

• 医院药学 •

基于多元回归-灰色模型的非小细胞肺癌靶向药物用量预测分析

臧美彤, 吴春暖, 代爽, 张洁*

天津医科大学肿瘤医院 药学部 国家肿瘤临床医学研究中心 天津市“肿瘤防治”重点实验室 天津市恶性肿瘤临床医学研究中心, 天津 300060

摘要: **目的** 建立非小细胞肺癌(NSCLC)常用靶向药物的用量预测模型,为指导医疗机构抗肿瘤靶向药物的采购和库存管理提供数据支撑。**方法** 参考天津市肿瘤医院2019年各月的靶向药物用量数据,以4种临床常用的NSCLC靶向药物(吉非替尼、埃克替尼、奥希替尼和克唑替尼)为例,建立多元回归模型、GM(1,1)灰色模型以及多元回归-灰色组合模型,并对3种预测模型进行评价和验证。**结果** 多元回归模型在描述靶向药物用量的波动变化方面具有优势,灰色模型可以更好地描述靶向药物用量的增长趋势,而组合模型兼备描述用量变化趋势和短期波动的能力。吉非替尼、埃克替尼、奥希替尼和克唑替尼4种靶向药物运用组合模型得到的预测值与实际值误差分别为4.30%、2.87%、3.62%、4.42%。**结论** 多元回归-灰色组合模型运行良好,与单一模型相比表现出更高的精准度,可应用于医疗机构靶向药物的用量预测,从而实现靶向药物的精准化管理。

关键词: 非小细胞肺癌; 靶向药物; 多元回归模型; 灰色模型; 组合预测模型; 吉非替尼; 埃克替尼; 奥希替尼; 克唑替尼
中图分类号: R979.1 **文献标志码:** A **文章编号:** 1674-5515(2021)01-0165-05

DOI: 10.7501/j.issn.1674-5515.2021.01.034

Prediction of the usage of targeted drugs for non-small cell lung cancer based on multiple regression-gray model

ZANG Mei-tong, WU Chun-nuan, DAI Shuang, ZHANG Jie

Department of Pharmacy, National Clinical Research Center for Cancer, Tianjin Key Laboratory of "Cancer Prevention and Therapy",
Tianjin Clinical Research Center for Cancer, Tianjin Medical University Cancer Institute and Hospital, Tianjin 300060, China

Abstract: **Objective** To establish a prediction model of the commonly used targeted drugs for non-small cell lung cancer (NSCLC), providing data support for the procurement and management of targeted anti-tumor drugs in medical institutions. **Methods** The multiple regression model, gray model and the combined model were established with four commonly used targeted drugs for non-small cell lung cancer (gefitinib, icotinib, osimertinib and crizotinib), based on monthly usage of Tianjin Medical University Cancer Institute and Hospital in 2019, and the prediction models were evaluated and verified. **Results** The advantage of the multiple regression model is to describe the fluctuation of the usage of the targeted drugs. The gray model can reflect the increasing trend, while the combined model has the characteristics of describing the increasing trend and short-term fluctuations. The error between the predicted value and the actual value of the combined prediction model for gefitinib, icotinib, osimertinib and crizotinib was 4.30%, 2.87%, 3.62% and 4.42%, respectively. **Conclusion** Compared with the single model, the multiple regression-gray combined prediction model works better and shows higher accuracy. It can be applied to the prediction of the usage of targeted drugs in medical institutions, which is conducive to the accurate management for targeted drugs.

Key words: non-small cell lung cancer; targeted drugs; multiple regression model; gray model; combined prediction model; gefitinib; icotinib; osimertinib; crizotinib

收稿日期: 2020-08-13

基金项目: 国家自然科学基金青年基金资助项目(81703454)

作者简介: 臧美彤, 药师, 硕士, 研究方向为医院药学。Tel: (022)23340123 E-mail: tongtongzmt@126.com

*通信作者: 张洁, 副主任药师, 硕士, 研究方向为临床药理学及药事管理方向。Tel: (022)23340123 E-mail: jiezhang1224@163.com

肺癌是全球发病率和死亡率最高的恶性肿瘤,其中约 85% 为非小细胞肺癌 (NSCLC) [1]。近年来,随着 NSCLC 相关驱动基因被发现,NSCLC 治疗进入个体化、精准化的靶向治疗时代 [2-3]。2017 年国家人社部发布《关于将 36 种药品纳入国家基本医疗保险药品目录乙类范围的通知》,多种抗肿瘤靶向药物被纳入医保目录,使得靶向药物的使用量逐步攀升。靶向药物的采购供应和库存管理关系着患者的用药安全,医疗机构必须保证充足的备药量。然而靶向药物价格昂贵,人为预测药物用量难以满足多变的临床需求,并可能造成药物的存储和积压风险。因此,需要运用科学合理的数据统计工具对靶向药物的用量进行预测,有利于医疗机构实现靶向药物的精准化管理。

天津市肿瘤医院 (天津医科大学肿瘤医院) 是三级甲等肿瘤专科医院,在肿瘤靶向治疗方面具有丰富的临床经验,抗肿瘤靶向药物的临床需求大,且使用量呈现逐年上升的趋势 [4]。本研究以天津市肿瘤医院 4 种临床常用的 NSCLC 靶向药物为例,采用多元回归模型、GM (1, 1) 灰色模型以及多元回归-灰色组合模型对药物用量进行预测分析,以期为医疗机构科学管理靶向药物提供理论指导。

1 研究方法

1.1 数据来源

本研究数据来源于天津市肿瘤医院信息系统数据库,调取 2019 年 1 月 1 日—2019 年 12 月 31 日 NSCLC 靶向药物相关用量数据,通过排序统计筛选出 4 种临床上用量较大的 NSCLC 靶向药物 (吉非替尼、埃克替尼、奥希替尼和克唑替尼)。将 2019 年 1~10 月的数据用于建立预测模型,2019 年 11 月和 12 月的数据用于验证模型的预测精度与适用性。

1.2 模型建立

回归分析方法是利用数据统计原理对数据进行归一化处理,确定因变量与自变量的相关关系,并建立多元回归方程。拟合后的回归方程可以通过拟合优度 (R^2 , 也称判定系数) 和显著水平 (significant F) 进行检验。 R^2 的取值范围为 [0, 1], R^2 值越大表明模型拟合越好,一般应大于 0.6。Significant F 值主要验证因变量与自变量间的线性关系是否显著,当 F 值小于 0.05 时表示回归方程显著。多元线性回归模型的一般形式如式 (1) 所示。

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + \varepsilon \quad (1)$$

Y 为因变量, $X_1, \cdots, X_k$ 为自变量, $\beta_0, \cdots, \beta_k$ 为自变量系数, ε 为误差项

灰色模型以灰色系统理论为基础,通过原始数据序列对信息不完全、不透明的系统进行分析 [5]。灰色系统理论认为,尽管客观表象复杂,但整体功能必然蕴含某种内在规律,通过生成灰色序列弱化其随机性,进而找到规律。通过求解时间响应方程式,如式 (2) 所示,再进行累减和反向平移就可得到预测值。灰色模型的预测精度用后验差比值 (C 值) 和小误差概率 (P 值) 进行检验,当 $C \leq 0.35$, 且 $P \geq 0.95$ 时,认为模型满足一级精度。

$\tilde{x}^{(1)}(k) = \left(x^{(0)}(1) - \frac{u}{a}\right)e^{-a(k-1)} + \frac{u}{a}, k = 2, 3, \cdots, n \quad (2)$
 u, a 为方程系数, $x^{(0)}(1)$ 为原始序列第一个值, $\tilde{x}^{(1)}(k)$ 为预测一阶累加序列第 k 个值

在多元回归模型和灰色模型的基础上,有学者 [6] 提出了基于多元回归模型和灰色模型的并联组合模型。并联组合模型通过对多元回归模型和灰色模型的结果进行加权组合,从而对预测结果进行优化。如式 (3) 所示,其中 $\tilde{y}, \tilde{y}_1, \tilde{y}_2$ 分别为组合模型预测值、多元回归模型预测值和灰色模型预测值, f_1 和 $(1-f_1)$ 分别为多元回归模型和灰色模型的权值系数。

$$\tilde{y}(k) = f_1 \tilde{y}_1(k) + (1-f_1) \tilde{y}_2(k), k = 1, 2, \cdots, n \quad (3)$$

在模型的建立和验证过程中,采用平均绝对百分误差 (MAPE) 对模型精度和预测准确度进行评价,MAPE 值越小说明预测效果越好,计算方法如式 (4) 所示。

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{\text{实际值} - \text{预测值}}{\text{实际值}} \right| * 100\% \quad (4)$$

1.3 数据处理

统计 2019 年 1~12 月内 4 种靶向药物的各月用量数据,计算每月的工作日天数,并利用 Excel 软件建立数据库。分别采用 Excel 软件对 4 种靶向药物进行多元回归模型建模分析,采用 MATLAB 软件进行 GM (1, 1) 灰色模型建模和多元回归-灰色组合模型建模。

2 结果

2.1 多元回归模型分析

以 2019 年 1~10 月靶向药物实际使用量为因变量 (Y), 上月靶向药物使用量 ($X_{\text{上月}}$) 和本月工作日天数 ($X_{\text{工作日}}$) 作为自变量进行回归分析,模型拟合公式及检验结果如表 1 所示。可见吉非替尼和埃克替尼的 R^2 值分别为 0.77、0.60,处于合格范围内,而奥希替尼和

表1 多元回归模型分析结果

Table 1 Results of multiple regression analysis

药品	拟合公式	R^2 值	F 值
吉非替尼	$Y = -766.54 + 0.65 X_{\text{上月}} + 46.56 X_{\text{工作日}}$	0.77	0.006
埃克替尼	$Y = -468.16 + 0.29 X_{\text{上月}} + 62.53 X_{\text{工作日}}$	0.60	0.039
奥希替尼	$Y = -158.22 + 0.85 X_{\text{上月}} + 9.14 X_{\text{工作日}}$	0.96	1.4×10^{-5}
克唑替尼	$Y = -48.27 + 0.82 X_{\text{上月}} + 2.80 X_{\text{工作日}}$	0.86	0.001

克唑替尼的 R^2 值均大于 0.86, 拟合程度优, 同时回归方程的 F 值均小于 0.05, 表示回归方程具有显著性。

2.2 GM (1, 1) 灰色模型分析

对 2019 年 1~10 月 4 种靶向药物的使用数据进行平移变换处理, 使得数据级比满足建模要求。利用 MATLAB 软件对处理后的数据进行建模, 并做模型检验, 结果见表 2, 可见 4 种药物灰色模型的 C 值均 < 0.35 , P 值均 > 0.95 , 模型均满足一级精度。

2.3 多元回归-灰色组合模型分析

以 MAPE 最小作为优化目标, 对组合模型进行 MATLAB 建模, 确定组合模型的最优 MAPE 和此时的多元回归模型与灰色模型的权重比, 如表 3 所示。

可见组合模型的 MAPE 低于单一模型, 同时单一模型在组合模型中的权重与模型精度呈正比, 低 MAPE 值的模型在组合模型中权值较大。靶向药物实际用量与 3 种模型拟合值、预测值的对比如图 1 所示。

表2 灰色模型分析结果

Table 2 Results of gray model analysis

药品	模型时间响应方程参数	C 值	P 值
吉非替尼	$a = -0.042, u = 801.468$	0.329	1
埃克替尼	$a = -0.027, u = 1\,413.432$	0.333	1
奥希替尼	$a = -0.023, u = 554.792$	0.056	1
克唑替尼	$a = -0.008, u = 517.173$	0.137	1

表3 组合模型权重比及各模型 MAPE

Table 3 Weight ratio of combined model and the MAPE of each model

药品	回归模型与灰色模型组合权重比	MAPE/%		
		回归模型	灰色模型	组合模型
吉非替尼	0.94 : 0.06	13.73	16.42	13.53
埃克替尼	0.37 : 0.63	6.96	5.17	4.47
奥希替尼	0.79 : 0.21	4.81	6.63	4.16
克唑替尼	0.50 : 0.50	8.81	8.07	6.28

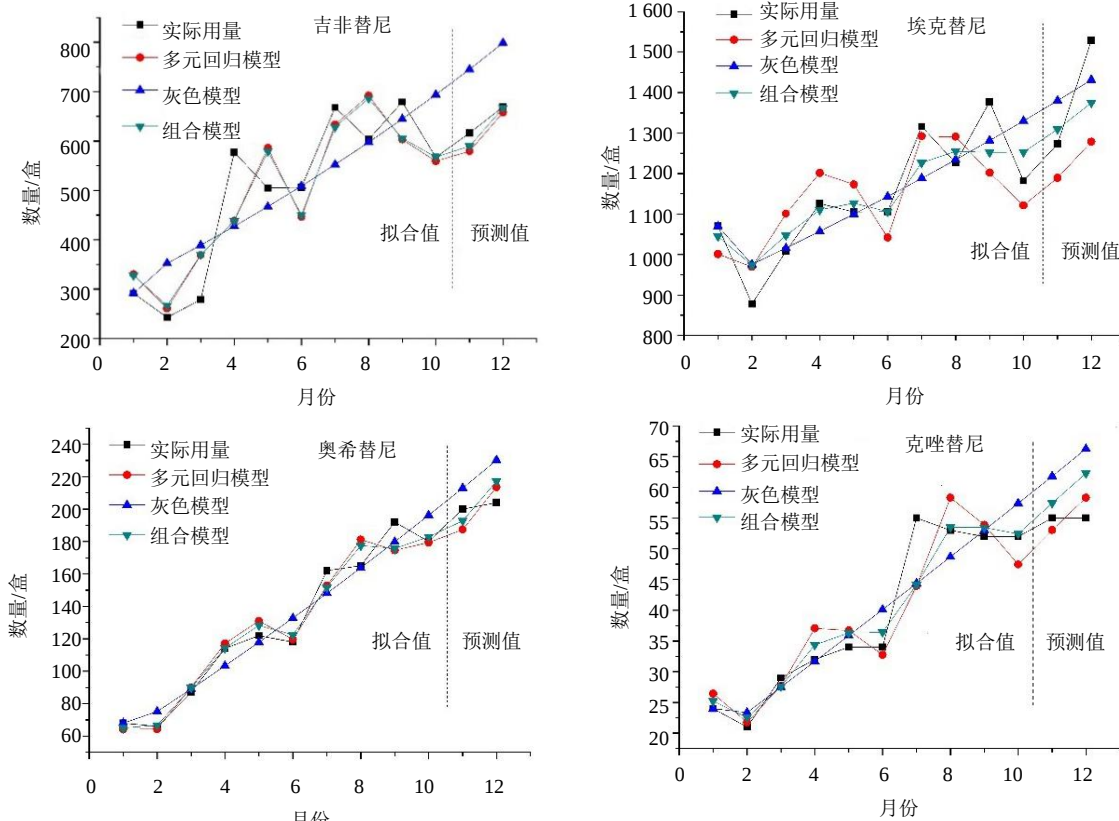


图1 药品实际用量与3种模型拟合值、预测值对比图

Fig. 1 Comparison of actual usage with fitted value and predicted value of three models

2.4 预测结果分析

使用多元回归-灰色组合模型对 11 月和 12 月的 4 种靶向药物用量进行预测, 结果如表 4 所示。可以看出, 4 种药物 11 月用量的预测误差范围为 2.87%~4.42%, 精度较高; 4 种药物 12 月用量预测误差范围为 0.39%~13.23%, 精度较低, 且波动变大。

表 4 组合模型对 11 月和 12 月靶向药物用量的预测误差
Table 4 Prediction error of combined model for targeted drug usage in November and December

月份	用量的预测误差/%			
	吉非替尼	埃克替尼	奥希替尼	克唑替尼
11 月	4.30	2.87	3.62	4.42
12 月	0.39	10.08	6.41	13.23

3 讨论

目前, 国内外文献对预测模型已有报道, 主要方法包括回归分析法、时间序列分析法、灰色模型、神经网络、遗传算法等^[7-10]。由于药品需求影响因素的各异性, 且靶向药物纳入医保时间较短, 导致用量数据贫乏, 时间序列分析、神经网络和遗传算法等数据需求量大的预测模型应用受限, 因此本研究选取数据量需求较少的多元回归模型和灰色模型进行建模分析, 并通过 2 种模型的组合提高建模精度和预测精度。

通过对天津市肿瘤医院 2019 年 1~10 月 4 种常用 NSCLC 靶向药物 (吉非替尼、埃克替尼、奥希替尼和克唑替尼) 的使用量进行多元回归模型、GM (1, 1) 灰色模型以及组合模型建模分析, 可以看到 4 种靶向药物的多元回归模型的拟合优度和显著性均符合统计学要求 ($R^2 \geq 0.6$, $F < 0.05$), 并且能够反映使用量的波动特性, 但模型会出现波动加剧, 这是因为多元回归模型使用上月的药物实际用量作为自变量之一, 当实际用量发生波动时多元回归模型会在惯性作用下延续这个波动, 同时多元回归模型对变化趋势描述不足, 从而造成预测误差。GM (1, 1) 灰色模型对 4 种靶向药物的预测精度均为 I 级, 模型近似线性, 可以对药物的实际用量变化趋势进行描述, 不受数据短期波动的影响, 但其无法预测工作日天数不均衡带来的数据波动, 从而产生预测误差。组合模型通过对单一模型的加权, 使得预测结果同

时具备描述药物使用量变化趋势和短期波动的能力, 因而使预测结果更为准确。

从建模精度看, 吉非替尼、埃克替尼、奥希替尼和克唑替尼 4 种药物的组合预测模型 MAPE 分别为 13.53%、4.47%、4.16%、6.28%, 具有较高精度。从预测结果看, 组合模型对 4 种药物的短期 (11 月) 预测误差分别为 4.30%、2.87%、3.62%、4.42%, 远期 (12 月) 预测误差分别为 0.39%、10.08%、6.41%、13.23%, 短期预测精度显著优于远期预测精度。这是因为, 11 月的预测值是以 10 月真实使用量为基础进行运算得到, 而 12 月预测值是以 11 月预测值为基础运算得到, 造成误差传递并以代数和形式叠加, 当 11 月与 12 月相对误差反向时, 就会产生如吉非替尼低预测误差 (MAPE 为 0.39%), 当 11 月与 12 月相对误差同向时, 就会产生如其他 3 种药物高预测误差 (MAPE 为 10%左右)。因此, 在实际应用中应不断对输入数据和模型进行更新, 以确保预测结果的准确性。

本研究表明, 多元回归-灰色组合模型可以有效预测短期内靶向药物的使用量, 为医疗机构的药品采购和库存管理提供数据参考, 从而实现科学精准备药。医疗机构可以根据药物的实际使用情况, 考虑将本模型作为应用模块嵌入医院信息系统 (HIS), 通过监测实时药品用量和库存数据, 来指导医疗机构药品管理, 保证患者的用药需求和用药安全。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] Zhu C N C, Zhuang W H, Chen L M, et al. Frontiers of ctDNA, targeted therapies, and immunotherapy in non-small-cell lung cancer [J]. *Transl Lung Cancer Res*, 2020, 9(1): 111-138.
- [2] Shah R, Lester J F. Tyrosine kinase inhibitors for the treatment of EGFR mutation-positive non-small-cell lung cancer: A clash of the generations [J]. *Clin Lung Cancer*, 2020, 21(3): 216-228.
- [3] Masuda N, Ohe Y, Gemma A, et al. Safety and effectiveness of alectinib in a real-world surveillance study in patients with ALK-positive non-small-cell lung cancer in Japan [J]. *Cancer Sci*, 2019, 110(4): 1401-1407.
- [4] 杨檬檬, 张 洁, 缪 玮, 等. 2017—2019 年天津市肿瘤医院抗肿瘤分子靶向药物的使用情况分析 [J]. 现代

- 药物与临床, 2020, 35(12): 2459-2466.
- [5] 邓聚龙. 灰预测与灰决策 [M]. 武汉: 华中科技大学出版社, 2002: 60-68.
- [6] 林 强, 陈一梅. 灰色多元回归模型在港口吞吐量预测中的应用 [J]. 水运工程, 2008, 417(7): 77-80, 84.
- [7] 曾 亮. 基于分数阶累加灰色 GM(1, 1)模型的生物药品销售收入预测 [J]. 高师理科学刊, 2016, 36(10): 8-11.
- [8] 宫 剑, 李洁玲. 基于神经网络的药品需求组合预测研究与应用 [J]. 中国管理信息化, 2014, 17(8): 84-88.
- [9] 柳 攀. 基于时间序列分析的销售预测方法研究 [D]. 大连: 大连理工大学, 2018.
- [10] 褚淑贞, 张环宇, 陈 露. 基层医疗机构基本药物采购资金的预测研究-基于 BP 神经网络 [J]. 中国新药杂志, 2016, 25(3): 258-261, 276.

【责任编辑 刘东博】