 - 7.172 X_4 - 3.524 X_6 + 11.695 X_{10} + 1.097 X_{13} - 1.868 X_{15} + 38.254 X_{16} - 0.261 X_{17} - 0.023 X_{20} + 0.022 X_{22}$$

$$Y_2 = -721.689 - 31.254 X_1 + 0.321 X_3 - 9.692 X_4 - 3.349 X_6 + 11.563 X_{10} + 1.253 X_{13} - 0.753 X_{15} + 38.386 X_{16} - 0.268 X_{17} - 0.059 X_{20} + 0.024 X_{22}$$

表 2 2009—2015 年甘肃省所有病原检测的 749 例 FRS 病例病原检出情况

病原	阳性		构成比 (%)	病原	阳性		构成比 (%)
	例数	率 (%)			例数	率 (%)	
人流感病毒	163	21.76	36.22	肺炎链球菌	419	55.94	68.35
鼻病毒	119	15.89	26.44	流感嗜血杆菌	157	20.96	25.61
人冠状病毒	68	9.08	15.11	肺炎支原体	12	1.60	1.96
呼吸道合胞病毒	42	5.61	9.33	肺炎克雷伯菌	11	1.47	1.79
人副流感病毒	29	3.87	6.44	肺炎衣原体	8	1.07	1.31
人偏肺病毒	15	2.00	3.33	铜绿假单胞菌	4	0.53	0.65
人腺病毒	8	1.07	1.78	金黄色葡萄球菌	1	0.13	0.16
人博卡病毒	6	0.80	1.33	军团菌	1	0.13	0.16
合计	450	60.08	100.00	合计	613	81.84	100.00

注: 包括数据缺失和混合感染者

表 3 FRS 主要 4 种病原感染病例临床指标选入和移出 Wilks-Lambda 检验

步骤	入选变量	统计量	$df_1$	$df_2$	$df_3$	精确 F 检验			
						统计量	$df_1$	$df_2$	P 值
1	流涕	0.703	1	3	431	60.806	3	431.000	0.000
2	胸痛	0.529	2	3	431	53.740	6	860.000	0.000
3	乏力	0.427	3	3	431	48.560	9	1 044.224	0.000
4	胸腔积液	0.357	4	3	431	44.905	12	1 132.673	0.000
5	咽喉痛	0.319	5	3	431	40.250	15	1 179.160	0.000
6	淋巴细胞百分比	0.301	6	3	431	35.457*	18	1 205.395	0.000
7	肺部啰音	0.281	7	3	431	30.354	21	1 220.330	0.000
8	体温 (°C)	0.272	8	3	431	29.005	24	1 230.330	0.000
9	发热	0.263	9	3	431	26.585	27	1 236.020	0.000
10	血小板计数	0.256	10	3	431	12.395	30	1 239.329	0.000
11	白细胞计数	0.251	11	3	431	7.492	33	1 241.047	0.000

注: \*为近似 F 值

$$Y_3 = -709.847 - 30.529 X_1 - 3.844 X_3 - 9.528 X_4 - 0.170 X_6 + 8.479 X_{10} + 2.691 X_{13} - 0.404 X_{15} + 38.059 X_{16} - 0.165 X_{17} - 0.065 X_{20} + 0.022 X_{22}$$

$$Y_4 = -736.438 - 31.423 X_1 - 4.142 X_3 - 8.623 X_4 - 1.962 X_6 + 10.488 X_{10} + 2.544 X_{13} - 1.200 X_{15} + 38.746 X_{16} - 0.197 X_{17} - 0.055 X_{20} + 0.0226 X_{22}$$

应用已建立的判别函数对所有病原检测病例的 4 种主要病原进行判别效果评价。初始验证的总体正确率为  $(101 + 64 + 107 + 46) / 435 = 73.1\%$ , 4 种主要病原正确率分别为 81.4%、65.3%、74.3%、66.7%; 交叉验证的总体正确率为  $(98 + 62 + 103 + 44) / 435 = 70.6\%$ , 4 种主要病原正确率分别为 79.0%、63.3%、71.5%、63.8% (表 4)。

### 讨论

2009—2015 年甘肃省 FRS 主要感染病原为流感病毒、鼻病毒、肺炎链球菌以及流感嗜血杆菌。其中病毒阳性检出率为 23.75%, 细菌阳性检出率为 13.41%。病毒检测中阳性率最高的是流感病毒 (13.79%), 与全国范围内呼吸道感染病毒检测情况一致<sup>[4]</sup>; 其次是鼻病毒 (8.63%), 表明鼻病毒是引起呼吸道感染的重要病原之一, 成年人约 50% 的普通感冒是由其感染所致<sup>[5]</sup>。细菌检测中主要病原为肺炎链球菌和流感嗜血杆菌, 阳性率为 44.41%、18.07%, 其他细菌阳性率均不到 3%, 可能与当前抗生素的普遍使用有关。

分析显示, FRS 病例中 4 种主要病原的 Bayes 判别分析正确率较高, 在 FRS 病原分类诊断中有一定的应用价值。本文 Bayes 判别分析初始验证正确率为 73.10%, 交叉验证正确率为 70.60%。Bayes 判别是以概率为判别准则, 根据已建立的判别函数计算每一类的后验概率, 将判别对象判为后验概率最大的一类<sup>[6]</sup>。该判别分析方法不仅考虑每个总体出现的先验概率, 同时也考虑了因误判所形成的损失, 判别效能更优且灵敏性和稳定性更高<sup>[7-8]</sup>。甘肃省 FRS 主要病原感染病例的 Bayes 判别分析, 其初始验证和交叉验证正确率均  $> 70\%$ , 判别效果较好。在 4 种主要病原判别分析中, 鼻病毒 (初始验证正确率为 65.31%, 交叉验证正确率为 63.30%) 和流感嗜血杆菌 (初始验证正确

表4 FRS主要4种感染病原Bayes判别函数分析

感染病原分类	预测分组				合计
	Y <sub>1</sub>	Y <sub>2</sub>	Y <sub>3</sub>	Y <sub>4</sub>	
初始验证					
流感病毒	101(81.4)	7(5.6)	6(4.8)	10(8.2)	124(100.0)
鼻病毒	17(17.4)	64(65.3)	5(5.1)	11(11.2)	98(100.0)
肺炎链球菌	11(7.6)	8(5.6)	107(74.3)	18(12.5)	144(100.0)
流感嗜血杆菌	6(8.7)	6(8.7)	11(15.9)	46(66.7)	69(100.0)
交叉验证					
流感病毒	98(79.0)	10(8.1)	6(4.8)	10(8.1)	124(100.0)
鼻病毒	18(18.4)	62(63.3)	5(5.1)	13(13.2)	98(100.0)
肺炎链球菌	11(7.6)	8(5.6)	103(71.5)	22(15.3)	144(100.0)
流感嗜血杆菌	6(8.7)	6(8.7)	13(18.8)	44(63.8)	69(100.0)

注:括号外数据为例数,括号内数据为构成比(%)

率为66.67%,交叉验证正确率为63.80%)判别正确率均略低于其余2种病原,原因可能与Bayes判别分析要求样本量较大有关<sup>[9]</sup>,而鼻病毒和流感嗜血杆菌样本病例相对较少,使得判别正确率不够稳定。在本文中还应应用逐步判别法探讨了各临床指标在4种主要病原感染分类诊断中的价值,通过逐步筛选获得11个临床指标有统计学意义;它们对判别函数贡献大小依次为流涕、胸痛、乏力、胸腔积液、咽喉痛、淋巴细胞百分比、肺部啰音、体温、发热、血小板计数、白细胞计数。值得注意的是,有研究表明血小板计数对鉴别呼吸道感染病原体没有特异性诊断价值<sup>[10]</sup>,然而逐步判别分析显示血小板计数为有统计学意义的变量,对判别函数有贡献,说明血小板计数在鉴别呼吸道感染病原时有一定的价值。

本文存在局限性。如将Bayes判别分析法只应用在4种主要病原中,而未对所有病原进行判别分类,一方面是由于判别分析在分类越多的情况下其判别效果越难保证,另一方面部分病原检出阳性例数少且临床资料缺失较多,使得判别分析样本量难以保证;此外,本文也未涉及判别分析中是否存在病原亚型的分类应用价值,还有待进一步研究。

总之,2009—2015年甘肃省FRS病原检测中流感病毒、鼻病毒、肺炎链球菌以及流感嗜血杆菌4种主要病原阳性率均较高,应进行长期监测。Bayes逐步判别分析能够很好的分析这4种主要病原阳性病例的临床指标,筛选出对判别函数有意义的变量,并建立较为稳定的判别函数,对呼吸道感染病原的早期诊断有一定的积极作用。

利益冲突 无

参 考 文 献

- [1] Wallace DJ, Arquilla B, Heffernan R, et al. A test of syndromic surveillance using a severe acute respiratory syndrome model[J]. Am J Emerg Med, 2009, 27(4): 419-423. DOI: 10.1016/j.ajem.2008.03.020.
- [2] Feikin DR, Njenga MK, Bigogo G, et al. Etiology and Incidence of viral and bacterial acute respiratory illness among older children and adults in rural western Kenya, 2007-2010[J]. PLoS One, 2012, 7(8): e43656. DOI: 10.1371/journal.pone.0043656.
- [3] Nair H, Nokes DJ, Gessner BD, et al. Global burden of acute lower respiratory infections due to respiratory syncytial virus in young children: a systematic review and Meta-analysis [J]. Lancet, 2010, 375 (9725) : 1545-1555. DOI: 10.1016/S0140-6736(10)60206-1.
- [4] Feng LZ, Li ZJ, Zhao SW, et al. Viral etiologies of hospitalized acute lower respiratory infection patients in China, 2009-2013 [J]. PLoS One, 2014, 9(6) : e99419. DOI: 10.1371/journal.pone.0099419.
- [5] Greenberg SB. Update on rhinovirus and coronavirus infections [J]. Semin Respir Crit Care Med, 2011, 32(4) : 433-446. DOI: 10.1055/s-0031-1283283.
- [6] 孙振球, 徐勇勇. 医学统计学[M]. 4版. 北京: 人民卫生出版社, 2014: 307-308.  
Sun ZQ, Xu YY. Medical statistics[M]. 4<sup>th</sup> ed. Beijing: People's Medical Publishing House, 2014: 307-308.
- [7] 李晓毅. Bayes判别分析及其在疾病诊断中的应用[J]. 中国卫生统计, 2004, 21(6) : 356-357, 374. DOI: 10.3969/j.issn.1002-3674.2004.06.014.  
Li XY. Application of Bayes discrimination analysis in disease diagnosis [J]. Chin J Health Stat, 2004, 21(6) : 356-357, 374. DOI: 10.3969/j.issn.1002-3674.2004.06.014.
- [8] 陶红兵, 叶建君, 苗卫军, 等. 活动性肺结核密切接触者发病判别模型研究[J]. 中华流行病学杂志, 2009, 30(7) : 676-678. DOI: 10.3760/cma.j.issn.0254-6450.2009.07.007.  
Tao HB, Ye JJ, Miao WJ, et al. The incidence discriminant model for close contacts of active tuberculosis patients [J]. Chin J Epidemiol, 2009, 30(7) : 676-678. DOI: 10.3760/cma.j.issn.0254-6450.2009.07.007.
- [9] Hora SC. Sample size determination in Bayesian discriminant analysis[J]. J Am Stat Associat, 1978, 73(363) : 569-572. DOI: 10.1080/01621459.1978.10480056.
- [10] ten Oever J, Tromp M, Bleeker-Rovers CP, et al. Combination of biomarkers for the discrimination between bacterial and viral lower respiratory tract infections [J]. J Infect, 2012, 65(6) : 490-495. DOI: 10.1016/j.jinf.2012.08.004.

(收稿日期: 2016-12-21)

(本文编辑: 张林东)