

# 活动性肺结核密切接触者发病判别模型研究

陶红兵 叶建君 苗卫军 侯双翼 熊光练 余瑶 郭石林 陈璞

**【摘要】** 目的 建立肺结核发病判别模型,为肺结核密切接触者早期的诊断和发现提供科学依据。**方法** 使用 logistic 逐步回归分析筛选关键影响因素,采用逐步 Bayes 判别法对肺结核密切接触者发病的关键因素进行筛选并建立判别模型。**结果** 密切接触者未发病判别函数为: $Y_1 = -39.831$  (常数) + 1.927  $X_1$  (患者痰涂片阳性次数) + 3.528  $X_2$  (接触者文化程度) + 0.309  $X_3$  (接触时间) + 5.893  $X_4$  (咳嗽时回避) + 2.140  $X_5$  (居室通风) + 8.706  $X_6$  (接触者咳嗽) + 30.970  $X_7$  (接触者发热); 密切接触者发病判别函数为: $Y_2 = -57.875$  (常数) + 2.343  $X_1$  (患者痰涂片阳性次数) + 3.965  $X_2$  (接触者文化程度) + 0.361  $X_3$  (接触时间) + 6.296  $X_4$  (咳嗽时回避) + 1.348  $X_5$  (居室通风) + 12.984  $X_6$  (接触者咳嗽) + 36.555  $X_7$  (接触者发热)。**结论** 模型的建立有助于肺结核密切接触的早期诊断分析,并进行早期的干预和治疗。

**【关键词】** 肺结核; 接触者; Bayes 判别模型

**The incidence discriminant model for close contacts of active tuberculosis patients** TAO Hong-bing\*, YE Jian-jun, MIAO Wei-jun, HOU Shuang-yi, XIONG Guang-lian, YU Yao, GUO Shi-lin, CHEN Pu. \*School of Medical and Health Management, Tongji Medical College of Huazhong University of Science & Technology, Wuhan 430030, China

**【Abstract】** **Objective** To establish a discriminant model and to provide a relatively accurate scientific basis for the early diagnosis of tuberculosis (TB) and detection of the close contacts. **Methods** Through logistic regression analysis, key factors were selected according to Bayes theory and key factors of TB incidence of the close contacts were screened as well as a discriminant model was established. **Results** The non-TB incidence discriminant function of the close contacts was described as:  $Y_1 = -39.831$  (constant) + 1.927  $X_1$  (sputum-frequency) + 3.528  $X_2$  (education) + 0.309  $X_3$  (contact time) + 5.893  $X_4$  (evade) + 2.140  $X_5$  (ventilation) + 8.706  $X_6$  (cough) + 30.970  $X_7$  (fever). The discriminant function of non-TB incidence of the close contacts was as:  $Y_2 = -57.875$  (constant) + 2.343  $X_1$  (sputum-frequency) + 3.965  $X_2$  (education) + 0.361  $X_3$  (contact time) + 6.296  $X_4$  (evade) + 1.348  $X_5$  (ventilation) + 12.984  $X_6$  (cough) + 36.555  $X_7$  (fever). **Conclusion** The discriminant model might be used to contribute to the early diagnosis, early intervention and timely treatment on those close contacts of tuberculosis cases.

**【Key words】** Tuberculosis; Contacts; Bayes discriminant model

对山东、湖北、广东省 983 例结核病例 2637 名密切接触者进行结核相关检查和密切接触者调查,通过 logistic 逐步回归分析筛选密切接触者发病的相关影响因素,利用筛选后的关键因素建立判别模型,为感染肺结核早期诊断和发现提供科学依据。

## 对象与方法

1. 对象:采用整群随机抽样,在山东、湖北、广东省抽取经济相对发达和直接督导短程化疗(DOTS)

项目开展相对较好的县市 1~3 个。根据三省结核病流行现状和流行特点,每个调查点依据 2005 年 9 月至 2007 年 9 月活动性肺结核患病总例数 50% 的比例随机抽取一定数量的活动性肺结核病例及其密切接触者作为研究对象。密切接触者是指与痰涂片阳性的活动性肺结核患者传染期内累计接触 3 个月以上的患者家人、邻居或朋友。

## 2. 方法:

(1) 收集资料;查找结核病患者病历资料,收集一般情况和发病情况(包括病变部位、症状、痰涂片培养结果、X 线检查和治疗情况等),确定活动性肺结核患者的传染期。收集密切接触者与活动性肺结核患者接触方式、接触半径、接触频率以及接触者年龄、性别、健康状况等相关因素,同时通过接触者的身体状况或临床症状并结合应用结核菌素(PPD)试

DOI: 10.3760/cma.j.issn.0254-6450.2009.06.007

基金项目:国家自然科学基金(30571599)

作者单位:430030 武汉,华中科技大学同济医学院医药卫生管理学院(陶红兵、苗卫军、余瑶、郭石林、陈璞);湖北省疾病预防控制中心传染病防治研究所(叶建君、侯双翼);华中科技大学同济医学院公共卫生学院(熊光练)

验、X线检查和痰涂片检查对高危人群进行筛选(在当地医院进行),通过快速、特异的斑点金免疫渗透试验(DIGFA)对高危人群的感染情况进一步确认。

(2)问卷调查:设计统一的调查表,由经过统一培训的调查员入户对患者进行采访,搜集传染期的经常活动地点和接触者的准确信息,根据资料确定密切接触者,采访密切接触者(包括一般情况、结核病相关知识的知晓程度等),讲解对密切接触者调查目的、肺结核传播途径、传播危险以及对疾病进行检查治疗和随访的重要性,评价密切接触者的环境特征。

(3)接触者评价:密切接触者均行 PPD 试验、3 次痰检和胸部 X 线透视(胸透),胸透异常或 PPD 阳性者或有症状者再行 X 线拍片。对密切接触者或高危人群做 DIGFA 检测,暴露 10~12 周内初始皮试阳性的接触者于窗口期之后重新检查。

(4)判断标准:①PPD 结果判定标准:成年人局部无反应或硬结  $\leq 5$  mm 为阴性,5~9 mm 为弱阳性,10~19 mm 为阳性,硬结  $\geq 20$  mm 甚至有水泡或溃烂为强阳性;儿童局部无反应或硬结  $\leq 5$  mm 为阴性,5~15 mm 为阳性,硬结  $\geq 15$  mm 甚至有水泡或溃烂为强阳性。②胸部 X 线检查:分为絮状影、片状影、索条影、空洞与空腔影、钙化影、结节影、肿块影、肺门异常影;对活动性肺结核密切接触者采用 DIGFA 进行诊断。

3. 统计学分析:采用 EpiData 软件建立数据库,SPSS 15.0 软件进行统计分析,使用 logistic 逐步回归分析筛选关键影响因素,采用逐步 Bayes 判别法对肺结核的密切接触者发病的关键因素进行筛选并建立判别模型。

## 结 果

1. 基本情况:本研究 983 例结核患者中,痰涂片阳性 857 例(其中 396 例痰检为 2+ 或 2+ 以上)。X 线检查结果 490 例为双侧肺结核,488 例为一侧肺结核;919 例有咳嗽症状;631 例咳白色黏液痰;219 例有咯血史;445 例有发热症状。2637 名密切接触者中,发病 68 例,患病率为 2.58%。肺结核患者及其密切接触者性别、年龄及城乡分布情况见表 1。

### 2. 判别模型的建立:

(1)变量的赋值及筛选:将影响活动性肺结核密切接触者发病四个主要方面(患者病情、密切接触者健康状况、密切接触者与患者的接触情况及密切接触者的人口学因素)的 29 个相关影响因素作为自变

表 1 山东、湖北、广东三省 983 例肺结核患者及 2637 名密切接触者基本情况

特征	肺结核患者(n=983)	密切接触者(n=2637)
性别 男	684(69.6)	1213(46.0)
女	299(30.4)	1424(54.0)
年龄(岁) 0~	1(0.1)	233(8.8)
15~	543(55.2)	1436(54.5)
45~	196(19.9)	696(26.4)
$\geq 60$	243(24.7)	272(10.3)
城乡 城市	264(26.9)	583(22.1)
农村	719(73.1)	2054(77.9)

注:括号外数据为例数,括号内数据为构成比(%)

量 X, 密切接触者是否发病作为因变量 Y, 进行 logistic 回归分析(有条件的向前逐步回归法)( $\alpha = 0.05$ ), 筛选出患者痰涂片阳性次数、接触者文化程度、接触者平均每天和患者接触时间、咳嗽时回避情况、居室通风情况、接触者咳嗽、接触者发热 7 个关键影响因素(表 2, 3)。

(2)判别模型的建立:采用判别分析法对 logistic 回归分析筛选出的 7 个关键影响因素进行逐步判别分析, 7 个变量均被纳入判别函数(表 4), 这些变量对正确判断分类均有作用。根据纳入判别函数的 7 个变量, 计算出判别函数系数(表 5)。

根据判别函数系数, 建立肺结核密切接触者 Bayes 判别函数:①密切接触者未发病判别函数:  $Y_1 = -39.831$  (常数) +  $1.927 X_1$  (痰涂片阳性次数) +  $3.528 X_2$  (文化程度) +  $0.309 X_3$  (接触时间) +  $5.893 X_4$  (咳嗽时回避) +  $2.140 X_5$  (居室通风) +  $8.706 X_6$  (咳嗽) +  $30.970 X_7$  (发热)。②密切接触者发病判别函数:  $Y_2 = -57.875$  (常数) +  $2.343 X_1$  (痰涂片阳性次数) +  $3.965 X_2$  (文化程度) +  $0.361 X_3$  (接触时间) +  $6.296 X_4$  (咳嗽时回避) +  $1.348 X_5$  (居室通风) +  $12.984 X_6$  (咳嗽) +  $36.555 X_7$  (发热)。

## 讨 论

我国肺结核病患者 80% 是在农村地区<sup>[1]</sup>。本研究依据 2005 年 9 月至 2007 年 9 月活动性肺结核患者总人数 50% 的比例随机抽取一定数量的活动性肺结核病例, 使农村样本量远远大于城市的样本量, 抽样的结果显示城市肺结核接触者 583 例(22.1%), 农村肺结核接触者 2054 名(77.9%), 是符合我国肺结核患者集中在农村, 其密切接触者也会相对集中在农村的分布情况<sup>[2]</sup>。

本研究采用 logistic 逐步回归法与 Bayes 判别法结合的方法对肺结核密切接触者是否发病进行判定

表 2 变量赋值

对象及变量	赋 值
患者	
咳嗽	无咳嗽=1, 轻微咳嗽=2, 剧烈干咳=3, 轻微咳嗽咳痰=4, 剧烈咳嗽咳痰=5
咳痰	无咳痰=1, 白色黏液痰=2, 黄色黏液性痰=3, 咳出干酪样物质=4
咯血	无咯血=1, 痰中带少量血=2, 少量咯血=3, 大咯血=4
发热	无发热=1, 低热=2, 中度发热=3, 高热=4
胸痛	有=1, 无=2
体重减轻	明显减轻=1, 有一点减轻=2, 无减轻=3, 不清楚=4
痰涂片阳性次数	实际次数
延迟治疗	延迟治疗时间
接触者	
性别	男=1, 女=2
年龄(岁)	0~ =1, 15~ =2, 45~ =3, ≥60=4
职业	工人=1, 农民=2, 学生=3, 退休人员=4, 下岗工人=5, 无业人员=6, 其他=7
文化程度	文盲=1, 小学=2, 初中=3, 中专及高中=4, 大专以上=5
接触时间	平均每天接触时间
接触次数	非常多=1, 比较多=2, 一般=3, 偶尔=4, 没有=5
口罩	每次都戴=1, 大多数时间戴=2, 有时候戴=3, 没有戴过=4
咳嗽时回避	每次都回避=1, 大多数时间回避=2, 有时回避=3, 没有回避=4
居室通风	很好=1, 较好=2, 一般=3, 差=4
居住潮湿	很潮湿=1, 较潮湿=2, 一般=3, 不潮湿=4
日照	很好=1, 较好=2, 一般=3, 差=4
咳嗽	接触者无咳嗽=1, 接触者有咳嗽=2
发热	接触者无发热=1, 接触者有发热=2
环境	非粉尘环境=1, 粉尘环境=2
卡介苗	接种=1, 无接种=2, 不知道=3
吸烟	不吸=1, 吸=2
饮酒	不饮=1, 饮=2
药物	无药物依赖=1, 有药物依赖=2
疲劳	无过度疲劳感=1, 有过度疲劳感=2
情绪	无情绪低落感=1, 有情绪低落感=2
城乡	城市=1, 农村=2

表 3 logistic 回归分析结果

变量	$\beta$	$s_e$	Wald $\chi^2$ 值	P 值	OR 值(95%CI)
痰涂片阳性次数	0.409	0.167	6.027	0.014	1.506(1.086 ~ 2.088)
文化程度	-0.405	0.155	6.814	0.009	1.499(1.106 ~ 2.033)
经常接触	0.062	0.025	6.067	0.014	1.064(1.013 ~ 1.118)
咳嗽时回避	0.602	0.215	7.850	0.005	1.826(1.198 ~ 2.781)
居室通风	-1.076	0.273	15.593	0.000	0.341(0.200 ~ 0.582)
咳嗽	2.317	0.375	38.157	0.000	10.140(4.862 ~ 21.148)
发热	2.058	0.497	17.152	0.000	7.827(2.956 ~ 20.723)
常数	-11.617	1.387	70.183	0.000	0.000

分析,考察究竟是哪些关键因素影响密切接触者发病,并分别建立密切接触者未发病与发病的判别模型,采用交互验证法(cross-validated)与自身验证法对模型的判别效能进行验证,从交互验证结果看,判别函数对密切接触者未发病的判别正确率达 98%,对密切接触者发病的判别正确率达 71.9%,总

表 4 纳入判别模型的变量

序号	变量	统计量	F 检验	
			统计量	P 值
1	咳嗽	0.952	115.601	0.000
2	发热	0.936	77.512	0.000
3	居室通风	0.928	58.829	0.000
4	痰涂片阳性次数	0.923	47.322	0.000
5	文化程度	0.921	39.208	0.000
6	咳嗽时回避	0.918	33.699	0.000
7	经常接触	0.917	29.570	0.000

表 5 Bayes 判别函数系数

变量	未发病( $Y_1$ )	发病( $Y_2$ )
痰涂片阳性次数( $X_1$ )	1.927	2.343
文化程度( $X_2$ )	3.528	3.965
接触时间( $X_3$ )	0.309	0.361
咳嗽时回避( $X_4$ )	5.893	6.296
居室通风( $X_5$ )	2.140	1.348
咳嗽( $X_6$ )	8.706	12.984
发热( $X_7$ )	30.970	36.555
常数	-39.831	-57.875

正确率为 96.3%,与其自身验证结果一致,说明该判别模型有比较高的灵敏性和稳定性,能够较为准确地对密切接触者是否发病加以判别。通过文献研究,国内已有对输卵管妊娠早期诊断的判别模型<sup>[3]</sup>,国外有关于肺结核患者接触者调查的决策树研究<sup>[4]</sup>,但尚未有肺结核密切接触者早期主动发现的判别模型。

通过使用该判别模型针对任一新观测到的密切接触者,都可以将其患者痰涂片阳性次数、接触者的文化程度、接触者平均每天和患者接触时间、咳嗽时回避情况、居室通风情况、接触者咳嗽情况、接触者发热情况依据表 2 进行赋值,然后将赋值分别带入两个判别函数,分别计算出  $Y_1$ 、 $Y_2$ ,若  $Y_1 > Y_2$ ,则该密切接触者未发病;反之,则该密切接触者发病。

本研究中 Bayes 判别模型的建立是希望能够对肺结核的早期发现和发现提供科学准确依据,能够从量化角度对结核病的控制策略进行探索性研究。该判别模型的实用价值,在将来的实际应用中得到实证验证和效果评价。

参 考 文 献

[1] 全国结核病流行病学抽样调查技术指导组. 第四次全国结核病流行病学抽样调查报告. 中华结核和呼吸杂志, 2002, 25(1): 3-7.

[2] 杨本付, 宋红梅, 高立, 等. 272 例涂阳肺结核病例的人口学及临床特征分析. 中国热带医学, 2007, 7(11): 2005-2006.

[3] 魏伏冰, 苏宜香, 何锐志. Bayes 判别分析对输卵管妊娠早期诊断价值的探讨. 广东医学, 2006, 27(11): 1687-1688.

[4] Gerald LB, Tang S, Bruce F, et al. A decision tree for tuberculosis contact investigation. Am J Respir Crit Care Med, 2002, 166: 1122-1127.

(收稿日期: 2008-09-11)

(本文编辑: 张林东)