

利用两水平 logistic 模型探讨麻疹监测系统病例分类

黎翰东 敖睿 彭琳 李迪 张菊英

【摘要】 目的 建立判别麻疹临床症状的 logistic 回归模型, 比较麻疹与其他发热出疹性疾病临床症状的差异。方法 收集 2011 年 3 月至 2012 年 2 月四川省 4 个县级以上医院报告的所有麻疹、风疹、猩红热、幼儿急疹等病例, 以实验室确诊麻疹为反应变量, 各临床症状为自变量建立 logistic 回归方程, 并绘制 ROC 曲线, 选择最佳截断点。结果 共收集 551 名病例, 麻疹和风疹临床诊断与实验室诊断一致性均较差, $Kappa$ 值=0.349; 根据两水平 logistic 回归分析, 与麻疹确诊病例有关的症状有咳嗽 ($OR=5.75$)、结膜炎 ($OR=3.00$)、柯氏斑 ($OR=7.52$)、淋巴结肿大 ($OR=0.07$)、热退疹出 ($OR=0.07$); 回归方程 ROC 曲线的曲线下面积为 0.97, 最佳截断点为 0.249。结论 通过临床症状建立了预测效果较好的 logistic 回归模型, 但应开展更大范围的调查和实验室检测, 明确与麻疹相关的主要发热出疹性疾病类型及构成。

【关键词】 麻疹; 监测; 两水平 logistic 回归; ROC 曲线

Case classification in measles surveillance system under the Two-level Logistic Model Li Handong¹, Ao Rui², Peng Lin¹, Li Di¹, Zhang Juying¹. 1 Department of Epidemiology and Biostatistics, West School of Public Health, Sichuan University, Chengdu 610041, China; 2 Sichuan Provincial Center for Disease Control and Prevention

Corresponding author: Zhang Juying, Email: juying109@163.com

【Abstract】 Objective To study the prevalence rates of Rash and Febrile Illnesses (RFIs) including measles, rubella, scarlet fever, exanthema subitum and the differences among measles and other RFIs to tentatively formulate the logistic regression model through clinical manifestation. **Methods** All the suspected cases of measles, rubella, scarlet fever, exanthema subitum reported by the county/prefecture level hospitals at four counties were collected during March 2011 to February 2012. When setting laboratory confirmed measles as dependent variable and existed symptoms as independent variable, a logistic regression model was formulated and optimal operational point (OOP) chosen, according to the ROC curve. **Results** A total number of 551 cases were collected but the consistency of measles diagnosis between clinical and laboratory was not satisfied, with $Kappa$ value=0.349, same to the diagnosis of rubella. As for the result from the two-level logistic regression model, symptoms that related to the confirmation of measles would include cough ($OR=5.75$), conjunctivitis ($OR=3.00$), Koplik spot ($OR=7.52$), lymphadenectasis ($OR=0.07$), rash after fever ($OR=0.07$). The area under ROC curve was 0.97 and the optimal operational point was 0.249. **Conclusion** A logistic regression model was formulated using the clinical symptoms which was resulted in better performance on prediction. As the sample size of this survey was small, the expansion on the scale of investigation and laboratory testings were needed before the types and components of measles-related RFIs be clarified.

【Key words】 Measles; Surveillance; Two-level logistic regression; ROC curve

麻疹病毒的传播力极强, 在麻疹疫苗使用前, 几乎每个儿童均难以幸免感染^[1]。2005 年我国政府向世界卫生组织西太平洋区承诺于 2012 年实现消除

麻疹^[2], 而高质量的麻疹监测系统成为检验是否达到消除麻疹标准的重要依据^[3-5]。根据我国目前麻疹监测方案要求, 县级疾病预防控制中心 (CDC) 在经过流行病学调查和实验室检测后, 需对每例麻疹疑似病例进行最终分类, 即分类为麻疹病例和排除麻疹病例。但在四川省麻疹监测中, 散发麻疹疑似病例实验室诊断的比例较低, 2010 年散发麻疹疑似病例血清采集率仅为 80.52%。而且如果仅在病例

DOI: 10.3760/cma.j.issn.0254-6450.2014.01.014

作者单位: 610041 成都, 四川大学华西公共卫生学院卫生统计学教研室 (黎翰东、彭琳、李迪、张菊英); 四川省疾病预防控制中心 (敖睿)

通信作者: 张菊英, Email: juying109@163.com

出疹后3 d内采集一份血清标本,又未能采集第二份血清,由于采血过早,约有30%的麻疹病例血清IgM抗体不足以检测到^[6],这样检测的结果即使为阴性,也不能完全排除麻疹。2010年四川省在已采集血清的病例中,该类病例占36.42%,故全省约有48.81%的散发麻疹疑似病例无实验室诊断依据,导致县级CDC对病例分类缺乏依据。为了对缺乏实验室诊断的病例进行分类,本研究拟采用两水平logistic模型探讨麻疹疑似病例最终分类。

资料与方法

1. 资料来源:2011年3月至2012年2月以四川省攀枝花市东区和仁和区、泸州市合江县和江阳区为调查地区,在县级及以上医院选取所有临床诊断为麻疹、风疹、猩红热、幼儿急疹的病例,采用本研究的个案调查表收集病例的基本情况、临床表现等信息,并采集血清和咽拭子标本。

2. 标本检测:由各医院采集上述所有病例的血清和咽拭子;攀枝花市和泸州市CDC麻疹网络实验室负责采用ELISA检测血清中麻疹和风疹IgM(检测试剂由珠海海泰生物制药有限公司提供);四川省CDC麻疹实验室负责采用RT-PCR方法检测麻疹和风疹病毒核酸(PCR核酸引物由上海雅培制药有限公司提供)。血清IgM检测和咽拭子PCR检测中任一项麻疹或风疹阳性,则判定为麻疹或风疹确诊病例;两者均阴性,则判定为麻疹或风疹排除病例。

3. 统计学分析:比较临床诊断与实验室检查结果,运用单因素分析方法筛选变量,建立两水平logistic回归模型以预测确诊麻疹的概率,并绘制ROC曲线,确定最佳截断点,两水平logistic回归采用MLwiN 2.10软件,采用SigmaPlot 12.0软件绘制ROC曲线。

结 果

1. 病例概况:共收集到551例发热、出疹性病例,其中男性329例(59.69%),女性222例(40.31%),男女性别比为1.48;<7岁病例比例较高(69.69%)。将年龄划分为0~岁组(代表麻疹疫苗预防接种完成前的儿童病例)、2~岁组(代表麻疹疫苗预防接种完成后的儿童病例)和15~岁组(代表成人病例)。经 χ^2 检验,各年龄组麻疹确诊比例的差异有统计学意义,其中15~岁组麻疹确诊比例较高(表1)。

2. 临床诊断分类:调查病例临床诊断分类为麻疹、风疹、猩红热、幼儿急疹,而实验室检测被分类为

麻疹确诊病例、风疹确诊病例及其他疾病。因实验室仅完成了麻疹和风疹的检测,未开展猩红热和幼儿急疹检测,故本研究将其他疾病列为排除麻疹和风疹的其他疾病。根据列联表的关联性分析,可认为临床和实验室两种诊断方法之间有关联,Pearson列联系数 $r=0.622$ 。

表1 麻疹确诊病例的年龄分布

年龄组(岁)	报告病例	麻疹确诊病例	确诊率(%)	χ^2 值	P值
0~	186	87	46.77	118.390	<0.001
2~	268	27	10.07		
15~	97	60	61.86		
合计	551	174	31.58		

为比较临床和实验室两种方法诊断麻疹的一致性,将临床诊断中的风疹、幼儿急疹、猩红热合并为“临床排除麻疹”,同时将实验室确诊的风疹和其他合并为“实验室排除麻疹”,再对麻疹临床诊断与实验室诊断的一致性进行分析,Kappa值=0.349,一致性较差。而且临床诊断麻疹的灵敏度仅为59.20%,特异度仅为76.66%。两种诊断比较见表2。

表2 调查病例临床诊断和实验室诊断分类比较

临床诊断	实验室确诊病例			合计	χ^2 值	P值
	麻疹	风疹	其他疾病			
麻疹	103	54	34	191	348.294	<0.001
风疹	57	17	4	78		
幼儿急疹	11	4	79	94		
猩红热	3	5	180	188		
合计	174	80	297	551		

3. 建立麻疹临床症状判别的logistic回归模型:

(1)单因素分析:设定 $\alpha=0.1$ 。单因素分析结果显示,在麻疹和排除病例中除出疹外,发热、咳嗽、卡他性鼻炎、结膜炎、柯氏斑、淋巴结肿大、关节痛、悬雍垂溃疡、杨梅舌、口周苍白圈、热退疹出症状比例的差异有统计学意义(表3)。

表3 麻疹临床症状比例(%)的单因素分析

临床症状	麻疹(n=174)	排除麻疹(n=377)	OR值(90%CI)
发热	98.28	94.69	3.19(1.20~8.45)
出疹	100.00	100.00	-
咳嗽	88.51	35.54	13.96(9.61~20.29)
卡他性鼻炎	65.52	30.77	4.28(3.14~5.84)
结膜炎	72.99	15.65	14.56(3.78~5.50)
柯氏斑	41.38	4.24	15.93(10.57~24.02)
淋巴结肿大	2.30	11.67	0.18(0.08~0.39)
关节痛	5.75	2.39	2.49(1.18~5.26)
悬雍垂溃疡	3.45	14.59	0.21(0.11~0.41)
杨梅舌	2.30	34.22	0.05(0.03~0.09)
口周苍白圈	1.72	22.55	0.06(0.03~0.13)
热退疹出	2.30	25.46	0.07(0.04~0.14)

(2)logistic 回归模型:根据对麻疹确诊和排除病例中临床症状差异的分析结果,以是否确诊麻疹为反应变量(0=排除,1=确诊),各临床症状(0=无,1=有)为自变量,拟合 logistic 回归模型。其中,因自变量“出疹”在麻疹确诊和排除病例中比例的差异无统计学意义,故未将其纳入回归方程。因不同年龄段确诊麻疹或排除麻疹病例的临床症状可能有一定差异,成年人或免疫力低下人群可能出现不典型麻疹^[7],且不同年龄组麻疹阳性率不同^[8],所以在分析麻疹和非麻疹病例的临床症状差异时,年龄可能是混杂因素,故将年龄与各临床症状一同纳入方程。另外,本研究数据来自 16 所医院,各医院医生专业水平(诊断疾病和记录临床症状)、病例来源亦不尽相同,传染病专科医院收治的病例构成与其他普通综合医院也有较大差异。在进行 logistic 回归分析时,有必要将“医院”这一因素作为第二个水平纳入分析,建立两水平 logistic 方差成分模型,将由医院不同所导致的差异在回归方程中加以区分。

首先建立零模型。本文以“医院”作为高水平即水平 2,每名病例作为低水平即水平 1,采用 2 阶 PQL 算法拟合两水平不含任何解释变量的零模型,固定尺度参数为 1。结果显示不同医院之间的差异有统计学意义,故有必要建立两水平模型(表 4)。

表 4 麻疹临床症状判别 logistic 回归零模型

参数	β	s_i	χ^2 值	P 值
固定部分(截距) β_0	-1.115	0.438	6.479	0.011
随机部分 水平 2 σ_0^2	8.619	2.485	12.028	<0.001
水平 1 σ_i^2	1.000	0.000	-	-

将上述变量纳入两水平 logistic 回归分析后,“发热”、“卡他性鼻炎”、“悬雍垂溃疡”、“杨梅舌”、“口周苍白圈”未被选入回归方程,故使用其他变量建立 logistic 回归方程,结果与麻疹确诊病例有关的症状有咳嗽、结膜炎、柯氏斑、淋巴结肿大、热退疹出(表 5)。

根据 logistic 回归分析结果建立模型

$$\text{logit}\left(\frac{P}{1-P}\right) = \beta_0 - 2.411X_1 - 0.753X_2 + 1.749X_3 + 1.097X_4 + 2.017X_5 - 2.660X_6 - 2.697X_7$$

$$\beta_0 = -1.054 + \mu_0$$

$$\mu_0 \sim N(0, 4.825)$$

式中 P 为确诊麻疹的预测概率, X_1 为 2~ 岁年龄组, X_2 为 15~ 岁年龄组, X_3 为咳嗽, X_4 为结膜炎, X_5 为柯氏斑, X_6 为淋巴结肿大, X_7 为热退疹出。

4. 麻疹临床症状判别的 ROC 曲线分析:ROC 分析是目前公认的衡量诊断信息和诊断决策质量最佳

表 5 麻疹临床症状判别两水平 logistic 回归分析

参数	估计值	标准误	P 值	OR 值(95%CI)
固定部分				
截距	-1.054	0.586		0.070
年龄(岁)				
0~(对照)				
2~	-2.411	0.538	<0.001	0.09(0.03~0.26)
15~	-0.753	0.499	0.130	0.47(0.29~0.78)
临床症状				
咳嗽	1.749	0.488	<0.001	5.75(3.53~9.37)
结膜炎	1.097	0.451	0.010	3.00(1.91~4.70)
柯氏斑	2.017	0.567	<0.001	7.52(4.26~13.25)
淋巴结肿大	-2.660	0.895	<0.001	0.07(0.03~0.17)
热退疹出	-2.697	0.914	<0.001	0.07(0.03~0.17)
随机部分				
水平 2 σ_0^2	4.825	1.668	<0.001	
水平 1 尺度参数	1.000	0.000	-	

方法^[9]。根据两水平 logistic 回归模型所得的预测概率绘制 ROC 曲线,即 ROC 曲线下面积为 0.97, 95%CI:0.951~0.984。选取约登指数最大的截断点为最佳截断点,即预测概率为 0.249 时,灵敏度为 94.25%,特异度为 91.25%,约登指数为 0.855(图 1;横线以上为预测阳性病例)。

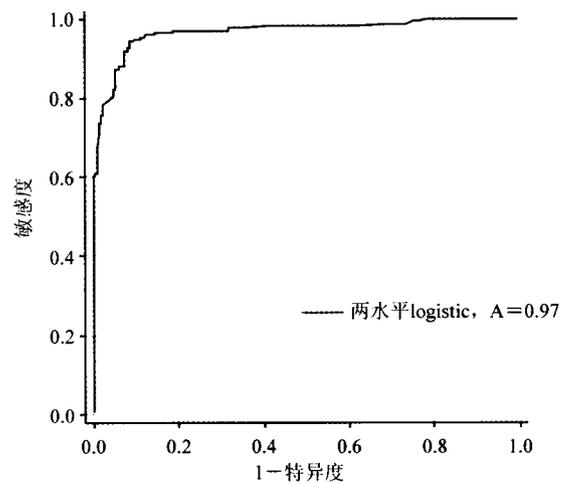


图 1 麻疹临床症状判别两水平 logistic 回归的 ROC 曲线

讨 论

四川省每年麻疹监测中,因未采集到血清或出疹后 3 d 内血清结果为阴性等原因,有大量麻疹疑似病例难以确诊。如以疫情预警为目的,可将所有不能确诊的疑似麻疹病例归类为麻疹临床诊断病例,这样提高了监测系统的敏感性,避免漏掉任何可能的麻疹病例;但如果使用该数据分析麻疹的流行特征,一些非麻疹病例可能混杂在麻疹病例中,从而误导分析结果,如幼儿急疹主要发病年龄为 <2 岁^[10],这部分病例如混杂在临床诊断的麻疹病例中可能增加 <2 岁组病例的年龄构成。本研究建立的 logistic

模型可对未确诊麻疹疑似病例分类,如发热病例出现咳嗽、结膜炎和柯氏斑而确诊为麻疹的可能性分别是无相应症状的 5.75、3.00、7.52 倍,而确诊病例中较少出现“淋巴结肿大”和“热退疹出”,仅为排除麻疹比例的 0.07 倍。

本研究建立的 logistic 回归模型 ROC 曲线下面积达到 97%,且在最佳截点处的灵敏度和特异度均远大于临床判断的灵敏度和特异度,表明具有较好的预测效果。实际运用时可将病例临床症状的取值代入公式,计算出患麻疹的预测概率,若所得的预测概率大于或等于最佳截点 0.249,则判定为麻疹;若所得的预测概率小于最佳截点则判定为排除麻疹。同其他诊断一样,阳性预测值会受到先验概率(即麻疹在发热出疹性疾病中构成比)的影响,如果其构成比下降,阳性预测值也会下降^[11]。

本研究建立的判别模型存在缺陷。如调查范围较小,样本不足,可能代表性相对较差;此外 logistic 回归的对照(即麻疹排除病例)是由多种疾病组成,如这些疾病的构成比例发生较大的变化,回归方程的判别效果可能受到影响,还应建立无序多分类 logistic 回归模型,将麻疹作为多分类比较的对照,计算回归方程,得出麻疹与主要鉴别疾病相对应的鉴别模型。

参 考 文 献

[1] Wu T. Epidemiological research of measles in vaccine era[J]. Zhejiang J Prev Med, 2000, 12(2): 1-3. (in Chinese)
吴霆. 疫苗时代麻疹流行病学研究[J]. 浙江预防医学, 2000, 12(2): 1-3.

[2] World Health Organization. Resolution (WPR/RC56.R8): Measles

elimination, hepatitis B control, and poliomyelitis eradication [M]. Manila, Philippines: World Health Organization, 2005.

[3] Su WN, Zhang YH. Discussion the strategy of controlling and eliminating the measles [J]. Chin J Vacc Immun, 1996, 2(2): 1-3. (in Chinese)
苏万年, 章以浩. 控制和消除麻疹策略研讨[J]. 中国计划免疫, 1996, 2(2): 1-3.

[4] Ren XY, Liu ML. Difficulties and countermeasures of eliminate the measles [J]. J Dis Monitor Control, 2010, 4(11): 641-642. (in Chinese)
任先云, 刘梦来. 消除麻疹面临的困难和应对措施[J]. 疾病监测与控制, 2010, 4(11): 641-642.

[5] World Health Organization. Progress towards elimination of measles in the Americas [J]. WER, 1998, 73: 81-88.

[6] Bhatt PN, Brandt CD, Weiss R, et al. Viral infections of monkeys in their natural habitat in southern India. II. Serological evidence of viral infection [J]. Am J Trop Med Hyg, 1966, 15(4): 561.

[7] Zhang HM, Yu SY, Wang AD, et al. Clinical analysis and serologic diagnosis of atypical measles [J]. Chin J Infect Dis, 1986, 4(2): 120, 87. (in Chinese)
张惠民, 于世芸, 王安定, 等. 非典型麻疹临床分析与血清学诊断[J]. 中华传染病杂志, 1986, 4(2): 120, 87.

[8] Ma C, Hao LX, Ma J, et al. Measles epidemiological characteristics and progress of measles elimination in China [J]. Chin J Vacc Immun, 2011, 17(3): 242-248. (in Chinese)
马超, 郝利新, 马静, 等. 中国 2010 年麻疹流行病学特征与消除麻疹进展[J]. 中国疫苗和免疫, 2011, 17(3): 242-248.

[9] Sun ZQ. Medical statistics [M]. Beijing: People's Medical Publishing House, 2002: 573-574. (in Chinese)
孙振球. 医学统计学 [M]. 北京: 人民卫生出版社, 2002: 573-574.

[10] Zhu FT, Hu YM, Jiang ZF. Textbook of pediatrics [M]. Beijing: People's Medical Publishing House, 2002: 728-740, 935. (in Chinese)
诸福棠, 胡亚美, 江载芳. 实用儿科学 [M]. 北京: 人民卫生出版社, 2002: 728-740, 935.

[11] Katz SL, Hutchins SS, Papania MJ, et al. Evaluation of the measles clinical case definition [J]. J Infect Dis, 2004, 189 Suppl 1: S153-159.

(收稿日期: 2013-07-08)

(本文编辑: 张林东)

中华流行病学杂志第六届编辑委员会通讯编委名单

- | | | |
|---------------------|-------------------|-----------------------|
| 陈 曦(湖南省疾病预防控制中心) | 窦丰满(成都市疾病预防控制中心) | 高 婷(北京市疾病预防控制中心) |
| 姜宝法(山东大学公共卫生学院) | 李 杰(北京大学医学部) | 李十月(武汉大学公共卫生学院) |
| 李秀央(浙江大学医学院公共卫生学院) | 廖苏苏(中国医学科学院基础医学院) | 林 玫(广西壮族自治区疾病预防控制中心) |
| 林 鹏(广东省疾病预防控制中心) | 刘爱忠(中南大学公共卫生学院) | 刘 刚(四川省疾病预防控制中心) |
| 刘 静(北京安贞医院) | 刘 莉(四川省疾病预防控制中心) | 刘 玮(军事医学科学院微生物流行病研究所) |
| 鲁凤民(北京大学医学部) | 欧剑鸣(福建省疾病预防控制中心) | 彭晓昱(北京市疾病预防控制中心) |
| 邱洪斌(佳木斯大学) | 赛晓勇(解放军总医院) | 苏 虹(安徽医科大学公共卫生学院) |
| 汤 哲(首都医科大学附属宣武医院) | 田庆宝(河北医科大学公共卫生学院) | 王 蓓(东南大学公共卫生学院) |
| 王素萍(山西医科大学公共卫生学院) | 王志萍(山东大学公共卫生学院) | 谢 娟(天津医科大学公共卫生学院) |
| 徐爱强(山东省疾病预防控制中心) | 徐慧芳(广州市疾病预防控制中心) | 严卫丽(新疆医科大学公共卫生学院) |
| 阎丽静(中国乔治中心) | 杨春霞(四川大学华西公共卫生学院) | 余运贤(浙江大学医学院公共卫生学院) |
| 曾哲淳(北京安贞医院) | 张 波(宁夏回族自治区卫生厅) | 张宏伟(第二军医大学) |
| 张茂俊(中国疾病预防控制中心传染病所) | 张卫东(郑州大学公共卫生学院) | 赵亚双(哈尔滨医科大学公共卫生学院) |
| 朱 谦(河南省疾病预防控制中心) | 祖荣强(江苏省疾病预防控制中心) | |