

## GM(1,1)模型在我国梅毒发病率预测中的应用

王雅文<sup>1</sup>, 沈忠周<sup>1</sup>, 杨银<sup>2</sup>

1. 北京协和医学院公共卫生学院, 北京 100730

2. 中国医学科学院基础医学研究所 北京协和医学院基础学院病原生物学系, 北京 100005

**摘要:** **目的** 探讨 GM(1,1)模型及其新陈代谢模型在我国梅毒发病率预测中的效果,为我国梅毒病情控制提供理论参考。**方法** 收集 2008–2017 年全国梅毒年发病率数据,用 R3.4.3 软件建立 GM(1,1)模型及其新陈代谢模型并预测,比较两类模型的拟合及预测效果。**结果** 拟建立的两个 GM(1,1)模型及两个新陈代谢模型分别为  $x^{(1)}(k+1) = 511.37e^{(0.048k)} - 491.88$ ,  $x^{(1)}(k+1) = 1\ 830.61e^{(0.016k)} - 1\ 803.75$ ,  $x^{(1)}(k+1) = 753.35e^{(0.036k)} - 730.28$ ,  $x^{(1)}(k+1) = 922.35e^{(0.016k)} - 895.48$ 。其拟合相对误差为 4.38%、0.79%、2.55%、1.96%,预测相对误差为 5.46%、5.51%、6.85%、1.48%,模型精度高。**结论** 新陈代谢 GM(1,1)模型的效果优于普通模型,用监测数据建立 GM(1,1)模型优于用预测数据建立模型。

**关键词:** 梅毒;GM(1,1);新陈代谢;预测**中图分类号:** R759.1 **文献标识码:** A **文章编号:** 1006-3110(2019)09-1069-04 **DOI:** 10.3969/j.issn.1006-3110.2019.09.012

## Application of GM (1,1) model to forecasting the incidence rate of syphilis in China

WANG Ya-wen<sup>1</sup>, SHEN Zhong-zhou<sup>1</sup>, YANG Yin<sup>2</sup>

1. School of Public Health, Peking Union Medical College, Beijing 100730, China;

2. Department of Medical Microbiology and Parasitology, Institute of Basic Medical Sciences CAMS,

School of Basic Medicine PUMC, Beijing 100005, China

Corresponding author: YANG Yin, Email: yyboby0706@163.com

**Abstract:** **Objective** To explore the performance of basic GM(1,1) model and metabolic GM(1,1) model in forecasting the incidence rate of syphilis in China so as to provide theoretical references for syphilis control. **Methods** The data about yearly incidence rates in China from 2008 to 2017 were collected, and R 3.4.3 software was used to develop models. The fitting and forecasting performance of the models was compared. **Results** The expression of two basic GM(1,1) models and metabolic GM(1,1) models was  $x^{(1)}(k+1) = 511.37e^{(0.048k)} - 491.88$ ,  $x^{(1)}(k+1) = 1\ 830.61e^{(0.016k)} - 1\ 803.75$ ,  $x^{(1)}(k+1) = 753.35e^{(0.036k)} - 730.28$ , and  $x^{(1)}(k+1) = 922.35e^{(0.016k)} - 895.48$ , respectively. The relative errors of fitting of these four models were 4.38%, 0.79%, 2.55% and 1.96%, respectively. The relative error of forecasting was 5.46%, 5.51%, 6.85% and 1.48%, respectively. The accuracy of the four models was high. **Conclusions** The performance of metabolic GM(1,1) model is better than that of basic GM(1,1) model, and the performance of GM(1,1) model built with actual data is superior to that of model built with forecasted data.

**Key words:** syphilis; GM(1,1); metabolism; prediction

梅毒,继艾滋病后的第二大性传播疾病,在全国甲乙类传染病发病顺位中仅次于乙肝和肺结核,近年来在全国及各省的发病率呈增长趋势,已发展成为重要的公共卫生问题<sup>[1-3]</sup>。目前梅毒尚无有效的疫苗预防,发病早期主要累及皮肤,随着病情发展可逐渐累及心血管系统等,造成严重不良影响。在梅毒疫情控制

方面,我国从 1987 年建立全国性病疫情监测系统开始,就将梅毒纳入为重点监测项目;2007 年 8 月,国家卫生部再次下发《关于进一步加强性病监测工作的通知》;2010 年,卫生部制定了《中国预防与控制梅毒规划(2010–2020)》,着重强调了梅毒预防与控制的重要性<sup>[4]</sup>。本研究将以此为背景,建立适用于我国梅毒疫情预测的灰色预测模型(GM(1,1))及其新陈代谢模型,评价两类模型在梅毒发病率时间序列方面的拟合及预测性能,为梅毒的预防与控制提供理论参考。

**基金项目:** 北京协和医学院 2016 年“青年教育学者计划”(2016zlgc0705)

**作者简介:** 王雅文(1994–),女,新疆呼图壁县人,在读硕士研究生,研究方向:流行病学。

**通信作者:** 杨银, E-mail: yyboby0706@163.com。

## 1 材料与方法

1.1 资料来源 全国梅毒发病率资料收集自国家统计局官方网站及国家卫生部(卫生计生委)疾病预防控制中心公布的年度全国法定传染病监测报告。基于上述数据建立我国梅毒年发病率(/10 万)数据库,用于梅毒预测模型的建立。

## 1.2 研究方法

1.2.1 研究数据 以 2008–2017 年共计十年的全国梅毒发病率作为建立及评价 GM(1,1)模型的数据。本研究共建立四个模型:模型 1 用 2008–2014 年数据建模,预测 2015 年的发病率并评价;模型 2 用 2009–2014 年的实测值和模型 1 预测的 2015 年发病率建立一次新陈代谢模型,预测 2016 年全国梅毒发病率并评价;模型 3 用 2010–2014 年实测值和模型 1、2 预测的 2015–2016 年发病率数据建立二次新陈代谢模型,预测 2017 年梅毒发病率并评价;模型 4 用 2010–2016 年实际数据建立 GM(1,1)模型,预测 2017 年梅毒发病率并评价。

1.2.2 GM(1,1)模型的建立过程<sup>[5]</sup> (1)对原始序列  $x^{(0)}$  进行一阶累加生成累加序列  $x^{(1)}$ ; (2)对序列  $x^{(1)}$  求均值,得到相邻均值生成序列  $y^{(1)}$ 。(3)根据累加序列建立一阶线性微分方程,该模型即 GM(1,1)模型; $dx^{(1)}/dt+ax^{(1)}=u$ ,其中  $a$  为发展系数, $u$  为灰色作用量,二者为模型的参数。(4)累减还原求解,得原始序列估计值序列  $\hat{x}^{(1)}(k+1)=[x^{(1)}(0)-u/a]e^{(-ak)}+u/a$ ,其中, $k=1,2,\dots,n$ 。

1.2.3 新陈代谢 GM(1,1)模型 随着社会的发展,新的扰动因素会不断出现并影响梅毒的发生发展趋势,此时旧的发病率信息的使用价值和意义将不断下降。新陈代谢理论即不断纳入新的信息,排除旧的信息,以此来保持模型较高的预测精度<sup>[6]</sup>。其建立过程为:设原始序列为  $x^{(0)}=\{x^{(1)}, x^{(2)}, \dots, x^{(n)}\}$ ,  $x^{(n+1)}$  为最新得到的数据,一阶新陈代谢模型即为新序列  $x^{(0)}=\{x^{(2)}, x^{(3)}, \dots, x^{(n)}, x^{(n+1)}\}$  所建立的预测模型。该新模型纳入了最新的数据,排除了老旧信息建立新模型称一次新陈代谢模型。在此基础上纳入  $x^{(n+2)}$ ,排除  $x^{(2)}$  所建立的模型称为二次新陈代谢模型。

1.2.4 拟合效果检验 GM(1,1)模型的常用评价指标为 后验差比值  $C$ ,  $C=Se/S_x$ ,  $Se$  表示残差序列标准差,  $S_x$  表示原始序列的标准差;小误差概率  $P$ ,  $P$  是满足残差与残差均值之差小于  $0.6475S_x$  的个数占总数的比。通常将二者联合评价模型的拟合及预测效果。 $C$  值越小,  $P$  值越大,模型拟合精度越高。 $P$  值与  $C$  值综合判断模型拟合精度标准见表 1。此外,相对误差检验也可较为直观的反映模型的拟合及预测效果,其

计算方法为实际值与预测值的差值除以实际值。

表 1 GM(1,1)模型的精度评判标准

模型精度等级	$P$	$C$
1 级(好)	$0.950 \leq P$	$C \leq 0.35$
2 级(合格)	$0.800 \leq P < 0.950$	$0.35 < C \leq 0.50$
3 级(勉强)	$0.700 \leq P < 0.800$	$0.50 < C \leq 0.65$
4 级(不合格)	$P < 0.700$	$0.65 < C$

1.3 统计学分析 运用 Excel 2016 建立 2008–2017 年全国梅毒发病率数据库,使用 R 3.4.3 软件编程建立 GM(1,1)灰色预测模型及其新陈代谢模型,并预测。

## 2 结果

2.1 GM(1,1)模型建立 2008–2017 年全国梅毒发病率呈明显上升趋势。模型 1 利用 2008–2014 年全国梅毒发病率监测数据建立灰色模型  $x^{(1)}(k+1)=511.37e^{(0.048k)}-491.88$ , 预测 2015 年梅毒发病率为 33.59/10 万人口。模型 4 利用 2010–2016 年全国梅毒发病率资料建模  $x^{(1)}(k+1)=1830.61e^{(0.016k)}-1803.75$ , 预测 2017 年该值为 32.59/10 万人。

2.2 新陈代谢模型 模型 2 利用 2009–2014 年的实测值和模型 1 预测的 2015 年发病率数据建模灰色预测模型并预测 2016 年我国梅毒发病率值;建立的模型为  $x^{(1)}(k+1)=753.35e^{(0.036k)}-730.28$ , 预测 2016 年梅毒发病率为 34.16/10 万。模型 3 利用 2010–2014 年实测值和模型 1、2 预测的 2015–2016 年发病率数据建立二次新陈代谢模型  $x^{(1)}(k+1)=922.35e^{(0.016k)}-895.48$ , 预测 2017 年全国梅毒发病率为 35.00/10 万人口。四个模型的拟合及预测情况见表 2。

表 2 梅毒发病率预测模型的拟合、预测及评价

时间(年度)	观测值	模型 1	模型 2	模型 3	模型 4
2008	19.49	19.49			
2009	23.07	25.15	23.07		
2010	26.86	26.40	27.59	26.86	26.86
2011	29.47	27.70	28.60	29.04	29.54
2012	30.44	29.07	29.63	29.96	30.02
2013	30.04	30.50	30.70	30.91	30.52
2014	30.93	32.01	31.82	31.88	31.02
2015	31.85	33.59	32.97	32.89	31.54
2016	31.97		34.16	33.93	32.06
2017	34.49			35.00	32.59
拟合相对误差(%)		4.38	2.55	1.96	0.79
$C$ 值		0.19	0.10	0.15	0.12
$P$ 值		>0.99	>0.99	>0.99	>0.99

2.3 模型评价 模型 1–4 的  $C$  值和  $P$  值反映模型的精度达到一级,四个模型的相对误差依次减小。在预测精度方面,模型 1–4 的预测误差分别为 5.46%、6.85%、1.48%、5.51%。通过观察模型 2 和模型 3 的拟合和预测结果可见,二次新陈代谢模型的精度高于

一次新陈代谢(图1);通过观察模型3和模型4的拟合和预测结果可知,用实际监测值建立模型的精度高于用预测值建立的模型(图2)。见表2、图1~图2。

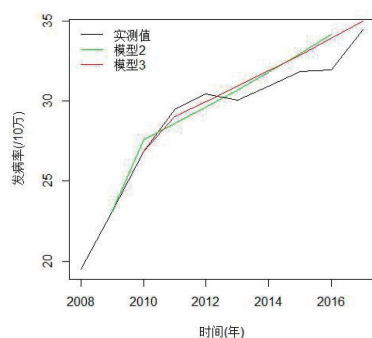


图1 梅毒发病率新陈代谢模型拟合及预测曲线

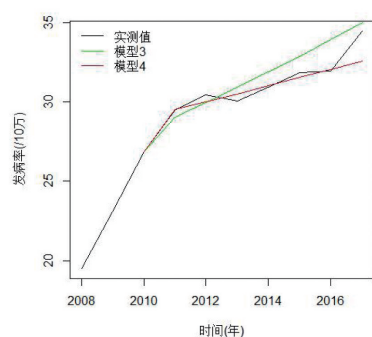


图2 梅毒发病率新陈代谢模型与GM(1,1)模型拟合及预测曲线

### 3 讨论

本研究建立了两个GM(1,1)模型及两个GM(1,1)新陈代谢模型,运用所建模型进行预测并评价,结果显示新陈代谢模型的拟合精度随着数据的不断更新而有所提升,但是精度略低于纳入更新的实际监测数据建立的模型。新陈代谢模型的预测误差随着代谢次数的增加而有所变化,但低于利用实际监测数据得到的模型的预测精度。

传染病发病的模型预测研究是对其发病预防和控制的重要组成部分<sup>[7-8]</sup>。梅毒作为重要法定传染病之一,对其预测研究也较为普遍。高礼勇等<sup>[9]</sup>运用GM(1,1)模型拟合并预测如皋市梅毒流行特征,预测精度为二级。马晓梅等<sup>[10]</sup>建立了ARIMA乘积季节模型和Holt-Winters季节模型拟合并预测全国梅毒月发病率情况。不同于本次研究,以月发病率数据建立模型,需考虑数据的周期性及季节趋势。本次研究以梅毒年发病率资料建立预测模型,在短时间内,影响梅毒发病的经济、医疗、文化等社会因素及季节、气候等呈周期性变化的自然因素不会发生较大变动。因此本研究采用年发病数据建模的影响因素更少,结果准确性更高。

李莉等<sup>[11]</sup>同样采用多种模型拟合并预测了甘肃省梅毒发病率情况,其中GM(1,1)模型的拟合误差达24.89%,误差高于本次研究结果。分析误差差异较大可能是由于数据特点造成的。GM(1,1)模型适用于呈指数变化、无波动或波动较小的序列<sup>[12]</sup>,本研究所用数据存在较小波动,故拟合效果相对较好。

时间序列分析中,由于时间是唯一可知影响观测值的变量,观测值随时间变化的趋势应不断被纳入模型学习,以维持模型的高预测性能,绝大多数时间序列的建模分析中均提出,数据需不断更新才能适应序列的发展趋势<sup>[13-14]</sup>。新陈代谢模型在疾病发病预测中的应用也较为普遍<sup>[15]</sup>,其采用动态更新原始数据的方式对自身数据进行修正,使模型更好的适应数据的变化规律和发展趋势,从而提高模型精度。本研究建立的两个新陈代谢模型,其预测误差逐渐降低,拟合精度始终保持在一级,这提示在GM(1,1)模型预测时数据修正的重要性。将模型3与模型4相比较,结果显示新陈代谢模型的误差较实测值建立的模型大,这样提示模型预测值作为建模值存在一定的缺点。综合本次研究的结果,利用GM(1,1)模型进行短期预测时,采用实测值建模将为模型拟合和预测提供保障,但是在利用该模型进行中长期预测时,需根据模型预测结果不断更新新数据,以准确判断序列的未来发展。

《中国预防与控制梅毒规划(2010-2020年)》目标旨在到2020年,全国一期和二期梅毒年报告发病率呈下降趋势,先天梅毒年报告发病率在15/10万活产数以下。根据本研究结果,到2017年全国梅毒发病率仍存在明显上升趋势,距离实现规划目标还有一定的距离。本研究结果可用于指导梅毒发病预测模型的选择,依据合适的模型预测合理的结果,在梅毒流行发生前做好准备预防其爆发等。本次研究尚存在一定的局限性,如本研究仅建立了GM(1,1)模型对全国梅毒发病率进行拟合和预测,并评价其效果,而未选取其他模型与之相比。时间序列分析及预测方法有很多种,本研究采用了其中一种进行拟合,并建立新陈代谢模型相比较,研究结果将仅反映该种模型在梅毒发病率预测中的应用,更多的模型仍有待后续研究。

### 参考文献

- [1] 赵鹏鹏, 胡锦涛, 孙修福, 等. 2006-2015年江苏淮安市梅毒流行病学特征分析[J]. 公共卫生与预防医学, 2017, 28(4): 73-76.
- [2] 张万宏, 徐业华, 石卫东, 等. 2010-2015年武汉市梅毒网络报告发病率与人群估计感染率对比分析[J]. 实用预防医学, 2018, 25(10): 1184-1186.
- [3] 陈福亭, 郑熠, 石萍, 等. 2008-2016年武汉市性传播疾病流行状况及趋势分析[J]. 实用预防医学, 2018, 25(7): 805-809.

(下转第1079页)