

# 基于多因素 logistic 回归模型和决策树模型的肺结核患者就诊延迟及确诊延迟影响因素分析

马晓格<sup>1</sup> 张立杰<sup>2</sup> 高汉青<sup>3</sup> 包城<sup>4</sup> 吴越<sup>3</sup> 吴思慧<sup>1</sup> 刘梦晗<sup>1</sup> 刘宇红<sup>2</sup> 李亮<sup>2</sup>

<sup>1</sup>山东大学齐鲁医学院公共卫生学院流行病学系, 济南 250012; <sup>2</sup>首都医科大学附属北京胸科医院/北京市结核病胸部肿瘤研究所/中国疾病预防控制中心结核病防治临床中心, 北京 101149; <sup>3</sup>北京市通州区疾病预防控制中心结核病防治所, 北京 101149; <sup>4</sup>北京市昌平区结核病防治所, 北京 102200

通信作者: 李亮, Email: liliang69@vip.sina.com

**【摘要】目的** 了解北京市通州区及昌平区肺结核患者就诊延迟及确诊延迟情况并分析其影响因素, 为早期发现及科学管理肺结核患者提供有效建议。**方法** 采用回顾性流行病学调查方法, 利用全国结核病管理信息系统收集北京市通州区及昌平区于 2021 年 1 月 1 日至 12 月 31 日登记的肺结核患者个案资料, 采用电话访谈对部分信息进行补充, 采用多因素 logistic 回归模型及决策树模型分析肺结核患者就诊延迟及确诊延迟的影响因素。**结果** 2021 年北京市通州区及昌平区肺结核患者就诊延迟时间  $M(Q_1, Q_3)$  为 11(5, 26)d, 就诊延迟率为 41.71%。多因素分析结果显示, 健康体检( $OR=0.033, 95\%CI: 0.008\sim 0.147$ )、就诊前咳嗽咳痰 < 2 周, 且有任一结核病症状( $OR=0.378, 95\%CI: 0.215\sim 0.665$ )、就诊前有其他症状( $OR=2.791, 95\%CI: 1.710\sim 4.555$ )及就诊期间无需上班或上学( $OR=2.990, 95\%CI: 1.419\sim 6.298$ )为就诊延迟的影响因素; 肺结核患者确诊延迟时间  $M(Q_1, Q_3)$  为 8(0, 18)d, 确诊延迟率为 35.20%。多因素分析结果显示, 确诊单位为结核病专科医院( $OR=0.426, 95\%CI: 0.236\sim 0.767$ )和结核病防治所( $OR=1.843, 95\%CI: 1.061\sim 3.202$ )以及患者来源为追踪( $OR=2.632, 95\%CI: 1.062\sim 6.521$ )为确诊延迟的影响因素。多因素 logistic 回归模型与决策树模型的整体效能相当, 决策树模型的灵敏度高于多因素 logistic 回归模型, 特异度低于多因素 logistic 回归模型。**结论** 2021 年北京市通州区及昌平区肺结核患者就诊延迟及确诊延迟处于较低水平, 但仍需加强宣传教育并积极开展主动筛查, 提高市民对结核病的防治意识, 同时进一步提升医疗服务水平和改善就医可及性, 减少患者就医及诊断延迟现象。

**【关键词】** 结核, 肺; 就诊延迟; 确诊延迟; 多因素 logistic 回归模型; 决策树模型

**基金项目:** 北京结核病综合防控关键技术研究(D18110000418005); 北京市卫生健康委员会高层次公共卫生技术人才项目(学科带头人-01-11)

## Analysis on influencing factors of medical care seeking delay and diagnosis delay of pulmonary tuberculosis patients based on logistic regression model and decision tree model

Ma Xiaoge<sup>1</sup>, Zhang Lijie<sup>2</sup>, Gao Hanqing<sup>3</sup>, Bao Cheng<sup>4</sup>, Wu Yue<sup>3</sup>, Wu Sihui<sup>1</sup>, Liu Menghan<sup>1</sup>, Liu Yuhong<sup>2</sup>, Li Liang<sup>2</sup>

<sup>1</sup>Department of Epidemiology, School of Public Health, Cheeloo College of Medicine, Shandong University, Jinan 250012, China; <sup>2</sup>Beijing Chest Hospital, Capital Medical University, Beijing

DOI: 10.3760/cma.j.cn112338-20240224-00080

收稿日期 2024-02-24 本文编辑 万玉立

引用格式: 马晓格, 张立杰, 高汉青, 等. 基于多因素 logistic 回归模型和决策树模型的肺结核患者就诊延迟及确诊延迟影响因素分析[J]. 中华流行病学杂志, 2024, 45(5): 721-729. DOI: 10.3760/cma.j.cn112338-20240224-00080.

Ma XG, Zhang LJ, Gao HQ, et al. Analysis on influencing factors of medical care seeking delay and diagnosis delay of pulmonary tuberculosis patients based on logistic regression model and decision tree model[J]. Chin J Epidemiol, 2024, 45(5): 721-729. DOI: 10.3760/cma.j.cn112338-20240224-00080.



Tuberculosis and Thoracic Tumor Research Institute, Clinical Center for Tuberculosis, Chinese Center for Disease Control and Prevention, Beijing 101149, China; <sup>3</sup>Institute for Tuberculosis Prevention and Control, Tongzhou District Center for Disease Prevention and Control, Beijing 101149, China; <sup>4</sup>Changping Institute for Tuberculosis Prevention and Treatment, Beijing 102200, China  
Corresponding author: Li Liang, Email: liliang69@vip.sina.com

**【Abstract】 Objective** To investigate the status of medical care seeking delay and diagnosis delay of pulmonary tuberculosis (PTB) patients in Tongzhou District and Changping District of Beijing, analyze the related factors and put forward suggestions for early detection and scientific management of PTB patients. **Methods** A retrospective epidemiological survey was conducted to collect the incidence data of PTB registered in Tongzhou and Changping from January 1 to December 31, 2021 by using the Chinese Tuberculosis Information Management System, and telephone interview were used for information supplement. Multivariate logistic regression model and decision tree model were used to analyze the influencing factors of medical care seeking delay and diagnosis delay of PTB patients. **Results** In 2021, the medical care seeking delay time  $M(Q_1, Q_3)$  in the PTB patients in Tongzhou and Changping was 11 (5, 26) days, with a delay rate of 41.71%. Results from multivariate logistic regression model analysis revealed that factors influencing the medical care seeking delay included regular health check-up ( $OR=0.033$ , 95% $CI$ : 0.008-0.147), coughing for less than 2 weeks or showing any symptom of PTB before medical care seeking ( $OR=0.378$ , 95% $CI$ : 0.215-0.665), showing other symptoms before medical care seeking ( $OR=2.791$ , 95% $CI$ : 1.710-4.555), no work or school in medical care seeking ( $OR=2.990$ , 95% $CI$ : 1.419-6.298). The diagnosis delay time  $M(Q_1, Q_3)$  in the PTB patients was 8 (0, 18) days, with a delay rate of 35.20%. Multivariate logistic regression model analysis revealed that the factors influencing the diagnosis delay of PTB included being diagnosed at a specialized tuberculosis hospital ( $OR=0.426$ , 95% $CI$ : 0.236-0.767) or a tuberculosis prevention and control institution ( $OR=1.843$ , 95% $CI$ : 1.061-3.202) and being traced as a source of infection ( $OR=2.632$ , 95% $CI$ : 1.062-6.521). The overall performance of the multivariate logistic regression model was comparable to that of the decision tree model, with the decision tree model exhibiting higher sensitivity but lower specificity. **Conclusions** The medical care seeking delay rate and diagnosis delay rate of tuberculosis in Tongzhou and Changping were at low levels in 2021. However, it is still necessary to strengthen the health education and active screening, improve the public awareness of PTB prevention and control, and further improve the level of medical services and medical access to reduce the medical care seeking delay and diagnosis delay of PTB patients.

**【Key words】** Tuberculosis, pulmonary; Medical care seeking delay; Diagnosis delay; Multivariate logistic regression model; Decision tree model

**Fund programs:** Key Techniques for Comprehensive Prevention and Control of Tuberculosis in Beijing (D181100000418005); High-level Public Health Technical Talents Project of Beijing Health Commission (Academic Leader-01-11)

结核病是一种慢性呼吸道传染病,位列全球死因第13位,严重危害人类的身心健康,在30个结核病高负担国家中,我国结核病估算发病人数位居全球第3,占全球结核病例的7.4%<sup>[1]</sup>。由于患者对结核病认识不足导致的就诊延迟及医疗卫生机构未及时确诊导致的确诊延迟增加了人群感染结核病的风险<sup>[2]</sup>。我国第五次结核病流行病学抽样调查结果显示,有症状的肺结核患者超过50%未及时就诊<sup>[3]</sup>。目前,国内针对肺结核患者延迟的研究大多采用多因素logistic回归模型分析<sup>[4]</sup>。决策树是数据挖掘中有效的分类规则学习方法之一<sup>[5]</sup>,其结果用树形图展现,不仅可以显示因变量的影响因素,还可以深入分析某变量在各亚层的影响方式,能够充分挖掘数据信息,目前广泛应用于公共卫生领

域<sup>[6]</sup>。本研究旨在使用多因素logistic回归模型和决策树模型,从不同角度探讨肺结核患者就诊及确诊延迟的影响因素,并为早发现肺结核患者及实施科学有效的防控计划提供理论依据。

## 资料与方法

1. 资料来源:来源于全国结核病管理信息系统(结核病专报系统),信息包含患者性别、年龄、职业、户籍、患者来源、诊断结果等信息。就诊前症状、首诊单位、确诊单位及转诊单位数量等,均通过电话访谈获取。

2. 研究对象:北京市通州区及昌平区2021年1月1日至12月31日在结核病专报系统中排除肺

外结核及信息不完全者共 621 例患者符合纳入条件,成功随访 547 例肺结核患者,排除电话访谈问卷回答不全及不愿参与调查的患者,收集有效问卷共 537 份。本研究通过首都医科大学附属北京胸科医院伦理委员会审查(批准文号:2018-30),研究对象均签署知情同意书。

### 3. 相关定义:

(1)病原学阳性:指痰菌检查结果为结核分枝杆菌涂片抗酸染色、培养或分子生物学检测任一阳性者;

(2)无病原学结果:指未接受过任何一种结核分枝杆菌相关病原学检测者;

(3)就诊延迟:指患者从出现症状至首次就诊于医疗卫生机构的时间间隔 $>14$  d;

(4)确诊延迟:指患者从首次就诊至确诊的时间间隔 $>14$  d<sup>[7-8]</sup>;

(5)重点人群:指 HIV/AIDS 患者、糖尿病患者、学校托幼机构人员、监管人员、养殖人员、接尘人员/尘肺病患者、养老院/福利院居住者、病原学阳性肺结核患者的密切接触者、医务人员或其他人群聚集场所人员;

(6)推介:指医务人员或有关人员将发现的肺结核可疑症状者推荐并督促其到结核病防治机构接受检查;

(7)转诊:指病例出现肺结核可疑症状后到医疗卫生机构(不包括结核病防治机构)就诊,经胸部 X 线或痰菌检查等诊断为肺结核或疑似肺结核后,患者携带医生填写的转诊单到结核病防治机构就诊;

(8)追踪:指疾病预防控制机构督促已进行疫情报告但未到结核病防治机构就诊的肺结核和疑似肺结核患者到结核病防治机构进行诊治<sup>[9]</sup>;

(9)任一结核病症状:指患者具有咯血、血痰、发热、盗汗、体重减轻、气短、胸痛和疲劳中任

一者<sup>[10]</sup>。

4. 研究方法:本研究为回顾性流行病学调查。采用电话访谈问卷调查法,经患者本人或知晓其具体病情的家属同意后展开调查。课题组成员经过统一的培训后开展电话访谈。患者的社会人口学信息(性别、职业、户籍等)和疾病特征信息(诊断结果、是否有合并症等)主要从结核病专报系统获得,患者就医行为(首诊单位、确诊单位及转诊单位数量等)的相关信息由电话访谈获得。

5. 统计学分析:采用 Excel 2016 软件对收集数据进行整理,计量资料服从近似正态分布的用  $\bar{x}\pm s$  表示,呈偏态分布时用  $M(Q_1, Q_3)$  表示,计数资料以率或构成比(%)表示。运用 SPSS 23.0 软件进行统计学分析。采用  $\chi^2$  检验、多因素 logistic 回归模型(变量赋值情况见表 1)及决策树模型分析肺结核患者的相关特征性因素与就诊及确诊延迟之间的相关关系,均以  $P<0.05$  为差异有统计学意义。本研究采用决策树模型的  $\chi^2$  自动交互检测法,先对变量进行过滤,再根据不同的模型进行参数设置,为防止模型过度拟合采用预修剪的方法进行剪枝,最后采用 MedCalc 20.0.4 软件绘制受试者工作特征曲线分析比较两种模型的预测价值,曲线检验方法为 Delong 检验。

## 结 果

1. 一般情况:本研究纳入 537 例肺结核患者,男性 344 例(64.1%);年龄为(40.68 $\pm$ 17.23)岁;职业以工人及农民(168 例,31.3%)和家务及待业人员(165 例,30.7%)为主;共 388 例(72.3%)患者来自于转诊;病原学阳性患者 264 例(49.2%);首诊单位以综合医院/中医院为主(417 例,77.7%);确诊单位以结核病专科医院为主(226 例,42.1%)。见表 2。

2. 延迟基本情况:537 例肺结核患者中,就诊

表 1 多因素 logistic 回归模型分析变量赋值情况

变 量	赋 值
户籍	0=外地;1=本地
患者来源	1=转诊;2=直接就诊;3=健康体检;4=推介;5=追踪
就诊前症状	1=无症状;2=咳嗽咳痰 $<2$ 周,且有任一结核病症状;3=咳嗽咳痰 $\geq 2$ 周;4=其他 <sup>a</sup>
首诊单位	1=结核病定点/专科医院;2=综合医院/中医院;3=基层医疗卫生机构;4=私人诊所;5=其他
确诊单位	1=结核病专科医院;2=结核病防治所;3=综合医院;4=其他
转诊单位数量	1=0;2=1;3=2;4= $\geq 3$
就诊期间是否坚持上班或上学	1=需要且坚持上班或上学;2=有时上有时不上;3=需要但未坚持上班或上学;4=无需上班或上学
共同居住人数	1=独居;2=1;3=2;4= $\geq 3$

注:<sup>a</sup>包括发热、胸闷、疼痛、咽喉不适、精神障碍、脖子红肿硬块、腹胀、心悸、发颤及腿部红斑

表2 肺结核患者就诊延迟及确诊延迟单因素分析

变 量	例数(构成比,%)	就诊延迟			确诊延迟		
		例数(率,%)	$\chi^2$ 值	P值	例数(率,%)	$\chi^2$ 值	P值
性别			0.67	0.412		0.23	0.628
男	344(64.1)	139(40.4)			123(35.8)		
女	193(35.9)	85(44.0)			65(33.7)		
年龄组(岁)			1.93	0.587		0.52	0.915
10~	182(33.9)	82(45.1)			65(35.7)		
30~	197(36.7)	82(41.6)			68(34.5)		
50~	106(19.7)	39(36.8)			35(33.0)		
70~	52(9.7)	21(40.4)			20(38.5)		
职业			3.91	0.689		7.85	0.249
干部职员、教师、医务人员	70(13.0)	29(41.4)			25(35.7)		
工人及农民	168(31.3)	69(41.1)			56(33.3)		
家务及待业人员	165(30.7)	75(45.5)			68(41.2)		
离退人员	47(8.8)	14(29.8)			14(29.8)		
商业服务人员	60(11.2)	25(41.7)			14(23.3)		
学生	21(3.9)	9(42.9)			8(38.1)		
其他	6(1.1)	3(50.0)			5(83.3)		
重点人群 <sup>a</sup>			0.15	0.703		0.03	0.859
是	78(14.5)	31(39.7)			28(35.9)		
否	459(85.5)	193(42.0)			160(34.9)		
户籍			1.30	0.254		4.26	0.039
本地	203(37.8)	91(44.8)			60(29.6)		
外地	334(62.2)	133(39.8)			128(38.3)		
患者来源			18.92	0.001		14.29	0.006
健康体检	32(6.0)	2(6.3)			8(25.0)		
直接就诊	43(8.0)	22(51.2)			19(44.2)		
转诊	388(72.3)	166(42.8)			125(32.2)		
推介	49(9.0)	22(44.9)			20(40.8)		
追踪	25(4.7)	12(48.0)			16(64.0)		
病原学诊断结果			0.00	0.983		0.96	0.326
阳性	264(49.2)	110(41.7)			87(33.0)		
阴性或无	273(50.8)	114(41.8)			101(37.0)		
合并症 <sup>b</sup>			0.27	0.606		0.44	0.508
有	51(9.5)	23(45.1)			20(39.2)		
无	486(90.5)	201(41.4)			168(34.6)		
就诊前症状			45.28	<0.001		4.14	0.247
咳嗽咳痰<2周,且有任一结核病症状	116(21.6)	27(23.3)			46(39.7)		
咳嗽咳痰≥2周	145(27.0)	62(42.8)			45(31.0)		
其他 <sup>c</sup>	77(14.3)	20(26.0)			22(28.6)		
无症状	199(37.1)	115(57.8)			75(37.7)		
首诊单位			13.66	0.008		5.39	0.250
结核病定点/专科医院	87(16.2)	48(55.2)			30(34.5)		
综合医院/中医院	417(77.7)	157(37.6)			144(34.5)		
基层医疗卫生机构	15(2.8)	10(66.7)			6(40.0)		
私人诊所	4(0.7)	2(50.0)			0(0.0)		
其他	14(2.6)	7(50.0)			8(57.1)		
确诊单位			5.15	0.161		54.42	<0.001
结核病专科医院	226(42.1)	95(42.0)			41(18.1)		
结核病防治所	210(39.1)	78(37.1)			108(51.4)		
综合医院	90(16.8)	46(51.1)			36(40.0)		
其他	11(2.0)	5(45.5)			3(27.3)		
转诊单位数量			5.79	0.122		14.71	0.002
0	37(6.9)	21(56.8)			9(24.3)		
1	257(47.9)	108(42.0)			73(28.4)		
2	199(37.0)	74(37.2)			87(43.7)		
≥3	44(8.2)	21(47.7)			19(43.2)		
就诊期间是否坚持上班或上学			11.64	0.009		6.95	0.074
需要且坚持上班或上学	116(21.6)	50(43.1)			34(29.3)		
有时上有时不上	55(10.2)	34(61.8)			23(41.8)		
需要但未坚持上班或上学	266(49.6)	105(39.5)			87(32.7)		
无需上班或上学	100(18.6)	35(35.0)			44(44.0)		
共同居住人数			1.90	0.592		8.21	0.042
独居	85(15.8)	41(48.2)			24(28.2)		
1	203(37.8)	83(40.9)			72(35.5)		
2	160(29.8)	63(39.4)			68(42.5)		
≥3	89(16.6)	37(41.6)			24(27.0)		

注:<sup>a</sup>指 HIV/AIDS 患者、糖尿病患者、学校托幼机构人员、监管人员、养殖人员、接尘人员/尘肺病患者、养老院/福利院居住者、病原学阳性肺结核患者的密切接触者、医务人员或其他人群聚集场所人员;<sup>b</sup>包括糖尿病、尘肺及其他相关疾病;<sup>c</sup>包括发热、胸闷、疼痛、咽喉不适、精神障碍、脖子红肿硬块、腹胀、心悸、发颤及腿部红斑



延迟时间  $M(Q_1, Q_3)$  为 11(5, 26)d, 最短就诊时间为 0 d, 最长就诊时间为 46 d, 就诊延迟率为 41.71%; 肺结核患者确诊延迟时间  $M(Q_1, Q_3)$  为 8(0, 18)d, 最短确诊时间为 0 d, 最长确诊时间为 35 d, 确诊延迟率为 35.20%。

3. 肺结核患者就诊延迟及确诊延迟单因素分析: 单因素分析结果显示, 患者来源、就诊前症状、首诊单位及就诊期间是否坚持上班或上学是就诊延迟的影响因素(均  $P < 0.05$ ); 户籍、患者来源、确诊单位、转诊单位数量及共同居住人数是确诊延迟的影响因素(均  $P < 0.05$ )。见表 2。

4. 肺结核患者就诊延迟及确诊延迟多因素 logistic 回归模型分析: 就诊延迟多因素结果发现, 健康体检( $OR=0.033, 95\%CI: 0.008\sim 0.147$ ), 就诊前咳嗽咳痰  $< 2$  周, 且有任一结核病症状( $OR=0.378, 95\%CI: 0.215\sim 0.665$ )为就诊延迟的保护因素; 就诊前有其他症状( $OR=2.791, 95\%CI: 1.710\sim 4.555$ )及就诊期间无需上班或上学( $OR=2.990, 95\%CI: 1.419\sim 6.298$ )为就诊延迟的危险因素。见表 3。无需上班或上学人群的平均年龄为 60.89 岁, 并且以家务及待业人员(34.0%)和离退休人员(31.0%)为主。

确诊延迟多因素分析结果显示, 确诊单位为结核病专科医院( $OR=0.426, 95\%CI: 0.236\sim 0.767$ )为确诊延迟的保护因素; 患者来源为追踪( $OR=2.632, 95\%CI: 1.062\sim 6.521$ )及确诊单位为结核病防治所( $OR=1.843, 95\%CI: 1.061\sim 3.202$ )为确诊延迟的危

险因素。见表 4。

5. 肺结核患者就诊延迟及确诊延迟决策树模型分析: 以就诊延迟为因变量建立的影响因素决策树模型共有 7 个节点, 5 个终结点, 树的深度为 2 层, 进入模型的因素依次是就诊前症状、患者来源。其中以就诊前无症状、患者来源为直接就诊或追踪的患者发生就诊延迟的概率最高, 为 88.9%。见图 1。

以确诊延迟为因变量建立的影响因素决策树模型共有 7 个节点, 4 个终结点, 树的深度为 2 层, 进入模型的因素依次是户籍、确诊单位。其中以户籍为外地、确诊单位为结核病防治所、综合医院及其他的患者发生确诊延迟的概率最高, 为 48.8%。见图 2。

6. 2 种模型结果比较: 就诊延迟多因素 logistic 回归模型的曲线下面积(AUC)、特异度和约登指数均高于决策树模型; 但决策树模型的灵敏度高于 logistic 回归模型。见表 5。2 种模型 AUC 均接近 0.750, 表明分类预测效果有实际意义, 并且模型分类均具有一定的准确性。两者比较, 差异无统计学意义( $Z=1.737, P=0.082$ )。见图 3。

确诊延迟多因素 logistic 回归模型的 AUC、特异度和约登指数均高于决策树模型; 决策树模型的灵敏度高于多因素 logistic 回归模型。见表 6。2 种模型 AUC 均接近 0.700, 表明分类预测效果有实际意义, 并且模型分类均具有一定的准确性。两者比

表 3 肺结核患者就诊延迟多因素 logistic 回归模型分析

变 量	$\beta$ 值	$s_x$	$\chi^2$ 值	P 值	OR 值	95%CI
患者来源						
转诊					1.000	
直接就诊	0.081	0.371	0.05	0.827	1.084	0.524~2.242
健康体检	-3.402	0.758	20.16	<0.001	0.033	0.008~0.147
推介	0.001	0.341	0.00	0.998	1.001	0.513~1.951
追踪	0.097	0.451	0.05	0.830	1.102	0.455~2.666
就诊前症状						
无症状					1.000	
咳嗽咳痰 $< 2$ 周, 且有任一结核病症状	-0.972	0.287	11.43	0.001	0.378	0.215~0.665
咳嗽咳痰 $\geq 2$ 周	-0.592	0.331	3.20	0.074	0.553	0.289~1.059
其他*	1.027	0.250	16.88	<0.001	2.791	1.710~4.555
就诊期间是否坚持上班或上学						
需要且坚持上班或上学					1.000	
需要但未坚持上班或上学	0.066	0.260	0.06	0.799	1.069	0.641~1.780
有时上有时不上	-0.010	0.321	0.00	0.975	0.990	0.528~1.856
无需上班或上学	1.095	0.380	8.30	0.004	2.990	1.419~6.298

注: \*包括发热、胸闷、疼痛、咽喉不适、精神障碍、脖子红肿硬块、腹胀、心悸、发颤及腿部红斑

表 4 肺结核患者确诊延迟多因素 logistic 分析

变量	β 值	<i>s<sub>e</sub></i>	χ <sup>2</sup> 值	P 值	OR 值	95%CI
患者来源						
转诊					1.000	
直接就诊	0.142	0.350	0.16	0.685	1.152	0.581~2.287
健康体检	-0.025	0.452	0.00	0.956	0.976	0.402~2.367
推介	-0.313	0.333	0.88	0.348	0.731	0.380~1.406
追踪	0.968	0.463	4.35	0.037	2.632	1.062~6.521
确诊单位						
综合医院					1.000	
结核病专科医院	-0.854	0.300	8.10	0.004	0.426	0.236~0.767
结核病防治所	0.611	0.282	4.70	0.030	1.843	1.061~3.202
其他	-0.471	0.738	0.41	0.523	0.624	0.147~2.651

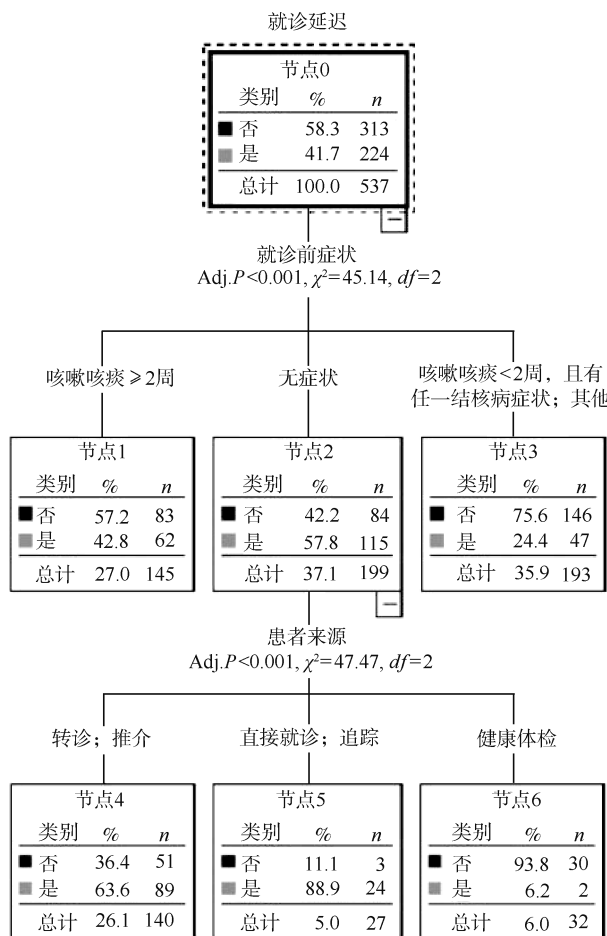


图 1 肺结核患者就诊延迟影响因素的决策树模型

较, 差异无统计学意义 ( $Z=1.521, P=1.034$ )。见图 4。

### 讨论

本研究结果显示, 2021 年北京市通州区及昌平区肺结核患者就诊延迟率为 41.71%, 低于同期国内外的一些研究结果<sup>[11-16]</sup>, 分析其原因可能与北

京市肺结核报告发病较低、经济发展水平较高、医疗卫生水平和可及性较高有关。

多因素 logistic 回归模型及决策树模型分析研究结果均表明, 就诊前具有咳嗽咳痰 < 2 周, 且有任一结核病症状是患者发生就诊延迟的保护因素, 该结果与已有研究结果一致<sup>[17-18]</sup>, 分析其原因可能是胸痛和体重减轻等较为严重症状的出现促使患者早期就医<sup>[19]</sup>。而就诊前出现发热及胸闷等容易与其他疾病相混淆的肺结核的非典型临床症状则会导致患者更容易出现就诊延迟<sup>[20]</sup>。相较于直接就诊、推介及追踪等被动发现模式, 以健康体检为主的主动发现模式发生就诊延迟的概率更低。有相关报道, 主动病例发现在及时发现结核病患者方面是有效的<sup>[21]</sup>, WHO 指南<sup>[22]</sup>以及我国结核病防治规划<sup>[9]</sup>均要求应进一步优化以控制传染源为主的策略基础上, 增加和强化新的干预措施, 开展重点人群筛查, 加强患者主动发现。多因素 logistic 回归模型显示, 相较于就诊期间坚持上班或上学的患者, 无需上班或上学的患者发生就诊延迟的可能性更大, 这部分人群大多是离退休人员、家务及待业人员, 可能与该群体就诊意愿低下有关<sup>[23]</sup>。

确诊延迟多因素分析结果显示, 结核病专科医院确诊延迟的可能性更小, 这与专科医院医务人员的鉴别能力及筛查技术有关。专科医院的医务工作人员往往对结核病的警惕性及鉴别能力较强, 同时, 专科医院大多已经使用更灵敏、更快速的分子生物学检测技术(如 GeneXpert)能在一定程度上减少确诊延迟<sup>[24]</sup>。本研究还发现确诊单位在结核病防治所是确诊延迟的危险因素, 分析其原因可能与综合医院未第一时间考虑肺结核而导致患者多次筛查和转诊最终才前往结核病防治所进行确诊有关<sup>[25]</sup>。

本研究结果显示, 多因素 logistic 回归模型和决策树模型性能相当, 均能对肺结核患者就诊延迟及确诊延迟的影响因素进行良好的分析, 但 2 种模型检验侧重点不同, 多因素 logistic 回归模型能通过 OR 值及其 95%CI 展示变量与延迟的相关性, 量化研究的影响因素。而决策树模型显示了变量间多种协同或叠加作用模式, 模型各节点因素可作为识别延迟高危人群的指标, 对于节点因素的干预有利于相应类别人群发生肺结核延迟的防控。2 种模型联合使用, 可以进行更为全面的分析, 从而增加解释和预测的有效性<sup>[26]</sup>。

本研究存在局限性。仅在北京市 2 个区进行,

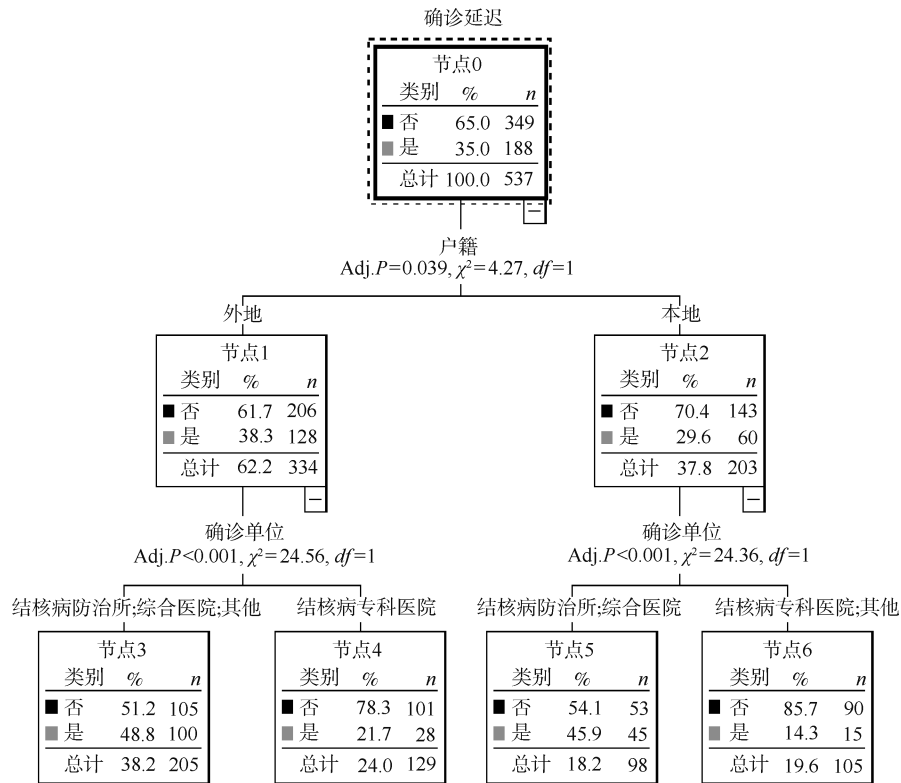


图2 肺结核患者确诊延迟影响因素的决策树模型

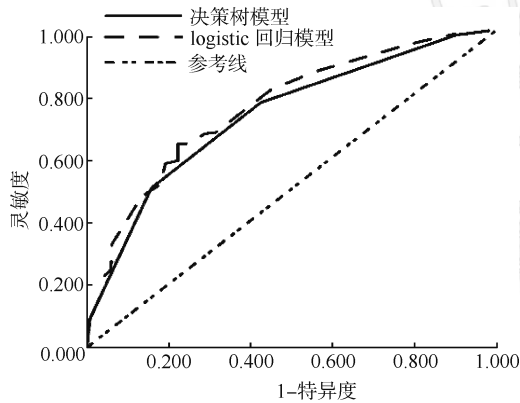


图3 肺结核患者就诊延迟多因素 logistic 回归模型和决策树模型的受试者工作特征曲线

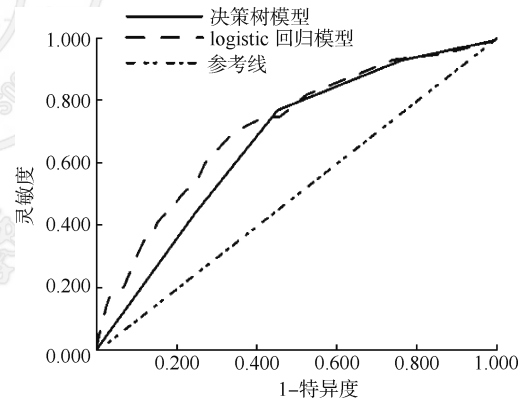


图4 肺结核患者确诊延迟多因素 logistic 回归模型和决策树模型的受试者工作特征曲线

虽然对2个区大部分登记在册的患者进行了调查，但是并不能代表全市的水平；部分患者可能存在回

忆偏倚。

综上所述，2021年北京市通州区及昌平区肺

表5 肺结核患者就诊延迟影响因素的模型性能比较

模型	AUC	95%CI	$s_x$	P值	灵敏度	特异度	约登指数
多因素 logistic 回归模型	0.756	0.715~0.797	0.021	<0.001	0.638	0.773	0.412
决策树模型	0.730	0.687~0.773	0.022	<0.001	0.781	0.562	0.344

注：AUC：曲线下面积

表6 肺结核患者确诊延迟影响因素的模型性能比较

模型	AUC	95%CI	$s_x$	P值	灵敏度	特异度	约登指数
多因素 logistic 回归模型	0.715	0.669~0.760	0.023	<0.001	0.713	0.636	0.349
决策树模型	0.671	0.625~0.718	0.024	<0.001	0.771	0.547	0.319

注：AUC：曲线下面积



结核患者就诊延迟及确诊延迟总体处于较低水平, 仍需采取针对性措施进一步改善。第一, 加强结核病的宣传教育力度, 提高公众对结核病危害、传播方式以及相关症状的知晓程度, 针对高风险人群实施个性化的宣传策略, 促进潜在肺结核患者的主动就诊意识, 减少就诊延迟现象; 第二, 提升综合性医疗机构、基层医疗卫生机构对结核病的警惕性, 加强培训, 推广新技术应用, 提高结核病早期鉴别能力; 第三, 采取措施简化结核病患者的就医路径, 减少转诊流程, 提高结核病诊断及鉴别诊断效率。

**利益冲突** 所有作者声明无利益冲突

**作者贡献声明** 马晓格: 研究设计、数据整理、统计分析、论文撰写; 张立杰: 研究设计/指导、组织实施、数据收集; 高汉青、吴思慧、刘梦晗: 数据整理、技术支持; 包城、吴越: 数据收集、技术支持; 刘宇红: 研究设计、组织实施、数据收集、技术支持、论文审阅; 李亮: 研究设计/指导、组织实施、论文审阅

### 参 考 文 献

- World Health Organization. Global tuberculosis report 2023-1. 1 TB disease burden[EB/OL]. (2023-11-07) [2024-02-22]. <https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2023/tb-disease-burden>.
- Marks GB, Nguyen NV, Nguyen PTB, et al. Community-wide screening for tuberculosis in a high-prevalence setting[J]. *N Engl J Med*, 2019, 381(14):1347-1357. DOI: 10.1056/NEJMoa1902129.
- 全国第五次结核病流行病学抽样调查技术指导组, 全国第五次结核病流行病学抽样调查办公室. 2010 年全国第五次结核病流行病学抽样调查报告[J]. *中国防痨杂志*, 2012, 34(8):485-508.  
Technical Guidance Group of the Fifth National TB Epidemiological Survey, The Office of the Fifth National TB Epidemiological Survey. The fifth national tuberculosis epidemiological survey in 2010[J]. *Chin J Antituberc*, 2012, 34(8):485-508.
- 赵钰雪, 朱慧, 宗屹, 等. 基于 logistic 回归和决策树的癌症患者及照顾者家庭功能影响因素分析[J]. *齐鲁护理杂志*, 2022, 28(11):23-30. DOI:10.3969/j.issn.1006-7256.2022.11.007.  
Zhao YX, Zhu H, Zong Y, et al. Analysis on family function influencing factors based on logistic regression and decision tree in cancer patients and caregivers[J]. *J Qilu Nurs*, 2022, 28(11):23-30. DOI:10.3969/j.issn.1006-7256.2022.11.007.
- 陈新. 数据挖掘分类方法中的决策树算法探究[J]. *电脑知识与技术*, 2013(35):7904-7905, 7910.  
Chen X. Data mining classification decision tree algorithms to explore[J]. *Comput Knowl Technol*, 2013(35):7904-7905, 7910.
- Al Fryan LH, Shomo MI, Alazzam MB, et al. Retracted: processing decision tree data using internet of things (IoT) and artificial intelligence technologies with special reference to medical application[J]. *BioMed Res Int*, 2022, 2022:8626234. DOI:10.1155/2022/8626234.
- Teo AKJ, Singh SR, Prem K, et al. Duration and determinants of delayed tuberculosis diagnosis and treatment in high-burden countries: a mixed-methods systematic review and meta-analysis[J]. *Respir Res*, 2021, 22(1):251. DOI:10.1186/s12931-021-01841-6.
- Chakma B, Gomes D, Filipe PA, et al. A temporal analysis on patient and health service delays in pulmonary tuberculosis in Portugal: inter and intra-regional differences and in(equalities) between gender and age[J]. *BMC Public Health*, 2022, 22(1): 1830. DOI: 10.1186/s12889-022-14216-3.
- 中国疾病预防控制中心结核病预防控制中心. 中国结核病防治工作技术指南[EB/OL]. (2021-06-13) [2024-02-22]. <https://tb.chinacdc.cn/xxjlg/202111/W020211119672904030470.pdf>.
- 中国防痨协会结核病控制专业分会, 中国防痨协会老年结核病防治专业分会, 《中国防痨杂志》编辑委员会. 中国社区肺结核主动筛查循证指南[J]. *中国防痨杂志*, 2022, 44(10): 987-997. DOI: 10.19982/j.issn.1000-6621.20220321.  
Tuberculosis Control Branch of Chinese Antituberculosis Association, Elderly Tuberculosis Control Branch of Chinese Antituberculosis Association, Editorial Board of Chinese Journal of Antituberculosis. Evidence-based guidelines for active screening of pulmonary tuberculosis in Chinese communities[J]. *Chin J Antituberc*, 2022, 44(10): 987-997. DOI: 10.19982/j.issn.1000-6621.20220321.
- 韩阳, 牛孝彬, 彭爱玲, 等. 2017-2021 年安徽省淮北市结核患者就诊延迟情况及影响因素[J]. *上海预防医学*, 2023, 35(8):758-763. DOI:10.19428/j.cnki.sjpm.2023.22860.  
Han Y, Niu XB, Peng AL, et al. Delayed medical visit among tuberculosis patients and risk factors in HuaiBei City, Anhui Province from 2017 to 2021[J]. *Shanghai J Prev Med*, 2023, 35(8): 758-763. DOI: 10.19428/j.cnki.sjpm.2023.22860.
- 郭净, 刘忠达, 丰银平. 2011-2021 年浙江省丽水市 2268 例肺结核患者就诊延迟影响因素分析[C]//中国防痨协会 2023 年第 34 届全国学术大会暨结核病防治新技术推广应用论坛论文汇编. 厦门: 中国防痨协会, 重庆智飞生物制品股份有限公司, 厦门致善生物科技股份有限公司, 2023: 4. DOI:10.26914/c.cnkihy.2023.009958.  
Guo J, Liu ZD, Feng YP. Analysis on influencing factors of delay in seeing a doctor in 2268 patients with pulmonary tuberculosis in Lishui City, Zhejiang Province from 2011 to 2021[C]//A compilation of papers from the 34th national academic conference of china tuberculosis association in 2023 and the forum on the promotion and application of new technologies for tuberculosis diagnosis and prevention. Xiamen: China National Tuberculosis Association, Chongqing Zhifei Biological Products Co., LTD., Xiamen Zhishan Biotechnology Co., LTD., 2023:4. DOI:10.26914/c.cnkihy.2023.009958.
- Zhu SJ, Wang YL, Chu TC, et al. Patient delay of patients with pulmonary tuberculosis in Yantai, China, 2013 to 2022: Analysis of a case series[J]. *Medicine (Baltimore)*, 2024, 103(9): e37188. DOI: 10.1097/MD.00000000000037188.
- de Moraes M, Sousa S, Marques J, et al. Investigating the role of symptom valorisation in tuberculosis patient delay in urban areas in Portugal[J]. *BMC Public Health*, 2023, 23(1):2421. DOI:10.1186/s12889-023-17319-7.
- Otero L, Boldoo T, Purevdagva A, et al. Delays in health seeking, diagnosis and treatment for tuberculosis patients in Mongolia: an analysis of surveillance data, 2018-2021[J]. *Western Pac Surveill Response J*, 2024, 15(1):1-9. DOI:10.5365/wpsar.2024.15.1.1074.
- Istamov K, Beglaryan M, Goncharova O, et al. Delays in treatment initiation and treatment outcomes in patients with tuberculosis in the Kyrgyz republic: are there differences between migrants and non-migrants? [J]. *Trop Med Infect Dis*, 2023, 8(8):412. DOI:10.3390/tropicalmed8080412.



[17] Butson A, Suggaravetsiri P, Tesana N. Delay of treatment among new smear-positive pulmonary tuberculosis patients in Thai-Cambodia border: cases study in Surin and Sisaket province, Thailand[J]. Res J Med Sci, 2010, 4(6):340-345.

[18] Kunjok DM, Mwangi JG, Mambo S, et al. Assessment of delayed tuberculosis diagnosis preceding diagnostic confirmation among tuberculosis patients attending Isiolo County level four hospital, Kenya[J]. Pan Afr Med J, 2021, 38:51. DOI:10.11604/pamj.2021.38.51.21508.

[19] Vigneswaran N, Parnis R, Lowbridge C, et al. Factors leading to diagnostic delay in tuberculosis in the tropical north of Australia[J/OL]. Intern Med J, 2023. DOI: 10.1111/imj.16223.

[20] 徐静, 贺晓新, 周虹九, 等. 2011-2020 年北京某市级定点医院肺结核患者特征分析[J]. 首都公共卫生, 2022, 16(3):138-141. DOI:10.16760/j.cnki.sdggws.2022.03.001. Xu J, He XX, Zhou HJ, et al. Characteristics of pulmonary tuberculosis patients in a municipal designated medical institution of Beijing, 2011-2020[J]. Cap J Public Health, 2022, 16(3):138-141. DOI:10.16760/j.cnki.sdggws.2022.03.001.

[21] Mhimbira FA, Cuevas LE, Dacombe R, et al. Interventions to increase tuberculosis case detection at primary healthcare or community-level services[J]. Cochrane Database Syst Rev, 2017, 11(11): CD011432. DOI: 10.1002/14651858.CD011432.pub2.

[22] World Health Organization. WHO operational handbook on tuberculosis: Module 2: Screening - Systematic screening for tuberculosis disease[EB/OL]. (2023-04-22) [2024-02-22]. <https://www.who.int/publications/i/item/9789240022614>.

[23] Sabawoon W, Sato H, Kobayashi Y. Delay in the treatment of pulmonary tuberculosis: a report from Afghanistan[J]. Environ Health Prev Med, 2012, 17(1): 53-61. DOI: 10.1007/s12199-011-0219-9.

[24] Zhang Q, Zhang Q, Sun BQ, et al. GeneXpert MTB/RIF for rapid diagnosis and rifampin resistance detection of endobronchial tuberculosis[J]. Respirology, 2018, 23(10): 950-955. DOI:10.1111/resp.13316.

[25] 梁达, 商越, 王兆芬, 等. 青海省 2014-2019 年肺结核患者就诊延迟情况及影响因素[J]. 中华疾病控制杂志, 2021, 25(7):763-767. DOI:10.16462/j.cnki.zhjbkz.2021.07.005. Liang D, Shang Y, Wang ZF, et al. Analysis of the health-care seeking delay and influencing factors among pulmonary tuberculosis patients in Qinghai Province, 2014-2019[J]. Chin J Dis Control Prev, 2021, 25(7): 763-767. DOI:10.16462/j.cnki.zhjbkz.2021.07.005.

[26] 张燕, 彭伶俐, 张磊, 等. 基于 logistic 回归和决策树法预测颅脑肿瘤患者术后病情恶化风险[J]. 护理学杂志, 2022, 37(7):15-19. DOI:10.3870/j.issn.1001-4152.2022.07.015. Zhang Y, Peng LL, Zhang L, et al. Risk prediction of postoperative deterioration for patients with craniocerebral tumors based on logistic regression and decision tree method[J]. J Nurs Sci, 2022, 37(7): 15-19. DOI:10.3870/j.issn.1001-4152.2022.07.015.

## 中华流行病学杂志第八届编辑委员会组成人员名单

(按姓氏汉语拼音排序)

顾问	高福	顾东风	贺雄	姜庆五	陆林	乔友林
	饶克勤	汪华	徐建国			
名誉总编辑	郑锡文					
总编辑	李立明					
副总编辑	邓瑛	冯子健	何纳	何耀	卢金星	沈洪兵
	谭红专	吴尊友	杨维中	詹思延		

### 编辑委员(含总编辑、副总编辑)

安志杰	白亚娜	毕振强	曹广文	曹卫华	曹务春	陈坤	陈可欣
陈万青	陈维清	代敏	戴江红	党少农	邓瑛	丁淑军	段广才
段蕾蕾	方利文	方向华	冯子健	龚向东	何纳	何耀	何剑峰
胡东生	胡永华	胡志斌	贾崇奇	江宇	阚飙	阚海东	李琦
李群	李敬云	李立明	李秀央	李亚斐	李中杰	林鹏	刘静
刘民	刘玮	刘殿武	卢金星	栾荣生	罗会明	吕繁	吕筠
吕嘉春	马军	马伟	马家奇	马文军	毛琛	孟蕾	米杰
缪小平	潘凯枫	潘晓红	彭晓霞	邱洪斌	任涛	单广良	邵中军
邵祝军	沈洪兵	施小明	时景璞	宋志忠	苏虹	孙业桓	谭红专
唐金陵	陶芳标	汪宁	王蓓	王岚	王丽	王璐	王金桃
王丽敏	王全意	王素萍	王伟炳	王增武	王长军	王子军	魏文强
吴凡	吴静	吴涛	吴先萍	吴尊友	武鸣	项永兵	徐飏
徐爱强	许汴利	许国章	闫永平	杨维中	么鸿雁	叶冬青	于普林
余宏杰	俞敏	詹思延	张建中	张顺祥	张卫东	张作风	赵方辉
赵根明	赵文华	赵亚双	周脉耕	朱凤才	庄贵华		