

DOI: 10. 12138/j. issn. 1671-9638. 20211135

· 论 著 ·

## 基于 SMOTE 算法的化疗肿瘤患者下呼吸道感染预警模型构建

王梅英<sup>1</sup>, 杨敏<sup>2</sup>, 刘佳微<sup>1</sup>, 张慧琳<sup>3</sup>

(1. 青海省人民医院感染管理科, 青海 西宁 810007; 2. 青海省人民医院肿瘤内科, 青海 西宁 810007; 3. 青海大学附属医院感染科, 青海 西宁 810007)

**[摘要]** **目的** 构建基于少数类样本合成过抽样技术(synthetic minority over-sampling technique, SMOTE)算法的化学治疗(化疗)肿瘤患者下呼吸道感染预警模型。**方法** 共纳入西宁市 4 所三级医院 2019 年 1 月—2021 年 6 月收治的 2 384 例接受化疗的肿瘤患者为研究对象, 将所收集病例按照 7:3 的比例随机分为建模组 1 668 例和验证组 716 例, 建模组数据用来建立模型, 验证组数据对所建立的模型进行验证, 利用单因素比较和 logistic 回归分析筛选下呼吸道感染影响因素, 基于 SMOTE 算法建立化疗肿瘤患者下呼吸道感染预警模型。**结果** logistic 回归分析可得, 年龄( $x_1$ )、身体质量指数(BMI)值是否正常( $x_2$ )、恶性肿瘤分期( $x_3$ )、吸烟史( $x_4$ )、合并糖尿病( $x_5$ )、合并肺部疾病( $x_6$ )均是化疗肿瘤患者下呼吸道感染的危险因素(均  $P < 0.01$ ), 获得原始数据预警模型:  $\text{Logit}(P) = 0.055x_1 + 0.967x_2 - 0.195x_3 + 1.383x_4 + 0.968x_5 + 0.939x_6 - 14.073$  和基于 SMOTE 算法的预警模型:  $\text{Logit}(P) = 0.090x_1 + 1.092x_2 - 0.249x_3 + 1.724x_4 + 1.136x_5 + 1.344x_6 - 14.859$ 。基于 SMOTE 算法预警模型 AUC 为 0.949 (95% CI: 0.937~0.961), 高于原始数据预警模型 AUC 0.780 (95% CI: 0.734~0.846)。**结论** 基于 SMOTE 算法所构建的预警模型能更准确预警化疗肿瘤患者下呼吸道感染, 有效解决感染与非感染患者样本数据不平衡所导致的预测误差, 基于预测模型可选择相应的对策进行应对。

**[关键词]** SMOTE 算法; 化疗; 肿瘤; 下呼吸道感染; 预警模型

**[中图分类号]** R181.3<sup>+</sup>2

## Construction of early warning model of lower respiratory tract infection in chemotherapy tumor patients based on SMOTE algorithm

WANG Mei-ying<sup>1</sup>, YANG Min<sup>2</sup>, LIU Jia-wei<sup>1</sup>, ZHANG Hui-lin<sup>3</sup> (1. Department of Infection Management, Qinghai Provincial People's Hospital, Xining 810007, China; 2. Department of Oncology, Qinghai Provincial People's Hospital, Xining 810007, China; 3. Department of Infection, Affiliated Hospital of Qinghai University, Xining 810007, China)

**[Abstract]** **Objective** To construct the early warning model of lower respiratory tract (LRT) infection in chemotherapy tumor patients based on synthetic minority over-sampling technique (SMOTE) algorithm. **Methods** 2 384 tumor patients treated with chemotherapy in 4 tertiary hospitals in Xining City from January 2019 to June 2021 were investigated, patients were randomly divided into modeling group ( $n = 1 668$ ) and validation group ( $n = 716$ ) according to the ratio of 7:3, data of modeling group was used to construct the model, data of validation group was used to verify the constructed model, influencing factors for LRT infection were screened by univariate comparison and logistic regression analysis, the early warning model of LRT infection of chemotherapy tumor patients was constructed based on SMOTE algorithm. **Results** Logistic regression analysis showed that age ( $x_1$ ), whether body mass index was normal (BMI,  $x_2$ ), stage of malignant tumor ( $x_3$ ), smoking history ( $x_4$ ), combined diabetes mellitus ( $x_5$ ) and combined pulmonary disease ( $x_6$ ) were all risk factors for LRT infection in chemotherapy tumor patients (all  $P <$

**[收稿日期]** 2021-07-05

**[基金项目]** 青海省卫生健康系统适宜推广技术项目(2019-wjtg-01); 昆仑英才高端领军创新创业人才(青人才字[2020]10号)

**[作者简介]** 王梅英(1973-), 女(汉族), 青海省西宁市人, 副主任护师, 主要从事医院感染管理研究。

**[通信作者]** 王梅英 E-mail: 2991365603@qq.com

0.01), the original data warning model :  $\text{Logit}(P) = 0.055x_1 + 0.967x_2 - 0.195x_3 + 1.383x_4 + 0.968x_5 + 0.939x_6 - 14.073$  and early warning model based on SMOTE algorithm:  $\text{Logit}(P) = 0.090x_1 + 1.092x_2 - 0.249x_3 + 1.724x_4 + 1.136x_5 + 1.344x_6 - 14.859$  were obtained. The AUC of early warning model based on SMOTE algorithm was higher than original data warning model (0.949[95%CI:0.937-0.961] vs 0.780 [95%CI:0.734-0.846]).

**Conclusion** The early warning model based on SMOTE algorithm can more accurately warn LRT infection in chemotherapy tumor patients, and effectively solve the warning error caused by the imbalance of the sample data of infected and non-infected patients, the corresponding countermeasures can be selected based on the warning model.

**[Key words]** SMOTE algorithm; chemotherapy; tumor; lower respiratory tract infection; early warning model

临床上,化学治疗(化疗)作为治疗肿瘤的重要手段之一,能有效遏制癌症发展,提高患者生存质量<sup>[1]</sup>,但大多数化疗肿瘤患者容易出现下呼吸道感染,影响患者康复及肺功能恢复<sup>[2]</sup>,严重感染可能导致患者死亡。因此,建立感染预警模型对预防化疗肿瘤患者下呼吸道感染具有重要意义。有研究<sup>[3-4]</sup>指出,化疗肿瘤患者下呼吸道感染主要受到年龄、合并基础疾病、营养状况和使用抗菌药物影响,由于其呼吸道感染发病率低于 8%~13%,造成感染和非感染病例数据之间的不平衡。传统统计分析方法对不平衡数据处理效果欠佳<sup>[5]</sup>,对多数类预测效果高,而对少数类预测效果较差,不适合喉癌患者术后下呼吸道感染预警模型构建。为解决数量不平衡数据分类问题,有研究者引入重采样技术重建数据平衡少数类与多数类的数据数量。少数类样本合成过抽样技术(synthetic minority over-sampling technique, SMOTE)由 Chawla 等<sup>[6]</sup>于 2011 年提出,其根据少数类属性特征随机产生相似样本,有效平衡少数类与多数类之间数量,减少数据集过度倾斜。相关研究已经通过 SMOTE 算法成功建立脑损伤继发精神障碍<sup>[7]</sup>、绝经后骨质疏松性骨折<sup>[8]</sup>和癫痫发作<sup>[9]</sup>等少数类预警模型。本研究调查 2019 年 1 月—2021 年 6 月西宁市 4 所三级医院收治的接受化疗的肿瘤患者临床资料,构建基于 SMOTE 算法的感染预警模型,为预防化疗肿瘤患者下呼吸道感染提供指导。

## 1 对象与方法

**1.1 临床资料** 本文选取 2019 年 1 月—2021 年 6 月西宁市 4 所三级医院收治的接受化疗的肿瘤患者临床资料,数据纳入标准:①经病理学和影像诊断确诊为恶性肿瘤;②年龄 $\geq 18$ 岁,临床资料完整;③行肿瘤化疗;④患者对本研究知情同意。排除标准:①合并呼吸系统疾病者,先天性免疫系统缺陷和未接受化疗的恶性肿瘤患者;②转入重症监护病房(ICU)的

昏迷、病危或病亡者;③转院或临床资料不完整者。

**1.2 下呼吸道感染诊断** 根据相关资料确定下呼吸道感染诊断标准<sup>[10]</sup>,包括:①胸部 X 线检出支气管或肺部炎症改变;②静脉血检出白细胞数和(或)嗜中性粒细胞比例上升;③下呼吸道分泌物培养检出病原菌,肿瘤患者化疗开始后出现同时满足①②、①③或①②③项均可判断为下呼吸道感染。化疗期间,由其主管医生和研究者每周通过微信或电话询问患者是否出现剧烈咳嗽、多痰、发热、胸痛、打喷嚏、流鼻涕、咽痛、低热、喘、呼吸困难等症状,若出现以上症状由主管医生、医院感染专家和研究者共同确诊是否为下呼吸道感染。

## 1.3 研究方法

**1.3.1 数据分组** 将所收集病例(2 384 例)按照 7:3 的比例随机分为建模组 1 668 例和验证组 716 例(随机种子为 20190820),建模组数据用来建立模型,验证组数据对所建立的模型进行验证。将建模组根据是否发生下呼吸道感染分为感染组 99 例和非下呼吸道感染组 1 569 例(包括非感染患者和其他部位感染患者),基于 SMOTE 算法建立模型,具体实施步骤为:设少数类样本数 = 下呼吸道感染组,增加样本数  $n = 1\ 569/99 \approx 16$ ,最近邻点数  $k = 16$ ,实施算法处理步骤如下<sup>[9]</sup>:第 1 步,计算数据集内各少数类样本的  $k$  个最近邻点,近邻选择为距离或相似系数;第 2 步,随机取少数类样本点  $i$  的  $k$  个最近邻点中 1 个样本  $j$ ;第 3 步,计算样本  $i$  和样本  $j$  全部属性的差值  $Q$ ;第 4 步,产生一个区间为  $[0,1]$  的随机数  $R$ ;第 5 步,产生少数类的合成样本:  $\text{Sample}_{\text{new}} = \text{Sample}_i + R \times Q$ ;第 6 步,重复第 1 步到第 5 步,直到少数类样本  $i$  增加  $n$  倍并达到要求后停止;第 7 步,重复第 1 步到第 6 步,直到  $B$  个少数类样本全部被处理完后停止。通过对类样本点与其最近邻样本点间进行随机插值,得实际抽为非下呼吸道感染组 = 1 569 例,感染组 = 1 584 例,下呼吸道感染组和非下呼吸道感染组比例近 1.01。

1.3.2 资料收集内容 使用自制量表收集患者基线资料(姓名、性别、年龄、学历、居住地、婚姻状态),生理状况[身体质量指数(BMI)值是否正常、吸烟史、饮酒史、有无合并高血压、糖尿病、肺部疾病],恶性肿瘤病情(恶性肿瘤分期、是否远处转移、是否接受肿瘤根治术),化疗情况(化疗时间),发生下呼吸道感染情况(是否发生感染、感染病原体类型)。

1.4 模型构建 应用 SPSS 25.0 和 R 3.5.2 进行模型构建,计数资料选择例数或百分比表示,组间比较选择卡方检验或 Fisher's 确切概率法检验,在建模组中,将单因素分析有意义的变量( $P < 0.05$ )纳入多因素 logistic 回归中分析下呼吸道感染的危险因素,选择 SMOTE 算法进行过抽样插值后构建下呼吸道感染预警模型, $P \leq 0.05$  为差异具有统计学意义。

1.5 模型验证 利用构建的风险预测评分模型对建模组和验证组患者进行评分,以患者是否发生下呼吸道感染为状态变量,以各患者的总风险评分值为检验变量,绘制受试者工作特征曲线(ROC 曲线)评价模型的区分度,采用 Hosmer-Lemeshow(H-L)检验评价模型的校准度,使用真正类率(TPR)、阳性预测值(PPV)、F 分数(F-score)和 ROC 曲线下面积(AUC)验证预测模型对验证组数据预测准确性,用以评估模型的临床应用价值。

## 2 结果

2.1 调查对象的基本特征 共收集 2019 年 1 月—

2021 年 6 月 2 488 例行化疗的肿瘤患者资料,依据纳入排除标准,剔除病历缺失或不完整病例,得到有效数据 2 384 例,其中发生下呼吸道感染者 152 例(6.38%),152 例下呼吸道感染患者未合并其他部位感染,共检出菌株 256 株,革兰阴性菌 176 株(68.75%),包括铜绿假单胞菌 96 株(37.50%)、肺炎克雷伯菌 40 株(15.63%)、大肠埃希菌 16 株(6.25%)、阴沟肠杆菌 12 株(4.69%)、鲍曼不动杆菌 12 株(4.69%);革兰阳性菌 80 株(31.25%),包括金黄色葡萄球菌 52 株(20.31%)和凝固酶阴性葡萄球菌 28 株(10.94%)。将所收集病例按照 7:3 的比例随机分为建模组 1 668 例和验证组 716 例(随机种子为 20190820),将建模组根据是否发生下呼吸道感染分为下呼吸道感染组 99 例(5.94%)和非下呼吸道感染组 1 569 例(94.06%),99 例感染患者检出菌株 162 株,革兰阴性菌 108 株(66.67%),包括铜绿假单胞菌 61 株(37.65%)、肺炎克雷伯菌 22 株(13.58%)、大肠埃希菌 11 株(6.79%)、阴沟肠杆菌 7 株(4.32%)、鲍曼不动杆菌 7 株(4.32%);革兰阳性菌 54 株(33.33%),包括金黄色葡萄球菌 38 株(23.45%)和凝固酶阴性葡萄球菌 16 株(9.88%)。

2.2 化疗肿瘤患者下呼吸道感染单因素分析 单因素比较,两组化疗肿瘤患者不同年龄、BMI 值是否正常、恶性肿瘤分期、吸烟史、合并糖尿病、合并肺部疾病、化疗时间比较,差异均有统计学意义(均  $P < 0.05$ )。见表 1。

表 1 建模组化疗肿瘤患者下呼吸道感染单因素分析[例(%)]

Table 1 Univariate analysis on LRT infection in chemotherapy tumor patients in modeling group (No. of cases[%])

项目	例数( $n = 1\ 668$ )	下呼吸道感染组( $n = 99$ )	非下呼吸道感染组( $n = 1\ 569$ )	$\chi^2$	$P$	
性别	男	912	52(5.70)	0.197	0.658	
	女	756	47(6.22)			709(93.78)
年龄(岁)	$\leq 45$	372	16(4.30)	19.275	$< 0.001$	
	46~60	516	23(4.46)			493(95.54)
	61~75	513	29(5.65)			484(94.35)
	$> 75$	267	31(11.61)			236(88.39)
学历	小学	434	26(5.99)	0.587	0.746	
	中学	742	47(6.33)			695(93.67)
	大学及以上	492	26(5.28)			466(94.72)

续表 1 (Table 1, Continued)

项目		例数( $n = 1\ 668$ )	下呼吸道感染组 ( $n = 99$ )	非下呼吸道感染组 ( $n = 1\ 569$ )	$\chi^2$	$P$
BMI 值是否正常	是	694	31(4.47)	663(95.53)	4.590	0.032
	否	974	68(6.98)	906(93.02)		
婚姻状态	未婚	159	13(8.18)	146(91.82)	-0.070	0.944
	已婚	1 138	64(5.62)	1 074(94.38)		
	离异	271	12(4.43)	259(95.57)		
	丧偶	100	10(10.00)	90(90.00)		
居住地	城市	827	57(6.89)	770(93.11)	2.691	0.101
	农村	841	42(4.99)	799(95.01)		
恶性肿瘤分期	II 期	335	18(5.37)	317(94.63)	7.632	0.022
	III 期	751	34(4.53)	717(95.47)		
	IV 期	582	47(8.08)	535(91.92)		
是否远处转移	是	336	18(5.36)	318(94.64)	0.252	0.616
	否	1 332	81(6.08)	1 251(93.92)		
是否接受肿瘤根治术	是	1 254	77(6.14)	1 177(93.86)	0.381	0.537
	否	414	22(5.31)	392(94.69)		
吸烟史	有	653	57(8.73)	596(91.27)	15.002	<0.001
	无	1 015	42(4.14)	973(95.86)		
饮酒史	有	1 039	58(5.58)	981(94.42)	0.615	0.433
	无	629	41(6.52)	588(93.48)		
高血压	有	744	52(6.99)	692(93.01)	2.672	0.102
	无	924	47(5.09)	877(94.91)		
糖尿病	有	468	63(13.46)	405(86.54)	66.002	<0.001
	无	1 200	36(3.00)	1 164(97.00)		
肺部疾病	有	433	65(15.01)	368(84.99)	9.289	<0.001
	无	1 235	34(2.75)	1 201(97.25)		
化疗时间(周)	<2	395	22(5.57)	373(94.43)	7.528	0.023
	2~3	733	33(4.50)	700(95.50)		
	4~6	540	44(8.15)	496(91.85)		

注: BMI 值正常范围选择中国标准( $18.5 \leq \text{BMI} < 24$ )。

### 2.3 化疗肿瘤患者下呼吸道感染的多因素分析

以是否有下呼吸道感染为因变量(否 = 0, 是 = 1), 选择单因素分析有统计学意义的变量为自变量引入多因素 logistic 回归分析, 设置  $a_{\lambda} = 0.05$ ,  $a_{\text{出}} = 0.10$ , 可获得年龄、BMI 值是否正常、恶性肿瘤分期、吸烟史、合并糖尿病、合并肺部疾病均是化疗肿瘤患者下呼吸道感染的独立危险因素(均  $P < 0.05$ ), 见表 2。

### 2.4 基于 SMOTE 算法的化疗肿瘤患者下呼吸道感染多因素分析

选择多因素 logistic 回归分析结果获得化疗肿瘤患者下呼吸道感染的影响因素, 利用 SMOTE 算法对感染组原样本进行 16 倍过抽样, 获得感染和非感染比例近似为 1(实际抽样为非

感染 1 569 例, 感染 1 584 例, 感染/非感染 = 1.01), 然后对过抽样后数据进行多因素 logistic 回归, 设置  $a_{\lambda} = 0.05$ ,  $a_{\text{出}} = 0.10$ , 见表 3。

### 2.5 基于 SMOTE 算法风险预测评分模型的构建

以化疗肿瘤患者是否下呼吸道感染为因变量  $y$  ( $y = 1$  为感染,  $y = 0$  为非感染), 年龄 ( $x_1$ )、BMI 值是否正常 ( $x_2$ )、恶性肿瘤分期 ( $x_3$ )、吸烟史 ( $x_4$ )、合并糖尿病 ( $x_5$ )、合并肺部疾病 ( $x_6$ ) 为自变量, 通过原始数据多因素 logistic 回归和基于 SMOTE 算法下重建数据多因素 logistic 回归分析, 可获得原始数据预警模型:  $\text{Logit}(P) = 0.055x_1 + 0.967x_2 - 0.195x_3 + 1.383x_4 + 0.968x_5 + 0.939x_6 - 14.073$  和基于

**表 2** 建模组化疗肿瘤患者下呼吸道感染 logistic 回归分析

**Table 2** Logistic regression analysis on LRT infection in chemotherapy tumor patients in modeling group

自变量	B	S. E	Wald $\chi^2$	P	OR	95%CI	
						下限	上限
常量	-14.073	3.806	13.674	0.000			
年龄( $\leq 45$ 岁 = 1; 46~60岁 = 2; 61~75岁 = 3; $> 75$ 岁 = 4)	0.055	0.027	4.079	0.043	1.057	1.002	1.114
BMI 值是否正常(否 = 0; 是 = 1)	0.967	0.218	19.602	$< 0.001$	2.630	1.714	4.035
恶性肿瘤分期(II 期 = 1; III 期 = 2; IV 期 = 3)	-0.195	0.094	4.364	0.037	0.822	0.685	0.988
吸烟史(无 = 0, 有 = 1)	1.383	0.484	8.172	0.004	3.988	1.545	10.295
合并糖尿病(无 = 0, 有 = 1)	0.968	0.486	3.969	0.046	2.633	1.016	6.823
合并肺部疾病(无 = 0, 有 = 1)	0.939	0.468	4.024	0.045	2.556	1.022	6.396

**表 3** 建模组化疗肿瘤患者基于 SMOTE 算法的下呼吸道感染多因素分析

**Table 3** Multivariate analysis on LRT infection in chemotherapy tumor patients in modeling group based on SMOTE algorithm

自变量	B	S. E	Wald $\chi^2$	P	OR	95%CI	
						下限	上限
常量	-14.859	1.749	72.136	0.000			
年龄( $\leq 45$ 岁 = 1; 46~60岁 = 2; 61~75岁 = 3; $> 75$ 岁 = 4)	0.090	0.014	43.597	$< 0.001$	1.094	1.065	1.123
BMI 值是否正常(否 = 0; 是 = 1)	1.092	0.240	20.620	$< 0.001$	2.980	1.860	4.775
恶性肿瘤分期(II 期 = 1; III 期 = 2; IV 期 = 3)	-0.249	0.046	29.354	$< 0.001$	0.780	0.713	0.853
吸烟史(无 = 0, 有 = 1)	1.724	0.223	59.558	$< 0.001$	5.606	3.618	8.685
合并糖尿病(无 = 0, 有 = 1)	1.136	0.229	24.563	$< 0.001$	3.115	1.988	4.883
合并肺部疾病(无 = 0, 有 = 1)	1.344	0.208	41.626	$< 0.001$	3.833	2.548	5.764

SMOTE 算法的预警模型:  $\text{Logit}(P) = 0.090x_1 + 1.092x_2 - 0.249x_3 + 1.724x_4 + 1.136x_5 + 1.344x_6 - 14.859$ 。分别以 2 个模型的预警概率  $P$  为诊断变量,以化疗肿瘤患者是否感染为金标准,对模型预警效应进行 ROC 曲线分析,获得基于 SMOTE 算法的预警模型对原始数据的 AUC 为 0.949(95%CI: 0.937~0.961),高于原始数据预警模型的 AUC 0.780(95%CI: 0.734~0.846),基于 SMOTE 算法的预警模型 H-L 检验  $P$  值( $P = 0.594$ )也高于原始数据预警模型( $P = 0.301$ ),见表 4。

**表 4** 预警模型 ROC 曲线的预测效应比较

**Table 4** Comparison of warning efficacy of ROC curve of early warning model

模型	AUC	P	95%CI	P(H-L 检验)
原始数据预警模型	0.780	$< 0.001$	0.734~0.846	0.301
基于 SMOTE 算法的预警模型	0.949	$< 0.001$	0.937~0.961	0.594

利用构建化疗患者下呼吸道风险预测评分模型对验证组患者进行验证,获得原始数据预警模型 AUC 为 0.757(95%CI: 0.708~0.805),基于 SMOTE 算法的预警模型 AUC 为 0.908(95%CI: 0.855~0.961),基于 SMOTE 算法的预警模型对验证组患者预测 TPR 值低于原始数据预警模型,而 PPV、F-score 均高于原始数据预警模型,见表 5、图 1~2。

**表 5** 预警模型对验证组患者预测概率验证

**Table 5** Validation of warning probability of patients in validation group by early warning mode

模型	AUC	95%CI	TPR	PPV	F-score
原始数据预警模型	0.757	0.708~0.805	0.722	0.601	0.654
基于 SMOTE 算法的预警模型	0.908	0.855~0.961	0.709	0.799	0.751

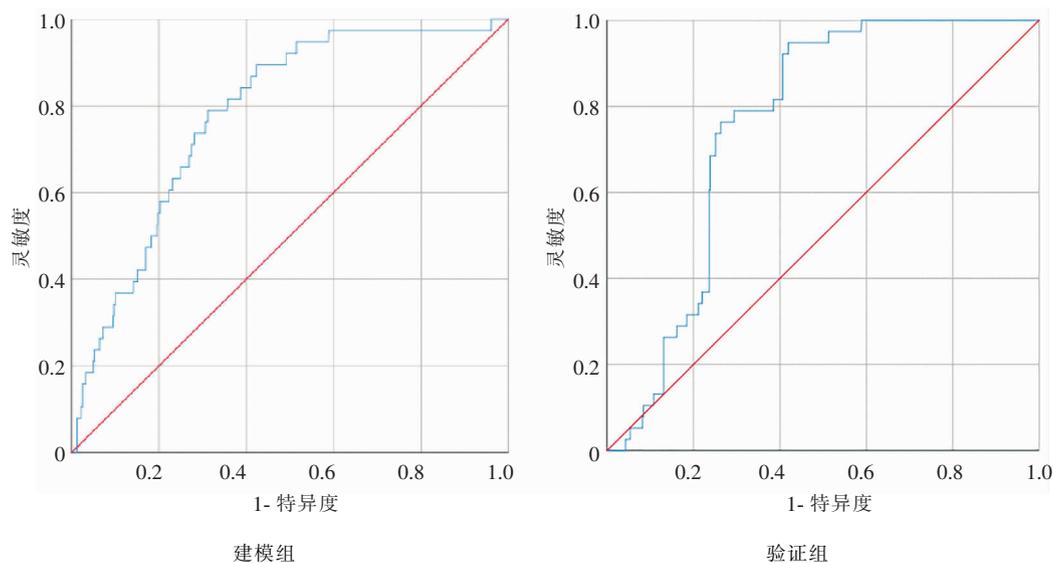


图 1 原始数据预警模型预警概率的 ROC 曲线

Figure 1 ROC curve of warning probability of original data early warning model

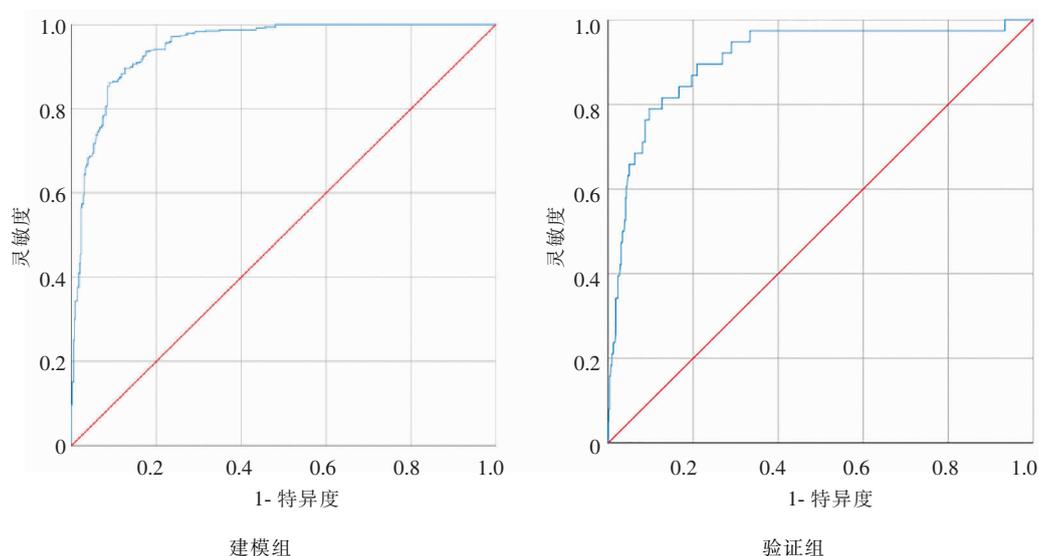


图 2 基于 SMOTE 算法预警模型预警概率的 ROC 曲线

Figure 2 ROC curve of early warning probability of early warning model based on SMOTE algorithm

### 3 讨论

本研究调查结果显示,2 384 例化疗肿瘤患者发生下呼吸道感染率为 6.38%,低于毛振宇等<sup>[11]</sup>研究中喉癌患者术后化疗期下呼吸道感染率 8.75%,表明该院对化疗肿瘤患者下呼吸道感染预防重视程度较高,使超过 50%的患者接受相应的抗菌药物治疗,有效降低下呼吸道感染率。152 例下呼吸道感染化疗肿瘤患者检出菌株 256 株,主要为革兰阴性菌 156 株,占 68.75%,与其他研究<sup>[12]</sup>结果相符。整

体检出菌类型集中于铜绿假单胞菌(37.50%)和金黄色葡萄球菌(20.31%),其能破坏化疗肿瘤患者机体,阻碍其代谢功能,从而降低其免疫功能,成为化疗肿瘤患者下呼吸道的主要感染菌种。

本研究结果表明,年龄、BMI 值是否正常、恶性肿瘤分期、吸烟史、合并肺部疾病、合并糖尿病均是化疗肿瘤患者下呼吸道感染的影响因素,与其他研究<sup>[13-15]</sup>结论相符,表现为高龄、BMI 值异常、有吸烟史、合并肺部疾病及糖尿病患者下呼吸道感染风险更高,这是因为:①化疗肿瘤患者年龄越大,其自身免疫功能下降程度和生理健康恶化程度越高,导致其身体抵

御病原体侵袭的能力下降,从而增加了其感染风险;

②若肿瘤患者 BMI 值过低,其身体营养状况不良程度越明显,从而自身免疫功能抗感染效率越被限制,研究<sup>[16-17]</sup>显示,过低 BMI 值甚至可降低恶性肿瘤患者生命质量,增加其病死率,而肿瘤患者 BMI 值过高,其存在肥胖及代谢功能疾病,合并糖尿病、冠心病及神经系统疾病概率较高,可造成免疫功能受到限制,增加下呼吸道感染率。

③患者恶性肿瘤分期越高,其身体受到肿瘤侵袭程度及损伤越重,导致其身体免疫功能越差,造成其下呼吸道感染风险越高。

④长期吸烟能损伤支气管黏膜,导致呼吸道纤毛变短、断裂和运动功能下降,降低支气管清除感染病原体的能力,且吸烟也能增加呼吸道内诱导痰内炎症因子表达,恶化感染程度。

⑤肺部疾病如慢性肺炎、支气管炎、慢性阻塞性肺疾病等可致呼吸系统组织及器官长期受到炎性环境的刺激,导致支气管黏膜腺体增生,刺激其分泌大量黏液,堵塞细支气管形成易感染菌定植环境,增加下呼吸道感染率。

⑥糖尿病患者本身就属于医院感染高危人群,其身体在强刺激源如伤害或恶性肿瘤作用下因激素调节和细胞因子分泌异常出现应激性高血糖,高糖环境能增高机体内反应性氧化物(ROS)表达,ROS 可诱导各种促炎因子表达上升,产生炎性级联反应诱发和恶化感染,因此合并糖尿病能进一步增加化疗肿瘤患者下呼吸道感染风险。因此,对有吸烟史、伴有慢性阻塞性肺疾病、慢性气管炎、慢性支气管炎和慢性肺炎等肺部疾病或炎症的肿瘤患者,监督其遵医嘱按时服药治疗肺部病症,术前指导其进行相应的呼吸功能训练,指导家属正确给患者翻身和叩背,鼓励患者进行深呼吸训练提升肺活量;整个化疗期为伴有高血压及糖尿病患者制定降血压及血糖控制的干预方案;指导家属为其准备富含蛋白质、维生素及低糖、低脂肪、低盐的食谱,给其补充足够营养,加快其损伤身体恢复;观察患者心理状态,视情况进行心理干预,降低其焦虑及抑郁心理。

本研究中化疗肿瘤患者发生下呼吸道感染者占 6.38%,非下呼吸道感染者占 93.62%,可获得感染和非下呼吸道感染患者数据数量存在明显不平衡。传统统计学预警模型对多数类如非感染人数预测效果高,而对少数类如感染人数预测效果较差<sup>[18]</sup>,是因为当统计数据高度不平衡时,筛查特异度能显著影响模型预测精度,导致其对阳性个体预测结果误差较高,因此,需要解决本研究中感染与非下呼吸道感染患者数量不平衡数据分类问题,本文引入过抽

样方法即 SMOTE 算法对原始数据进行重采样技术重建,获得感染和非感染比例近似为 1 新样本数据,结合多因素 logistic 回归分析结果,得到基于 SMOTE 算法的预警模型。以预警概率  $P$  为诊断变量和化疗肿瘤患者是否感染为金标准,对模型预警效应进行 ROC 曲线分析,结果显示 SMOTE 算法预警模型对原始数据预警效率 AUC 为 0.949 (95%CI:0.937~0.961),高于原始数据预警模型 AUC 面积 0.780 (95%CI:0.734~0.846),利用构建化疗患者下呼吸道风险预测评分模型对验证组患者进行验证,获得原始数据预警模型 AUC 为 0.757 (95%CI:0.708~0.805),基于 SMOTE 算法的预警模型 AUC 为 0.908 (95%CI:0.855~0.961),基于 SMOTE 算法的预警模型对验证组患者预测 TPR 值低于原始数据预警模型,而 PPV、F-score 均高于原始数据预警模型,说明使用 SMOTE 算法对本研究下呼吸道感染与非下呼吸道感染患者不平衡数据进行重建,使用重建数据拟合 logistic 预测模型,获得新预警模型预测准确率相对更高。

由本研究结果可知,化疗肿瘤患者发生下呼吸道感染受年龄、BMI 值是否正常、恶性肿瘤分期、吸烟史、是否合并肺部疾病、糖尿病的影响,结合上述影响因素基于 SMOTE 算法对数据进行重采样技术重建,获得的预警模型比传统 logistic 回归建立的模型预测效应更优。由于本研究因时间和环境所限,本研究仅收集了 4 所医院化疗肿瘤患者为研究对象,导致所选样本量选择较少,且未对其他不同感染部位患者的混杂因素进行比较分析,导致研究结果不完善,需要进一步扩大研究对象范围及纳入更多可能的影响因素,从而使研究结论更完善。

#### [参 考 文 献]

- [1] 滕兆霞,王志英,蔡珂. 消化系统肿瘤患者化疗后感染病原学分布及其影响因素分析[J]. 临床内科杂志, 2020, 37(2): 112-114.
- [2] 程鹏,郭莘,朱俊琳. 放化疗食管癌患者肺部感染的影响因素分析[J]. 中华医院感染学杂志, 2019, 29(16): 2436-2439.
- [3] 薛静,阳甜,陈明伟. 肺癌化疗患者下呼吸道感染病原菌分布及影响因素分析[J]. 现代肿瘤医学, 2017, 25(13): 2058-2061.
- [4] 李妍,史爱华. 急性白血病化疗后下呼吸道感染病原菌分布及耐药性分析[J]. 临床和实验医学杂志, 2018, 17(5): 548-550.
- [5] Blagus R, Lusa L. SMOTE for high-dimensional class-imbalanced data[J]. BMC Bioinformatics, 2013, 14: 106.

- [6] Chawla NV, Bowyer KW, Hall LO, et al. SMOTE: synthetic minority over-sampling technique[J]. J Artif Intell Res, 2002, 16: 321 - 357.
- [7] 孙秀彬, 辛涛, 薛付忠, 等. 基于 SMOTE 算法的颅脑损伤患者继发精神障碍预警模型[J]. 中国卫生统计, 2013, 30(6): 790 - 793.
- [8] 章轶立, 魏戎, 聂佩芸, 等. 基于 SMOTE 算法和决策树的绝经后骨质疏松性骨折分类模型建构[J]. 中国骨质疏松杂志, 2019, 25(1): 1 - 5.
- [9] 吴庆忠, 车峰远, 薛付忠. 基于非平衡数据的癫痫发作预警模型研究[J]. 山东大学学报(医学版), 2012, 50(2): 141 - 144, 148.
- [10] 胡必杰. 下呼吸道感染实验诊断规范[M]. 上海: 上海科学技术出版社, 2006: 45 - 46.
- [11] 毛振宇, 江涛. 喉癌术后并发下呼吸道感染的调查分析[J]. 中华全科医学, 2015, 13(8): 1382 - 1384.
- [12] 杨桂强, 刘文超, 刘滋润. 肺癌患者化疗后肺部感染病原菌及耐药分析[J]. 中国卫生检验杂志, 2017, 27(9): 1341 - 1343.
- [13] 孙瑶, 李乃坤, 武金凤, 等. 复发或转移性食管癌患者化疗后再应用紫杉类药物化疗引起肺炎的风险分析研究[J]. 医学与哲学(B), 2017, 38(6): 50 - 54.
- [14] 王海. 探讨化疗后肺癌患者合并下呼吸道感染的临床特点及影响因素[J]. 航空航天医学杂志, 2017, 28(7): 820 - 821.
- [15] 魏娟娟. 老年肺癌化疗前后肺部感染的病原菌培养及药敏分析[J]. 实用癌症杂志, 2017, 32(3): 466 - 468.
- [16] 刘联斌, 黄刚, 郭守俊, 等. 晚期肺癌患者化疗后肺部感染的病原菌分布和对免疫功能的影响研究[J]. 中华医院感染学杂志, 2017, 27(22): 5105 - 5108.
- [17] 陈卓, 贺亚宁, 陈琦, 等. 体重指数对乳腺癌新辅助化疗病理完全缓解及预后的影响[J]. 国际外科学杂志, 2019, 46(1): 20 - 25.
- [18] Gong HL, Zhou L, Wu HT, et al. Long-term clinical outcomes of supracricoid partial laryngectomy with cricohyoido-epiglottopexy for glottic carcinoma [J]. Acta Otolaryngol, 2019, 139(9): 803 - 809.

(本文编辑:陈玉华)

**本文引用格式:**王梅英, 杨敏, 刘佳微, 等. 基于 SMOTE 算法的化疗肿瘤患者下呼吸道感染预警模型构建[J]. 中国感染控制杂志, 2021, 20(12): 1094 - 1101. DOI: 10.12138/j.issn.1671-9638.20211135.

**Cite this article as:** WANG Mei-ying, YANG Min, LIU Jia-wei, et al. Construction of early warning model of lower respiratory tract infection in chemotherapy tumor patients based on SMOTE algorithm[J]. Chin J Infect Control, 2021, 20(12): 1094 - 1101. DOI: 10.12138/j.issn.1671-9638.20211135.