

我国包虫病报告病例数自回归移动平均模型预测研究

谭恩丽¹, 王正峰², 周文策², 李石柱³, 卢艳³, 艾琳³, 蔡玉春³, 滕雪娇³, 张顺先³,
党志胜³, 杨春利⁴, 陈家旭³, 胡薇^{3,5}, 周晓农³, 田利光^{3*}

【摘要】 目的 采用自回归移动平均模型(Autoregressive integrated moving average, ARIMA)对全国(不含港、澳、台地区)包虫病月报告病例数进行预测,为包虫病的防控提供科学参考。**方法** 通过SPSS 24.0软件,分别以2007–2015年和2007–2014年全国包虫病月报告病例数,分别建立最优的ARIMA模型,并进行模型比较。**结果** 2007–2015年全国包虫病月报告病例数的最优模型为ARIMA(1,0,0)(1,1,0)₁₂,预测相对误差为-13.97%,AR(1)=0.367($t=3.816$, $P<0.001$),SAR(1)=-0.328($t=-3.361$, $P=0.001$),Ljung-Box $Q=14.119$ ($df=16$, $P=0.590$)。2007–2014年全国包虫病月报告病例数的最优模型为ARIMA(1,0,0)(1,0,1)₁₂,预测相对误差为0.56%,AR(1)=0.413($t=4.244$, $P<0.001$),SAR(1)=0.809($t=9.584$, $P<0.001$),SMA(1)=0.356($t=2.278$, $P=0.025$),Ljung-Box $Q=18.924$ ($df=15$, $P=0.217$)。**结论** 时间序列不同,所建立的预测模型可能不同。数据积累越多、预测时间越短、预测误差越小的情况还需得到进一步验证。模型的建立和预测应用是动态过程,需要不断根据积累的数据进行调整,但同时要充分考虑影响传染病报告病例数相关工作(普查和专项调查等)的影响。

【关键词】 包虫病;月报告病例数;自回归移动平均模型;建模

【中图分类号】 R53 【文献标识码】 A

Study on the ARIMA model application to predict echinococcosis cases in China

TAN En-li¹, WANG Zheng-feng², ZHOU Wen-ce², LI Shi-zhu³, LU Yan³, AI Lin³, CAI Yu-chun³, TENG Xue-jiao³, ZHANG Shun-xian³, DANG Zhi-sheng³, YANG Chun-li⁴, CHEN Jia-xu³, HU Wei^{3,5}, ZHOU Xiao-nong³, TIAN Li-guang^{3*}

1 Department of Gerontal Respiratory Medicine, First Hospital of Lanzhou University, Lanzhou 730000, China; 2 Second General Surgery Department, First Hospital of Lanzhou University, China; 3 National Institute of Parasitic Diseases, Chinese Center for Disease Control and Prevention; Key Laboratory for Parasitology and Vector Biology, National Health and Family Planning Commission, WHO Collaborating Center for Tropical Diseases, National Center for International Research on Tropical Diseases, China; 4 Shanghai Dermatology Hospital, China; 5 Department of Microbiology and Microbial Engineering, School of Life Sciences, Fudan University, China

* Corresponding author

【Abstract】 Objective To predict the monthly reported echinococcosis cases in China with the autoregressive integrated moving average (ARIMA) model, so as to provide a reference for prevention and control of echinococcosis. **Methods** SPSS 24.0 software was used to construct the ARIMA models based on the monthly reported echinococcosis cases of time series from 2007 to 2015 and 2007 to 2014, respectively, and the accuracies of the two ARIMA models were compared. **Results** The model based on the data of the monthly reported cases of echinococcosis in China from 2007 to 2015 was ARIMA (1, 0, 0) (1, 1, 0)₁₂, the relative error among reported cases and predicted cases was -13.97%, AR (1) = 0.367 ($t=3.816$, $P<0.001$), SAR (1) = -0.328 ($t=-3.361$, $P=0.001$), and Ljung-Box $Q=14.119$ ($df=16$, $P=0.590$). The model based on the data of the monthly reported cases of echinococcosis in China from 2007 to 2014 was ARIMA (1, 0, 0) (1, 0, 1)₁₂, the relative error among reported cases and predicted cases was 0.56%, AR (1) = 0.413 ($t=4.244$, $P<0.001$), SAR (1) = 0.809 ($t=9.584$, $P<0.001$), SMA (1) = 0.356 ($t=2.278$, $P=0.025$), and Ljung-Box $Q=18.924$ ($df=15$, $P=0.217$). **Conclusions** The different time series may have different ARIMA models as for the same infectious diseases. It is needed to be further verified that

【基金项目】 国家自然科学基金(81473022)

【作者单位】 1 兰州大学第一医院老年呼吸科(兰州 730000);2 兰州大学第一医院普外二科;3 中国疾病预防控制中心寄生虫病预防控制所、国家卫生和计划生育委员会寄生虫病原与媒介生物学重点实验室、世界卫生组织热带病合作中心、国家级热带病国际联合研究中心;4 上海市皮肤病医院;5 复旦大学生命科学学院微生物学与微生物工程系

【作者简介】 谭恩丽,女,硕士,主治医师。研究方向:呼吸系统疾病

* 通信作者 E-mail: 13651914365@126.com

【数字出版日期】 2018-02-26 13:47

【数字出版网址】 <http://kns.cnki.net/kcms/detail/32.1374.R.20180226.0919.004.html>

the more data are accumulated, the shorter time of predication is, and the smaller the average of the relative error is. The establishment and prediction of an ARIMA model is a dynamic process that needs to be adjusted and optimized continuously according to the accumulated data, meantime, we should give full consideration to the intensity of the work related to infectious diseases reported (such as disease census and special investigation).

[Key words] Echinococcosis; Monthly reported cases; Autoregressive integrated moving average (ARIMA) model; Modeling

棘球蚴病(又称“包虫病”)是全球畜牧业地区常见的人畜共患病之一^[1],我国主要流行区包括西藏、新疆、内蒙古、四川、甘肃、青海和宁夏等畜牧业较为发达的地区^[2-3]。棘球蚴主要感染肝、肾、脑和肺等组织或器官^[4-5],棘球蚴病病程较长,危害严重,是流行区居民因病致贫、因病返贫的重要传染病之一,已成为当地高度关注和亟待解决的公共卫生问题之一^[6-7]。因此,掌握我国包虫病流行趋势,是开展相关健康教育、检测技能培训和配置必要卫生资源的前提和基础,也是包虫病防控的重要内容之一。本研究采用自回归移动平均模型(Autoregressive integrated moving average, ARIMA),对我国包虫病月报告病例数进行预测研究,为其防控提供科学参考。

材料与方法

1 资料来源

原始数据来源于中华人民共和国国家卫生和计划生育委员会网站中全国法定传染病疫情概况(每月公布)、《疾病监测》杂志每期公布(每月)的全国法定传染病疫情动态简介,浏览网站或下载文献,并摘录2007年1月-2016年12月全国包虫病月报告病例数(不含港、澳、台地区),建立Excel数据库。

2 ARIMA模型建立及预测比较过程

本研究采用时间序列乘积季节模型ARIMA(p, d, q) \times (P, D, Q) s ,其中 d 为一般差分阶数, p 和 q 为自回归移动平均阶数, D 为季节差分阶数, P 和 Q 为季节性自回归移动平均阶数, s 为季节周期。ARIMA的建模过程分5个步骤:数据的平稳化判断和处理、模型识别、参数估计及检验、模型的显著性检验和模型预测。以上5过程需反复进行,根据模型是否简洁(模型越简洁越好)、ARIMA中各参数是否有统计学意义($P < 0.05$)、模型检验指标(Box-Ljung统计量 Q 值及 P 值, $P > 0.05$)、平稳决定系数(R^2)(越大越好)、标准化贝叶斯准则(BIC)(越小越好)、均方根误差(越小越好)和实际预测误差(越小越好),筛选最优模型。预测比较过程:以2007年1月-2015年12月全国包虫病月报告病例数建立模型,用2006年1-12月数据进行验证,以2007年1月-2014年12月全国包虫病月报告

病例数建立模型,用2005年1月-2016年12月数据进行验证。通过比较两组序列预测误差的大小,判断两组序列的实际应用效果。

3 统计软件

采用Excel 2010软件对2007年1月-2016年12月全国包虫病月报告病例数进行整理,应用SPSS 24.0软件进行ARIMA模型的建立和应用。ARIMA各参数以 $P < 0.05$ 为差异有统计学意义。

结 果

1 原始数据特征

2007-2016年全国包虫病报告病例总数为44 329例(图1)。报告病例数位列前3位的年份为2008年(5 919例)、2007年(5 879例)和2016年(5 290例),报告病例数较少的年份为2012年(3 591例)、2011年(3 421例)和2009年(3 310例)(图1)。从图2可以看出:全国包虫病月报告病例数在每年的11月和12月都相对较多,尤其在2007年和2008年11月和12月;但报告病例数在每年的1月和2月相对较少,呈现一定的季节性特征,同时2007-2016年全国包虫病月报告病例数的自相关函数(Autocorrelation function, ACF)和偏自相关函数(Partial autocorrelation function, PACF)既不截尾也不拖尾,故考虑使用ARIMA季节模型,其一般表示为ARIMA(p, d, q) \times (P, D, Q) s ,其中 P, Q 一般不会超过2阶,可依次取0、1、2并由低阶到高阶逐个尝试。

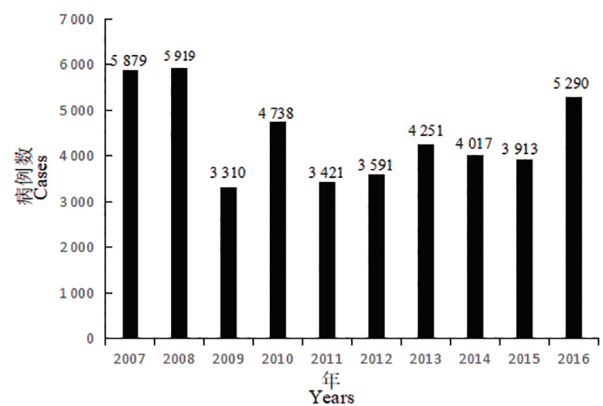


图1 2007-2016年全国包虫病报告病例数

Fig. 1 Prevalence trend of echinococcosis cases in China from 2007 to 2016

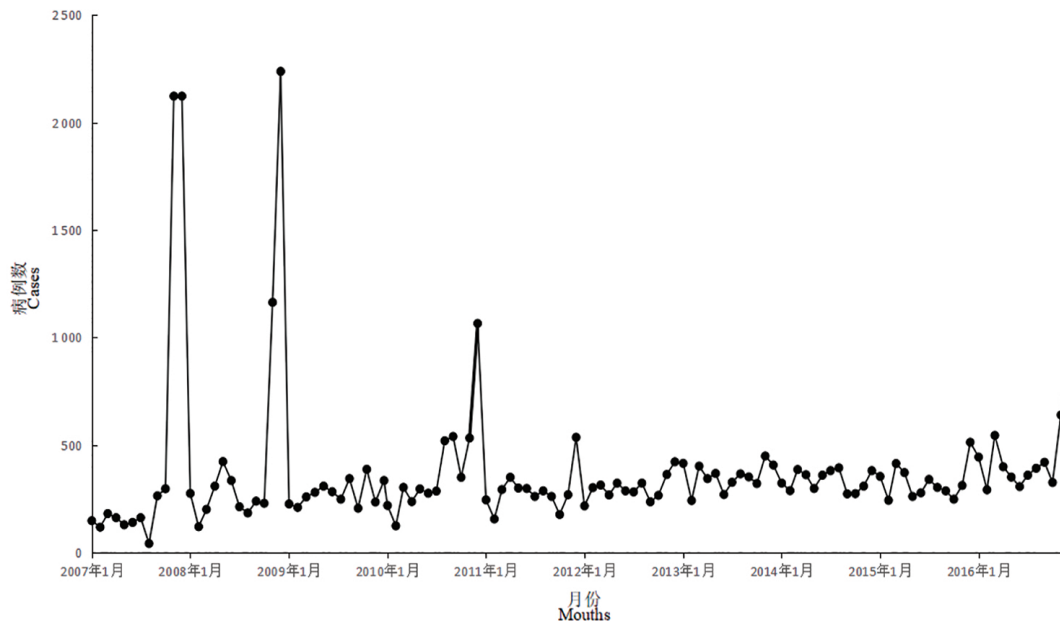


图2 2007–2016年全国包虫病月报告病例数时间序列

Fig. 2 Time series of monthly reported cases of echinococcosis in China from 2007 to 2016

2 根据2007–2015年数据建立、识别模型及预测应用

根据2007年1月–2015年12月全国包虫病月报告病例数,建立时间序列。经自然对数转化、一次季节差分和一次一般差分后的序列近似的平稳(ACF图和PACF图近似截尾)。因此,初步确定ARIMA模型为 $(0,0,0)(0,1,0)_{12}$ 。根据模型简洁、各参数有统计学意义等基本条件,筛选出11个基本模型(表1)。然后再从这些基本模型中选择平稳决定系数(R^2)最大ARIMA(1,0,0)(1,0,1)₁₂(0.272)、均方根误差最小ARIMA(0,0,2)(1,0,0)₁₂(230.654)、标化BIC最小ARIMA(0,0,2)(1,0,0)₁₂(11.072)、Ljung-Box Q 值最小ARIMA(1,0,0)(1,1,0)₁₂(14.119)作为备用模型(表2),分别用这些备用模型预测2016年1–12月全国包虫病月报告病例数(表2),并与实际报告病例数进行比较,发现ARIMA(1,0,0)(1,1,0)₁₂模型的预测相对误差最小(−13.97%),故确定其为最优模型,其主要参数包括 $AR(1)=0.367(t=3.816, P<0.001)$, $SAR(1)=-0.328(t=-3.361, P=0.001)$, Ljung-Box $Q=14.119(df=16, P=0.590)$ (表1)。

3 根据2007–2014年数据建立、识别模型及预测应用

根据2007年1月–2014年12月全国包虫病月报告病例数,建立时间序列。根据模型简洁、各参数有统计学意义($P<0.05$)和Ljung-Box统计量所对应的

$P>0.05$ 的基本条件,筛选出12个基本模型(表3)。从这些基本模型中选择平稳决定系数(R^2)最大ARIMA(0,1,1)(0,1,1)₁₂(0.277)、ARIMA(1,0,1)(1,0,1)₁₂(0.277)、ARIMA(1,1,1)(0,1,0)₁₂(0.277)、均方根误差最小ARIMA(0,0,2)(2,1,0)₁₂(239.945)、标化BIC最小ARIMA(0,0,2)(2,1,0)₁₂(11.255)、Ljung-Box Q 值最小ARIMA(1,0,0)(0,1,1)₁₂(14.466)的这5个模型作为备用模型(表4)。分别以这5种备用模型预测2015年1月–2016年12月全国包虫病月报告病例数(表4),并与实际报告病例数进行比较,发现ARIMA(1,0,0)(1,0,1)₁₂模型的预测相对误差最小(0.56%),故确定其为最优模型,其主要参数包括 $AR(1)=0.413(t=4.244, P<0.001)$, $SAR(1)=0.809(t=9.584, P<0.001)$, $SMA(1)=0.356(t=2.278, P=0.025)$ 和Ljung-Box $Q=18.924(df=15, P=0.217)$ 。

4 不同年代时间序列模型的比较

根据2007–2015年数据,建立的最优模型为ARIMA(1,0,0)(0,1,1)₁₂,预测2016年全国包虫病月报告病例数时预测相对误差为−13.97%(表2)。根据2007–2014年数据,建立的最优模型为ARIMA(1,0,0)(1,0,1)₁₂,预测2015年全国包虫病月报告病例数时预测相对误差为12.99%,预测2016年全国包虫病月报告病例数时预测相对误差为−11.88%,预测2015–2016年全国包虫病月报告病例数时预测相对误差为0.56%(表4)。

表1 几种ARIMA模型的参数估计(根据2007–2015年全国包虫病月报告病例数)

Table 1 Parameter estimations of several ARIMA models (according to the monthly reported cases of echinococcosis from 2007 to 2015)

模型 Model	参数 Parameter	估计值 Estimate	t 值 t value	P 值 P value	平稳 R^2 Stationary R^2	均方根误差 Root mean square error	标化BIC Normalized BIC	Ljung-Box Q	df	P 值 P value
ARIMA(0,0,1)(1,1,0) ₁₂	MA(1)	-0.289	-2.914	0.004	0.174	265.796	11.308	16.447	16	0.442
	SAR(1)	-0.327	-3.358	0.001						
ARIMA(0,0,2)(1,1,0) ₁₂	MA(1)	-0.306	-3.029	0.003	0.123	230.654	11.072	15.508	15	0.415
	MA(2)	-0.256	-2.567	0.012						
ARIMA(0,0,1)(0,0,1) ₁₂	SAR(1)	-0.366	-3.781	<0.001	0.269	286.039	11.442	23.869	16	0.092
	MA(1)	-0.431	-4.855	<0.001						
ARIMA(0,1,1)(0,1,0) ₁₂	SMA(1)	-0.503	-5.163	<0.001	0.229	315.989	11.607	25.652	17	0.081
ARIMA(1,0,0)(0,1,0) ₁₂	MA(1)	0.656	8.200	<0.001	0.133	250.516	11.142	24.779	17	0.100
ARIMA(1,0,0)(1,1,0) ₁₂	AR(1)	0.367	3.819	<0.001	0.203	250.469	11.189	14.119	16	0.590
	SAR(1)	0.367	3.816	<0.001						
ARIMA(1,0,0)(2,1,0) ₁₂	AR(1)	-0.328	-3.361	0.001	0.206	249.818	11.232	18.016	15	0.262
	SAR(1)	-0.455	-4.220	<0.001						
ARIMA(1,0,0)(2,1,1) ₁₂	SAR(2)	-0.280	-2.772	0.007	0.210	251.793	11.295	16.787	14	0.268
	AR(1)	0.401	4.126	<0.001						
ARIMA(1,1,0)(1,1,0) ₁₂	SAR(1)	-0.974	-4.231	<0.001	0.215	261.812	11.279	25.333	16	0.064
	SAR(2)	-0.451	-4.728	<0.001						
ARIMA(1,1,0)(1,0,1) ₁₂	SMA(1)	-0.593	-2.189	0.031	0.272	290.491	11.517	20.673	15	0.148
	AR(1)	-0.379	-3.934	<0.001						
ARIMA(1,0,0)(1,0,0) ₁₂	SAR(1)	-0.308	-3.113	0.002	0.270	286.135	11.443	22.453	16	0.129
	AR(1)	0.408	4.417	<0.001						
ARIMA(1,0,0)(1,0,0) ₁₂	SAR(1)	0.831	11.669	<0.001						
	SMA(1)	0.375	2.680	0.009						
ARIMA(1,0,0)(1,0,0) ₁₂	AR(1)	0.410	4.483	<0.001						
	SAR(1)	0.593	7.917	<0.001						

表2 根据2007–2015年数据建立模型,比较2016年全国包虫病月报告病例数预测值和实际值的差异

Table 2 ARIMA model established according to the monthly reported cases of echinococcosis from Jan 2007 to Dec 2015, and comparison of actual monthly echinococcosis cases with predicted echinococcosis cases from ARIMA model in 2016

月份 Month	报告 病例数 Reported number	ARIMA(0,0,2) (1,1,0) ₁₂		ARIMA(1,0,0) (1,1,0) ₁₂		ARIMA(1,0,0) (1,0,1) ₁₂	
		预测人数 Predicted number	相对误差 Relative error (%)	预测人数 Predicted number	相对误差 Relative error (%)	预测人数 Predicted number	相对误差 Relative error (%)
2016-Jan	443	365	-17.61	387	-12.64	339	-23.48
2016-Feb	291	300	3.09	289	-0.69	268	-7.90
2016-Mar	544	429	-21.14	432	-20.59	360	-33.82
2016-Apr	398	394	-1.01	395	-0.75	338	-15.08
2016-May	350	299	-14.57	299	-14.57	313	-10.57
2016-Jun	306	340	11.11	339	10.78	310	1.31
2018-Jul	359	386	7.52	387	7.80	328	-8.64
2016-Aug	391	371	-5.12	371	-5.12	347	-11.25
2016-Sep	419	300	-28.40	301	-28.16	306	-26.97
2016-Oct	325	280	-13.85	280	-13.85	291	-10.46
2016-Nov	639	333	-47.89	334	-47.73	340	-46.79
2016-Dec	825	473	-42.67	477	-42.18	428	-48.12
平均相对误差 (中位数) Average relative error (median)	-	-	-14.21	-	-13.24	-	-13.16
平均相对误差(均数) Average relative error (mean)	-	-	-14.21	-	-13.97	-	-20.15

表3 ARIMA 模型的参数估计(根据2007–2014年全国包虫病月报告病例数)

Table 3 Parameter estimation of ARIMA models (according to the monthly reported cases of echinococcosis from 2007 to 2014)

模型 Model	参数 Parameter	估计值 Estimate	<i>t</i> 值 <i>t</i> value	<i>P</i> 值 <i>P</i> value	平稳 R^2 Stationary R^2	均方根 误差 Root Mean square error	标化 BIC Normalized BIC	Ljung- Box <i>Q</i>	<i>df</i>	<i>P</i> 值 <i>P</i> value
ARIMA(0,0,1)(0,0,1) ₁₂	MA(1)	-0.434	-4.611	<0.001	0.274	301.924	11.563	22.838	16	0.116
	SMA(1)	-0.508	-4.895	<0.001						
ARIMA(0,0,1)(0,1,1) ₁₂	MA(1)	-0.304	-2.899	0.005	0.180	291.470	11.508	16.509	16	0.418
	SMA(1)	0.397	3.485	0.001						
ARIMA(0,0,1)(2,1,1) ₁₂	MA(1)	-0.325	-3.076	0.003	0.185	301.536	11.682	17.180	14	0.247
	SAR(1)	-1.032	-3.846	<0.001						
	SAR(2)	-0.466	-4.586	<0.001						
	SMA(1)	-0.670	-2.027	0.046						
	MA(1)	-0.370	-3.401	0.001						
ARIMA(0,0,2)(2,1,0) ₁₂	MA(2)	-0.250	-2.297	0.024	0.218	239.945	11.225	17.036	14	0.254
	SAR(1)	-0.493	-4.222	<0.001						
	SAR(2)	-0.289	-2.620	0.011						
ARIMA(0,1,1)(0,1,1) ₁₂	MA(1)	0.651	7.539	<0.001	0.277	356.309	11.911	22.005	16	0.143
	SMA(1)	0.311	2.752	0.007						
ARIMA(1,0,0)(0,1,1) ₁₂	AR(1)	0.381	3.702	<0.001	0.209	270.568	11.359	14.466	16	0.564
	SMA(1)	0.398	3.500	0.001						
ARIMA(1,0,0)(1,0,0) ₁₂	AR(1)	0.415	4.312	<0.001	0.275	302.622	11.568	20.375	16	0.204
	SAR(1)	0.575	7.183	<0.001						
	AR(1)	0.413	4.244	<0.001						
ARIMA(1,0,0)(1,0,1) ₁₂	SAR(1)	0.809	9.584	<0.001	0.277	307.031	11.644	18.924	15	0.217
	SMA(1)	0.356	2.278	0.025						
ARIMA(1,1,0)(0,1,1) ₁₂	AR(1)	-0.361	-3.483	0.001	0.218	284.005	11.458	23.631	16	0.098
	SMA(1)	0.351	3.080	0.003						
ARIMA(1,1,1)(0,1,0) ₁₂	AR(1)	0.386	3.416	0.001	0.277	327.923	11.745	19.737	16	0.250
	MA(1)	0.996	2.304	0.024						
ARIMA(1,0,0)(2,1,0) ₁₂	AR(1)	0.401	3.857	<0.001	0.207	271.581	11.420	16.846	15	0.328
	SAR(1)	-0.453	-3.916	<0.001						
	SAR(2)	-0.296	-2.709	0.008						
	AR(1)	0.964	24.113	<0.001						
	AR(2)	-0.967	-29.629	<0.001						
ARIMA(2,0,3)	MA(1)	0.494	4.574	<0.001	0.263	311.965	11.771	21.978	13	0.056
	MA(2)	-0.588	-5.861	<0.001						
	MA(3)	-0.338	-3.158	0.002						

表4 根据2007–2014年数据建立模型,比较2015–2016年全国包虫病月报告病例数预测值和实际值的差异
Table 4 ARIMA model established according to monthly reported cases of echinococcosis from Jan 2007 to Dec 2014, and comparison of actual monthly malaria cases with predicted malaria cases from ARIMA model from 2015 to 2016

月份 Month	报告 病例数 No. reported cases	ARIMA(0,0,2) (2,1,0) ₁₂		ARIMA(0,1,1) (0,1,1) ₁₂		ARIMA(1,0,0) (0,1,1) ₁₂		ARIMA(1,0,0) (1,0,1) ₁₂		ARIMA(1,1,1) (0,1,0) ₁₂	
		预测人数 Predicted number	相对误差 Relative error (%)	预测人数 Predicted number	相对误差 Relative error (%)	预测人数 Predicted number	相对误差 Relative error (%)	预测人数 Predicted number	相对误差 Relative error (%)	预测人数 Predicted number	相对误差 Relative error (%)
2015-Jan	353	319	-9.63	322	-8.78	352	-0.28	344	-2.55	337	-4.53
2015-Feb	242	325	34.30	269	11.16	309	27.69	307	26.86	305	26.03
2015-Mar	413	442	7.02	377	-8.72	446	7.99	391	-5.33	409	-0.97
2015-Apr	371	396	6.74	348	-6.20	412	11.05	368	-0.81	382	2.96
2015-May	260	384	47.69	314	20.77	380	46.15	344	32.31	314	20.77
2015-Jun	277	383	38.27	334	20.58	390	40.79	354	27.80	378	36.46
2015-Jul	339	408	20.35	363	7.08	419	23.60	373	10.03	399	17.70
2015-Aug	302	442	46.36	388	28.48	451	49.34	391	29.47	411	36.09
2015-Sep	286	332	16.08	294	2.80	346	20.98	323	12.94	284	-0.70
2015-Oct	247	338	36.84	290	17.41	338	36.84	317	28.34	284	14.98
2015-Nov	311	421	35.37	352	13.18	416	33.76	365	17.36	319	2.57
2015-Dec	512	480	-6.25	413	-19.34	487	-4.88	407	-20.51	393	-23.24
2016-Jan	443	407	-8.13	352	-20.54	401	-9.48	354	-20.09	349	-21.22
2016-Feb	291	342	17.53	300	3.09	342	17.53	319	9.62	315	8.25
2016-Mar	544	491	-9.74	431	-20.77	488	-10.29	387	-28.86	421	-22.61
2016-Apr	398	438	10.05	406	2.01	450	13.07	368	-7.54	393	-1.26
2016-May	350	423	20.86	374	6.86	414	18.29	349	-0.29	321	-8.29
2016-Jun	306	401	31.05	406	32.68	425	38.89	357	16.67	386	26.14
2018-Jul	359	443	23.40	450	25.35	457	27.30	372	3.62	406	13.09
2016-Aug	391	480	22.76	492	25.83	491	25.58	387	-1.02	418	6.91
2016-Sep	419	381	-9.07	380	-9.31	377	-10.02	331	-21.00	287	-31.50
2016-Oct	325	374	15.08	383	17.85	368	13.23	327	0.62	287	-11.69
2016-Nov	639	473	-25.98	475	-25.67	453	-29.11	366	-42.72	321	-49.77
2016-Dec	825	512	-37.94	569	-31.03	531	-35.64	400	-51.52	394	-52.24
平均相对误差(中位数)(2015–2016)	–	–	16.80		4.97		17.91		0.16	–	0.94
Average relative error (median)											
平均相对误差(均值)(2015–2016)	–	–	13.46		3.53		14.68		0.56	–	-0.67
Average relative error (mean)											
平均相对误差(中位数)(2015)	–	–	27.33		9.12		25.64		15.15	–	8.97
Average relative error (median)											
平均相对误差(均值)(2015)	–	–	22.76		6.53		24.42		12.99	–	10.68
Average relative error (mean)											
平均相对误差(中位数)(2016)	–	–	12.56		2.55		13.15		-4.28	–	-9.99
Average relative error (median)											
平均相对误差(均值)(2016)	–	–	4.16		0.53		4.94		-11.88	–	-12.02
Average relative error (mean)											

讨 论

本研究发现2016年全国包虫病月报告病例数的预测值和实际值的相对误差较小(13.97%),说明ARIMA模型在一定程度上可准确估计我国的包虫病疫情,这与其他回归移动平均模型预测研究的结论相似^[8]。但本研究发现ARIMA应用过程中,部分月份的实际病例数和预测病例数之间相对误差较大(有些月份的误差超过50%),提示时间序列自身变化蕴含的规律并不能完全解释数据的波动,这与我国手足口病重症患者数自回归移动平均模型预测研究的结论相似^[9]。ARIMA模型简单易用,它包括了诸多影响包虫病发生发展的自然因素,如动物迁徙、野外棘球蚴的存活和数量、人口流动、气温等。不可预知或突然变化的因素对模型的影响,模型只能以随机扰动项来表达,所以,当出现实际值超出或明显低于预测值的情况时,需要结合实际情况进一步分析,探索干预力度、疾病报告制度等因素是否发生重大变化。

本研究发现根据2007–2015年的数据建立预测模型预测2016年的全国包虫病月报告病例数时预测相对误差为-13.97%,2007–2014年数据建立的预测模型预测2015年的全国月报告病例数时相对误差为12.99%。尚未看到原始数据积累越多、预测时间越短则预测效果越好的情况。这一现象提示并不能简单地认为构建ARIMA模型的原始数据越多,构建出的模型就越优,模型的预测效果就越好^[9-10]。

ARIMA在急性传染病的短期预测中效果较好,但包虫病是慢性疾病,病例的发现主要与查病工作开展的情况有关,主动检查的越多,发现的病例数就越多,全国各地往往有先重后轻的查病工作安排,导致全国包虫病月报告病例数不能完全反映我国包虫病的流行态势和患病程度。如2007年和2008年,我国加强了包虫病的防治力度,进行了全国规模的包虫病普查工作,包虫病例数报告很高。同理,在2016年由于全国包虫病工作力度的加大导致病例数较高,2017年又政策性地加强了对西藏地区包虫病防治的工作,预计2017年全国报告病例数也相对较高,即包虫病的报告病例数与防控工作的力度有密切关联,这在一

定程度上限制了ARIMA模型的应用。加强全国包虫病主动监测,尽可能多地发现病例,依此建立的ARIMA模型预测效果可能会更好。

模型的建立和预测是个动态过程,需不断探索新的预测方法,如在人工智能时代,新算法(如支持向量机、卷积神经网络、深度置信网络、随机森林算法等)的引入和应用可提高预测精度^[11-13],从而为包虫病等传染病防控提供科学参考。

【参考文献】

- [1] 雷正龙,王立英.全国重点寄生虫病防治形势与主要任务[J].中国寄生虫学与寄生虫病杂志,2012,30(1):1-5.
- [2] 买买提江·吾买尔,阿斯亚·阿西木,王端明,等.新疆维吾尔自治区北部地区6~12岁儿童棘球蚴病血清学调查[J].中国寄生虫学与寄生虫病杂志,2017,35(3):246-249.
- [3] 买买提江·吾买尔,伊斯拉音·乌斯曼,阿迪力·司马义,等.新疆维吾尔自治区动物棘球蚴感染调查分析[J].中国寄生虫学与寄生虫病杂志,2017,35(2):145-149.
- [4] 王乐.新疆包虫病定点医院包虫病病人经济负担分析[D].乌鲁木齐:新疆医科大学,2010.
- [5] 贾铁武,周晓农.疾病负担(DALY)的评价与应用[J].中国寄生虫学与寄生虫病杂志,2005,23(5):304-308.
- [6] 张梦媛,伍卫平,官亚宜,等.藏区肝棘球蚴病患者手术医疗费用及其相关影响因素分析[J].中国寄生虫学与寄生虫病杂志,2017,35(3):250-253.
- [7] 王东,冯宇,李凡,等.甘肃省藏区人群棘球蚴病流行现状调查及分析[J].中国寄生虫学与寄生虫病杂志,2017,35(2):140-144.
- [8] 谭恩丽,侯慧玉,包海荣,等.采用自回归移动平均模型预测中国流感病例数[J].病毒学报,2017,33(5):699-705.
- [9] 张顺先,邱琪,王英.我国手足口病重症患者数自回归移动平均模型预测研究[J].病毒学报,2017,33(1):77-81.
- [10] 侯慧玉,宋华芹,张顺先,等.我国疟疾报告病例数自回归移动平均模型预测研究[J].中国血吸虫病防治杂志,2017,29(4):436-440,458.
- [11] 张晓晓,庞婷.浅谈人工智能特点及在医学中应用的几个阶段[J].福建电脑,2017,33(3):106,161.
- [12] 陆家发,张国明,陈安琪.基于深度学习的疾病诊断[J].医学信息学杂志,2017,38(4):39-43.
- [13] 唐燕,王苹.随机森林算法在中医药院校贫困生认定预测中的应用研究[J].中国医药导报,2017,14(14):164-168.

【收稿日期】2017-09-26 【编辑】朱宏儒