

2011—2018 年北京市门诊儿童轮状病毒 检出率的季节特征分析及模型预测

田祎, 严寒秋, 高志勇, 刘白薇, 陈艳伟, 李伟红, 贾蕾, 王全意

北京市疾病预防控制中心/北京市预防医学研究中心, 北京 100013

摘要: **目的** 分析 2011—2018 年北京市门诊儿童轮状病毒检出率的季节性分布规律, 探讨合适的预测模型。 **方法** 采用圆形分布法对 2011 年 7 月—2018 年 6 月北京市门诊儿童轮状病毒检出率的时间分布特征进行分析, 应用 Eviews 9.0 构建和优化 SARIMA 模型, 模型预测 2018 年 7—12 月北京市门诊儿童轮状病毒检出率, 并与实际值进行比较验证模型。

结果 2011—2018 年北京市门诊儿童轮状病毒的检出存在季节性 ($M=0.59$), 除 2012—2013 年度外, 流行趋势基本相同。高峰集中在冬春季 (11 月—次年 1 月), 每年均存在高峰日和高峰期, 但各年的高峰日不全相同 ($F=9.23, P<0.05$)。最优模型为 SARIMA (1, 1, 1) × (1, 0, 2)₁₂, 预测的平均绝对误差为 5.65%。 **结论** 2011—2018 年冬春季北京市门诊儿童轮状病毒检出率较高, SARIMA (1, 1, 1) × (1, 0, 2)₁₂ 模型拟合效果较好, 可用于北京市门诊儿童轮状病毒检出率的预测。

关键词: 轮状病毒; 圆形分布法; SARIMA 模型; 预测

中图分类号: R512.5 **文献标识码:** A **文章编号:** 1006-3110(2019)12-1430-05 **DOI:** 10.3969/j.issn.1006-3110.2019.12.006

Seasonal characteristics and model prediction of rotavirus detection rate among outpatient children in Beijing, 2011–2018

TIAN Yi, YAN Han-qiu, GAO Zhi-yong, LIU Bai-wei, CHEN Yan-wei, LI Wei-hong, JIA Lei, WANG Quan-yi

Beijing Center for Disease Control and Prevention/Beijing Research Center for Preventive Medicine, Beijing 100013, China

Corresponding author: WANG Quan-yi, E-mail: bjcdcxm@126.com

Abstract: **Objective** To analyze the seasonal distribution regularity of rotavirus detection rate among outpatient children in Beijing from 2011 to 2018, and to explore suitable prediction model. **Methods** The round distribution method was adopted to analyze the seasonal distribution characteristics of rotavirus detection rate among outpatient children in Beijing from July 2011 to June 2018. Eviews 9.0 software was used to build and optimize the ARIMA model. The model was used to forecast rotavirus detection rate among outpatient children in Beijing from July to December in 2018, and the model was verified through comparison of the predictive value and the actual value. **Results** There were seasonal characteristics in rotavirus detection among outpatient children in Beijing in 2011–2018 ($M=0.59$), and the epidemic tendency was basically similar except in 2012–2013. The peak season appeared in winter and spring (from November to January next year). There were peak days and peak period in each year, but the peak days were not all the same in each year ($F=9.23, P<0.05$). The optimal model was SARIMA (1, 1, 1) × (1, 0, 2)₁₂, and the average absolute prediction error was 5.65%. **Conclusions** The detection rates of rotavirus among outpatient children in Beijing in 2011–2018 were found to be higher in winter and spring. The established SARIMA (1, 1, 1) (1, 0, 2)₁₂ model is of good fitting effect, and it is suitable for predicting the detection rate of rotavirus among outpatient children in Beijing.

Key words: rotavirus; round distribution; SARIMA model; prediction

轮状病毒感染是全球儿童重症腹泻的常见病因, 是 5 岁以下儿童死亡的重要原因之一^[1-2]。我国轮状病毒感染以散发为主, 冬春季高发, 不同地区发病高峰不同, 同一地区不同年份的发病高峰也可能存在不

同^[3-4]。了解传染病的发病高峰, 拟合并运用适当的模型可更有效地预测和防控传染病的流行。本次研究旨在运用圆形分布分析 2011—2018 年北京市门诊儿童轮状病毒检出率的季节高峰, 同时结合季节性自回归移动平均模型 (seasonal auto regressive integrated moving average, SARIMA), 确定预测轮状病毒检出率趋势的最优模型, 为北京市儿童轮状病毒的防控提供科学依据。

基金项目: 国家科技重大专项 (2018ZX10714002-003-002)

作者简介: 田祎 (1987-), 女, 河北省任丘市人, 硕士, 主管医师, 研究方向: 传染病监测与预测预警。

通信作者: 王全意, E-mail: bjcdcxm@126.com。

1 资料与方法

1.1 资料来源 本研究资料来源于 2011 年 7 月—2018 年 12 月,以北京市三家门诊量最大的儿童医院为基础的门诊监测数据。肠道门诊监测的对象为急性腹泻病例,定义为 24 h 内出现腹泻 3 次及以上、5 岁及以下儿童病例,排除异地就医儿童病例及复诊病例。监测哨点医院每月至少采集腹泻儿童患者粪便标本 15 份,建议每月第 1、2、3 周各采样 5 份。由医院负责采集病例便标本及患儿基本信息,区级疾病预防控制中心进行信息复核及样本检测,北京市疾病预防控制中心负责复核样本检测。酶联免疫吸附测定法用于轮状病毒的检测。

1.2 分析方法 根据每月肠道门诊的监测病例总数和轮状病毒检测阳性数计算出检出率,检出率=轮状病毒阳性数/监测总数。纳入 2011 年 7 月—2018 年 6 月北京市门诊儿童轮状病毒检出率,使用 Excel 2010 建立分月报告病例信息数据库,按照每月 45 例平均采样量为标化数,计算出标化后的轮状病毒阳性病例数,进行时间的圆形分布分析。通过 Excel 2010 编制公式^[5],依次求出各年以及总的集中度 M、平均角 $\bar{\alpha}$ 、角度标准差 S,以及角度离散程度指标 γ ,运用 Raleigh's -test 检验平均角是否有统计学意义,有意义时由 $(\bar{\alpha} \pm S)$ 推算出流行高峰期。然后,应用 Watson-Williams 检验^[6]方法对不同年份门诊急性腹泻儿童患者季节高峰的差异进行假设检验。

1.3 模型构建 乘积季节模型 $ARIMA(p,d,q) \times (P,D,Q)_s$ 是在 ARIMA 模型的基础上,进一步考虑季节性因素。其中 p 、 d 和 q 分别表示自回归阶数、非季节差分次数和移动平均项数, P 、 D 分别表示季节性自回归和季节差分阶数, Q 表示季节性移动平均阶数, S 表示季节周期。模型的建模过程主要包括序列平稳性检验、模型识别、模型的参数估计、模型的检验及优化。应用 Eviews 9.0 构建和优化 SARIMA 模型。模型识别是模型构建的关键步骤之一,参数 d 和 D 可以根据序列达到平稳时的差分次数确定, p 、 q 和 P 、 Q 值可由自相关函数和偏自相关函数得到, p 、 q 和 P 、 Q 一般不超过 2 阶,故可以分别取 0、1、2,由低阶到高阶逐个进行试验,然后根据模型的确定系数 R^2 、AIC 最小准则、残差白噪声检验以及系数的显著性进行综合判断,以确定最优模型。使用 2018 年 7 月—2018 年 12 月北京市门诊儿童轮状病毒检出率进行模型验证。

2 结果

2.1 轮状病毒检出时间趋势 2011—2018 年北京市

门诊急性腹泻儿童患者中轮状病毒检出率时间分布情况见图 1。每年的检出高峰为冬春季,即当年的 11 月、12 月以及次年的 1 月检出率最高。但 2012—2013 年与其他年度不同,在 2013 年的 4 月和 5 月再次出现一个小高峰。

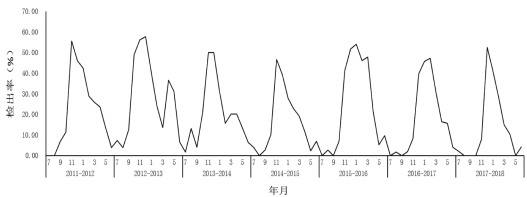


图 1 2011—2018 年北京市门诊儿童轮状病毒分月检出率

2.2 圆形分布结果

2.2.1 集中度和 Raleigh-test 检验平均角 2011—2018 年,分年集中度分别为 0.61、0.47、0.51、0.62、0.67、0.75 和 0.76,除 2012—2013 年度外,均大于 0.5,提示轮状病毒在北京市的检出呈较强的季节性。近 7 年北京市轮状病毒检出率合计,集中度为 0.59,亦提示有较强的季节性。Raleigh-test 检验平均角的结果显示,2011—2018 年北京市门诊急性腹泻儿童患者中轮状病毒的检出每年均存在高峰日和高峰期 ($P < 0.05$),7 年合计也存在高峰日和高峰期,见表 1。

表 1 Raleigh-test 检验 2011—2018 年北京市门诊腹泻儿童轮状病毒检出率平均角

年份	检出数	高峰角 (α°)	角标准差 (S)	角度离散程度 (γ)	Z 值	P 值
2011—2012	117	9.67	65.41	0.52	31.78	<0.01
2012—2013	152	-10.04	83.01	0.35	18.63	<0.01
2013—2014	112	-7.81	75.65	0.42	19.60	<0.01
2014—2015	88	-0.45	62.65	0.55	26.62	<0.01
2015—2016	129	23.70	57.35	0.61	47.38	<0.01
2016—2017	96	36.50	53.59	0.65	40.03	<0.01
2017—2018	75	18.30	46.57	0.72	38.74	<0.01
合计	769	11.81	67.00	0.50	195.89	<0.01

2.2.2 不同年份轮状病毒检出高峰的差异 2011—2018 年不同年份北京市门诊急性腹泻儿童患者中轮状病毒检出高峰日(高峰期)分别为 1 月 6 日(11 月 2 日—3 月 13 日)、12 月 17 日(9 月 25 日—3 月 11 日)、12 月 19 日(10 月 5 日—3 月 5 日)、12 月 27 日(10 月 26 日—2 月 28 日)、1 月 2 日(11 月 24 日—3 月 19 日)、2 月 2 日(12 月 11 日—3 月 28 日)、1 月 14 日(11 月 29 日—3 月 2 日),合计的高峰日为 1 月 8 日(11 月 2 日—3 月 16 日)。应用 Watson-Williams 检验法,检验各年的高峰日平均角是否存在差异,结果显示 2011—2018 年,北京市轮状病毒检出的高峰日不全相同,总体存在差异 ($F=9.23, P<0.05$)。

2.3 轮状病毒检出率模型预测 对 2011—2017 年北京市门诊急性腹泻儿童患者中轮状病毒分月检出率序列做相关图发现 ACF 呈拖尾, 衰减缓慢, ADF 单位根检验结果显示, 在 5% 水平下不能拒绝序列存在单位根的原假设, 即序列不平稳 ($t = -0.69, P = 0.41$)。对序列进行一次普通差分后, 再次做相关图和平稳性检验, 在 5% 水平下拒绝序列存在单位根的原假设, 即序列平稳 ($t = -8.12, P < 0.01$)。

原始序列经过一次普通差分后成为平稳序列, 因此 $d=1, D=0$ 。 p, q, P, Q 分别取 0、1、2, 由低阶至高阶拟合 $ARIMA(p, 1, q) \times (P, 0, Q)_{12}$ 模型。首先, 选择 $R^2 > 0.5$ 且 $P < 0.05$ 的模型, 见表 2。然后, 进一步使用 Breusch-Godfrey LM 检验进行残差白噪声检验, 判断所建模型的优劣。LM 检验的原假设为: 直至确定的滞后阶数, 残差序列无相关性, 此时对应的 F 值无统计学意义, P 值 > 0.05 。确定系数 R^2 越大, AIC 统计量越小, 且 LM 检验 P 值 > 0.05 的为最优模型, 故 $ARIMA(1, 1, 1) \times (1, 0, 2)_{12}$ 判定为较优模型。

$ARIMA(1, 1, 1) \times (1, 0, 2)_{12}$ 模型残差的 Ljung-Box 统计结果显示, 延迟 6 阶、12 阶、18 阶、24 阶时, 统计量 Q 均无统计学意义, 对应的 P 值均 > 0.05 , 最优模型的残差序列自相关和偏自相关系数均在 95% CI 内 (见图 2), 可认为所建模型的残差序列为白噪声, 表明所选模型是恰当的。 $ARIMA(1, 1, 1) \times (1, 0, 2)_{12}$ 模型对北京市门诊儿童轮状病毒检出率拟合及预测图见图 3。

应用 $ARIMA(1, 1, 1) \times ((1, 0, 2)_{12})$ 模型, 预测 2018 年 7—12 月北京市门诊急性腹泻儿童患者中轮状病毒的月检出率, 并与实际值进行比较, 预测绝对误

差比较结果见表 3。



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.093	0.093	0.6276	0.428
		2	0.142	0.135	2.1249	0.346
		3	-0.033	-0.058	2.2053	0.531
		4	0.038	0.027	2.3139	0.678
		5	-0.106	-0.102	3.1832	0.672
		6	-0.180	-0.179	5.7266	0.454
		7	0.056	0.126	5.9814	0.542
		8	-0.196	-0.189	9.1035	0.334
		9	-0.055	-0.054	9.3536	0.405
		10	-0.035	0.048	9.4594	0.489
		11	0.047	-0.012	9.6490	0.562
		12	0.121	0.139	10.923	0.536
		13	-0.057	-0.102	11.208	0.593
		14	-0.022	-0.139	11.251	0.666
		15	-0.008	0.081	11.258	0.734
		16	0.110	0.085	12.379	0.717
		17	-0.090	-0.121	13.154	0.726
		18	-0.007	0.026	13.159	0.782
		19	-0.039	-0.084	13.307	0.822
		20	-0.031	0.006	13.404	0.859
		21	-0.018	0.068	13.438	0.892
		22	0.093	0.047	14.347	0.889
		23	0.049	-0.032	14.604	0.908
		24	-0.010	0.011	14.616	0.931
		25	-0.025	-0.055	14.685	0.948
		26	-0.115	-0.124	16.206	0.931
		27	0.013	0.055	16.227	0.948
		28	-0.036	-0.038	16.379	0.960
		29	-0.062	-0.064	16.852	0.964
		30	-0.059	0.025	17.296	0.969
		31	0.027	0.006	17.390	0.977
		32	-0.086	-0.158	18.371	0.974

图 2 $ARIMA(1, 1, 1) \times (1, 0, 2)_{12}$ 的残差序列自相关和偏自相关图

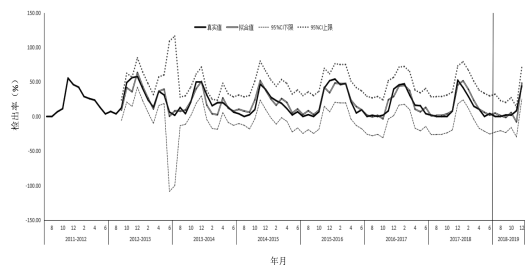


图 3 $ARIMA(1, 1, 1) \times ((1, 0, 2)_{12})$ 模型

对北京市门诊儿童轮状病毒检出率拟合及预测图

表 2 备选模型拟合优度比较、残差检验和参数估计

模型	拟合度		残差检验		参数估计			
	R^2	AIC	F(2)	P	参数	系数	t	P
$ARIMA(0, 1, 0) \times (0, 0, 2)_{12}$	0.51	7.46	0.83	0.44	MA(24)	0.87	38.01	0.00
$ARIMA(0, 1, 0) \times (1, 0, 2)_{12}$	0.71	6.96	1.54	0.22	AR(12)	0.75	8.58	0.00
					MA(24)	-0.89	-44.95	0.00
$ARIMA(0, 1, 0) \times (2, 0, 1)_{12}$	0.60	7.19	0.74	0.48	AR(24)	0.34	2.95	0.01
					MA(12)	0.87	28.54	0.00
$ARIMA(0, 1, 0) \times (2, 0, 2)_{12}$	0.55	7.31	0.16	0.85	AR(24)	0.52	4.39	0.00
					MA(24)	-0.86	-28.48	0.01
$ARIMA(0, 1, 1) \times (1, 0, 2)_{12}$	0.73	6.93	2.56	0.08	AR(24)	0.88	12.56	0.00
					MA(12)	-0.43	-3.94	0.00
					SMA(24)	-0.89	-42.25	0.00
$ARIMA(1, 1, 1) \times (0, 0, 1)_{12}$	0.55	7.44	1.78	0.18	AR(1)	0.58	6.40	0.00
					MA(1)	-0.99	-88.32	0.00
					SMA(12)	0.89	47.09	0.00

续表 2

模型	拟合度		残差检验		参数估计			
	R ²	AIC	F(2)	P	参数	系数	t	P
ARIMA(1,1,1)×(1,0,2) ₁₂	0.74	6.90	2.83	0.07	AR(1)	0.49	4.88	0.00
					SAR(12)	0.62	6.16	0.00
					MA(1)	-0.99	-137.88	0.00
					SMA(24)	0.88	39.99	0.00
ARIMA(1,1,2)×(1,0,2) ₁₂	0.74	6.92	0.01	0.99	AR(1)	-0.31	-2.55	0.01
					SAR(12)	0.86	11.67	0.00
					MA(1)	-0.32	-2.72	0.01
					SMA(24)	-0.89	-41.63	0.00
ARIMA(2,1,0)×(0,0,1) ₁₂	0.54	7.45	1.24	0.29	AR(2)	0.62	6.86	0.00
					MA(2)	-0.98	-117.18	0.00
					SMA(12)	0.90	48.42	0.00

注:F(2)表示滞后取2时,LM 检验的 F 统计量

表 3 2018 年 7—12 月北京市门诊儿童轮状病毒
分月检出率预测值与实际值比较

月份	实际值(%)	预测值(%)	绝对误差(%)
7	0	4.88	4.88
8	0	1.23	1.23
9	2.13	-1.43	3.56
10	1.79	5.71	3.92
11	8.33	-7.93	16.26
12	44.23	48.30	4.07
平均值			5.65

3 讨论

圆形分布可研究疾病的昼时性或季节性,计算可通过应用 Excel 2010 提供的函数实现,方法简便易行。此次研究结果显示,除 2012—2013 年度外,2011—2018 年北京市门诊儿童轮状病毒检出率的分年集中度以及近 7 年发病合计的集中度均大于 0.5,提示轮状病毒在北京市门诊儿童中的发病呈较强的季节性。据以往研究结果显示,2012 年和 2013 年的轮状病毒阳性率高于 2011 年,且在 2012—2013 年,北京市门诊儿童轮状病毒的主要流行基因型由 G3P[8]转变为 G9P[8]^[7],而此次研究结果显示 2012—2013 年度的发病集中度为 0.47,小于其他年份,提示轮状病毒主要流行株的改变会影响轮状病毒检出的季节性分布。

儿童轮状病毒感染与气候因素存在很强的相关性,在寒冷的季节轮状病毒阳性率较高^[8]。有文献报道,温度平均每升高 1℃,轮状病毒的发病率就会降低 10%^[9]。Raleigh-test 检验平均角的结果显示,2011—2018 年北京市门诊急性腹泻儿童患者中轮状病毒的检出每年及 7 年合计均存在高峰日和高峰期,高峰期为每年 11 月—次年 3 月,与蒋翠莲等^[10]的研究结果相似。但是,2012 年和 2013 年北京市门诊急性腹泻儿童患者中轮状病毒的检出率呈双高峰,与乌鲁木齐地

区^[11]的流行趋势相似,而且与北京市其他年份比,该年度高峰开始的比较早,分别为 9 月 25 日和 10 月 5 日,这也可能与 2012—2013 年轮状病毒主要流行株的改变相关。

据 WHO 统计,轮状病毒每年在世界范围内导致超过 50 万婴幼儿死亡,其中 85%以上的死亡病例来自非洲和亚洲的低收入国家^[12]。降低传染病导致的疾病负担,除了推行疫苗外,运用时间序列分析,挖掘前、后数据的相关性,找出传染病在时间上的变化规律,提前对传染病的发生、流行趋势以及高峰期作出预测,及时预警是最有效地防控方法^[13]。SARIMA 模型综合考虑了序列的趋势、周期变化,以及随机干扰等,同时在建模时将季节因素纳入,用模型参数将以上因素进行量化,能较好的拟合时间序列趋势和季节效应^[14],但需要多个季节周期的数据才可进行之后的预测,适用于短期预测。此次研究选择了 7 年的数据进行预测,结果显示,ARIMA(1,1,1)×(1,0,2)₁₂为较优模型,应用 2018 年 7—12 月的轮状病毒检出率进行模型验证,结果提示预测平均绝对误差为 5.65%。SARIMA 模型模型拟合时只纳入了时间序列和季节的影响因素,但其实际预测结果还会很大程度上受到监测资料质量、就医可及性、人群易感性、病毒变异等多因素影响^[15],故拟合值与实际值比较时,会存在绝对误差。

此次研究存在一定的局限性,选择的哨点医院以及监测样本量均较小,有待于扩大样本量进一步研究。同时,此次研究采用是北京市肠道门诊病毒性腹泻常规监测数据,有部分患者患病后并未及时就医,故该数据仅能反映轮状病毒在门诊腹泻儿童中的检出情况,不能完全真实地反映发病情况。故在实际应用中在保证有足够长的监测序列的前提下,应提高监测数据报