

# 基于中断时间序列设计的分段回归模型评价 抗菌药物管理防控院感的效果\*

山西医科大学卫生统计教研室(030001) 赵执扬 刘 静 史 楠 康建邦 段金菊 乔宇超 仇丽霞<sup>△</sup>

**【摘要】目的** 应用中断时间序列设计的分段回归模型分析抗菌药物管理防控院感的长短期效果,为公共卫生干预措施的效果评价提供方法学上的参考。**方法** 收集山西省某三甲医院 2017 年 7 月至 2020 年 6 月抗菌药物管理实施前后住院患者院内感染肺炎克雷伯菌(*klebsiella pneumoniae*, KPN)数据,采用调整异常值的中断时间序列的泊松分段回归模型进行分析。**结果** 实施抗菌药物管理后短期干预效果较好,干预之后感染 KPN 的风险是干预前的 81.25% ( $P < 0.05$ ),即感染 KPN 的风险下降了 18.75%;但长期干预效果较差,时间每增加一个月,感染风险会增加 1.55%;因新冠疫情期间重症患者占比大,院内 KPN 感染率升高,感染风险增加了 68.81%。**结论** 实施抗菌药物管理,仅在短期内降低了 KPN 感染的风险,需要长期严格执行才能持久地发挥作用。考虑了异常值的中断时间序列设计的分段回归模型既可以定量的评价短期、长期的干预效果,又可以对异常值导致的风险进行分析,为公共卫生干预措施的效果评价提供方法学上的参考。

**【关键词】** 中断时间序列设计 分段回归模型 抗菌药物管理 院内感染 克雷伯菌肺炎

**【中图分类号】** R195.1 **【文献标识码】** A **DOI** 10.11783/j.issn.1002-3674.2024.06.005

## Segmented Regression Model based on Interrupted Time Series Design to Evaluate the Effect of Antimicrobial Management on the Prevention and Control of Nosocomial Infection

Zhao Zhiyang, Liu Jing, Shi Nan, et al (Department of Health Statistics, Shanxi Medical university, Taiyuan 030001)

**【Abstract】 Objective** The segmented regression model designed by interruption time series was used to analyze the short-term and long-term effects of antimicrobial management and prevention of nosocomial infection, so as to provide methodological reference for the effect evaluation of public health interventions. **Methods** The data of nosocomial infection with *Klebsiella pneumoniae* (KPN) in a third class hospital in Shanxi Province from July 2017 to June 2020 were collected and analyzed by Poisson piecewise regression model with interrupted time series adjusted for outliers. **Results** The short-term intervention effect was better after the implementation of antibiotics management. The risk of KPN infection after the intervention was 81.25% before the intervention ( $P < 0.05$ ), and the risk of infection with *Klebsiella pneumoniae* decreased by 18.75%. However, the effect of long-term intervention is poor. For each additional month, the risk of infection will increase by 1.55%. Because of the large proportion of serious ill patients during the COVID-19 period, the KPN infection rate in the hospital increased, and the infection risk increased by 68.81%. **Conclusion** The implementation of antimicrobial drug management only reduces the risk of KPN infection in the short term, which needs to be strictly implemented in the long term to play a lasting role. Considering the interruption time series of outliers, the piecewise regression model can not only quantitatively evaluate the short-term and long-term intervention effects, but also analyze the risks caused by outliers, so as to provide methodological reference for the effect evaluation of public health interventions.

**【Key words】** Interrupt time series analysis; Piecewise regression model; Antibiotics management; Nosocomial infection; *Klebsiella pneumoniae*

公共卫生干预措施常常在某一群体中开展,采用自身前后对照,以时间序列的感染率、患病率等作为效果评价指标,时间序列数据可能存在季节周期性特征,而且可能受到其他突发事件的影响并有异常值的存在。过去研究人员通常将实施干预前作为对照组、实施干预后作为观察组,采用  $T$  检验、卡方检验等统计学方法进行分析<sup>[1-2]</sup>。然而,这会忽视季节周期性或异常值的影响,且无法客观地反映公共卫生干预措施的长、短期效果。中断时间序列设计(interrupted time series analysis, ITSA)是一种在明确的时间点来评估干

预措施有效性的准实验设计,近年来,中断时间序列设计常被用于定量地评价公共卫生措施效果<sup>[3-5]</sup>。分段回归分析是在中断时间序列研究中评价干预效果的一种强有力的统计方法,可以根据不同的数据类型构建不同的分段回归模型,如线性分段回归模型、泊松分段回归模型等<sup>[6]</sup>,将干预效果定量地分为短期和长期效果,通过自相关性检验来识别时间序列的季节自相关性,通过设置虚拟变量来调整数据的异常值,从而构建最优模型<sup>[7]</sup>。

本研究以山西省某三甲医院在 2019 年 1 月 1 日起开展的抗菌药物管理项目<sup>[8]</sup>为例,以干预前后该院每月住院患者人数、院内感染肺炎克雷伯菌(*Klebsiella pneumoniae*, KPN)<sup>[9]</sup>病例数作为评价指标,以泊松

\* 基金项目:国家自然科学基金资助项目(81973155);山西省科技厅重大专项(201803D31066)

<sup>△</sup>通信作者:仇丽霞, E-mail: qlx\_1126@163.com

回归构建分段回归模型<sup>[6]</sup>,通过调整数据的异常值,拟合得到最优模型,从而评价抗菌药物管理项目控制院内感染 KPN 的长、短期效果,为公共卫生干预措施的效果评价提供方法学上的参考。

### 资料与方法

#### 1.资料来源

从医院信息系统中提取出山西省某三甲医院 2017 年 7 月至 2020 年 6 月的住院患者资料,包括每月住院患者人数、入院 48 小时后的 KPN 感染人数;抗菌药物管理实施的时间为 2019 年 1 月 1 日,干预前阶段为 2017 年 7 月至 2018 年 12 月、干预后阶段为 2019 年 1 月至 2020 年 6 月。本研究的 KPN 月感染率均小于 0.05,为罕见事件,因此以泊松回归来建立分段回归模型;由于每月住院患者人数(观察单位数)不同,因此,通过在泊松分段回归模型中引入偏移量(offset)来抵消观察单位数不同对结果造成的影响<sup>[10]</sup>。

#### 2.研究方法

##### (1)时间序列图和序列的平稳性检验

使用时间序列图描述 2017 年 7 月至 2020 年 6 月 KPN 感染率的变动趋势。中断性时间序列分析要求序列平稳<sup>[11]</sup>,时间序列的平稳性采用 ADF 检验、自相关图判断,ADF 检验的 P 值小于 0.05 代表序列平稳,自相关系数快速衰减为 0 或在某阶后变为 0 可判断为平稳序列<sup>[12]</sup>。

##### (2)中断时间序列设计及分段回归模型

中断时间序列设计通过统计方程来评估干预措施对时间序列数据趋势的影响。其关键假设是干预前的数据趋势可以外推到未发生干预时的趋势,先采用干预前的时间序列建立一个潜在趋势,这种潜在的趋势被一个已知时间点的干预“打断”,如果干预没有发生,那么潜在的趋势将保持不变,这种情况被称为“反事实”,通过比较“反事实”与干预后实际的变化趋势,来评估干预措施的效果<sup>[6]</sup>。

分段回归模型是适合中断时间序列设计的统计方法,用于评估干预措施的瞬时效果和长期效果。分段回归中的“段”是指时间序列被干预措施分割成的部分。分段回归模型包括趋势参数、水平参数改变量、趋势参数改变量,见图 1。趋势参数指每一段回归模型的斜率,水平参数改变量指干预前后两段回归模型在干预前一刻的指标估计值之差,趋势参数改变量指干预前后两段斜率之差,通过分析干预前后水平参数和趋势参数的改变量,来评估干预的效果<sup>[7]</sup>。

泊松分段回归模型构建如下:

$$\text{模型 1: } \log(y_i) = \log(n_i) + \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$$

$$\text{或者 } \log(y_i/n_i) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$$

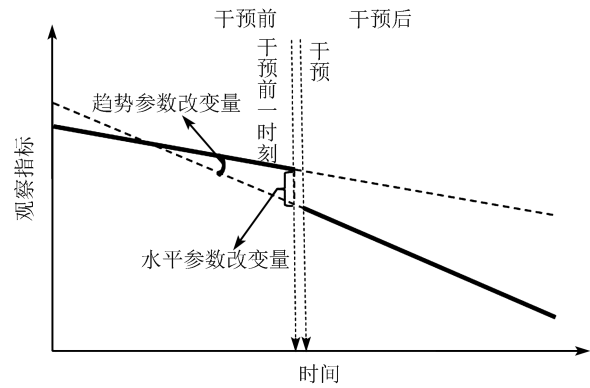


图 1 中断时间序列设计水平和趋势参数改变量图示

其中,  $X_1$ 、 $X_2$ 、 $X_3$  分别表示时间变量(本例依次为 1~36)、干预变量(干预前为 0,干预后为 1)、干预后时间变量(干预前为 0,干预后依次为 1~18)。因变量  $y_i$  为 KPN 每月的感染人数,  $n_i$  为每月的住院患者人数,由于每月住院患者人数不等,引入  $\log(n_i)$  作为偏移量来消除观察单位数不等对结果造成的影响;  $y_i/n_i$  为每月 KPN 感染率。  $\beta_0$  为干预前的截距,表示实施抗菌药物管理前的基线水平;  $\beta_1$  为干预前的趋势参数,即实施抗菌药物管理前回归线的斜率,  $\varepsilon$  表示随机误差;干预前  $X_2$ 、 $X_3$  均为 0,因此干预前的模型为:

$$\text{模型 2: } \log(y_i) = \log(n_i) + \beta_0 + \beta_1 X_1 + \varepsilon$$

模型 1 中,  $\beta_2$  为水平参数的改变量,即干预前后两段回归模型在干预前一刻指标估计值之差,即干预引起的指标值水平的变化,用于评价干预措施的短期效应;  $\beta_3$  为趋势参数的改变量,即实施抗菌药物管理前后斜率之差,用于评价干预措施的长期效应;  $\beta_1 + \beta_3$  为干预后的趋势参数,即实施抗菌药物管理后回归线的斜率;干预后  $X_2 = 1$ 、 $X_3 = X_1 - 18$ ,因此干预后的模型为:

$$\text{模型 3: } \log(y_i) = \log(n_i) + \beta_0 + \beta_1 X_1 + \beta_2 + \beta_3 (X_1 - 18) + \varepsilon = \log(n_i) + (\beta_0 + \beta_2 - 18\beta_3) + (\beta_1 + \beta_3) X_1 + \varepsilon$$

干预前最后一个月  $i = 18$ ,即  $X_1 = 18$ ,根据干预前模型 2 计算得到  $\log(y_{18})$  的估计值为  $\log(n_{18}) + \beta_0 + 18\beta_1 + \varepsilon$ ;而根据干预后模型 3 计算得到此月份  $\log(y_{18})$  的估计值为  $\log(n_{18}) + \beta_0 + \beta_2 + 18\beta_1 + \varepsilon$ ;两个估计值之差为  $\beta_2$ ,因此,  $\beta_2$  为干预后水平参数的改变量。

由于干预后的时间序列数据恰好包括了我国新冠疫情最严重的两个月,受疫情影响,轻、中度患者尽可能居家治疗,导致 2020 年 2 月、3 月住院患者重症比例较高,因此,出现了 KPN 感染率异常升高的现象。当时间序列存在已知原因的异常值,则引入控制异常值的虚拟变量  $X_4$ ,异常值出现时点设置为 1,其余时点设置为 0,分段回归模型可记为:

$$\text{模型 4: } \log(y_i) = \log(n_i) + \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$$

$$\text{或者 } \log(y_i/n_i) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$$

其中  $\beta_4$  表示在新冠疫情期间,住院患者重症比例

大引起的水平改变量,即在干预之后,有无异常值时的截距之差。

中断时间序列设计的部分数据及变量赋值见表 1。

表 1 中断时间序列设计的部分数据及变量赋值

时间	KPN 感染人数	住院患者人数	感染率 (‰)	$X_1$	$X_2$	$X_3$	$X_4$
2017 年 7 月	49	4408	11.12	1	0	0	0
2017 年 8 月	44	4611	9.54	2	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2018 年 12 月	46	5558	8.28	18	0	0	0
2019 年 1 月	47	5760	8.16	19	1	1	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2020 年 1 月	49	5430	9.02	31	1	13	0
2020 年 2 月	30	1711	17.53	32	1	14	1
2020 年 3 月	49	3372	14.53	33	1	15	1
2020 年 4 月	36	5712	6.30	34	1	16	0
2020 年 5 月	46	4610	9.98	35	1	17	0
2020 年 6 月	55	5245	10.49	36	1	18	0

若时间序列不平稳,存在季节自相关,则应用正、余弦变化来进行调整<sup>[13]</sup>,引入 $X_5$ 、 $X_6$ 两变量,其中 $X_5 = \sin(2\pi t_i/12)$ 、 $X_6 = \cos(2\pi t_i/12)$ ,泊松分段回归模型为:

$$\text{模型 5: } \log(y_i) = \log(n_i) + \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \varepsilon$$

$$\text{或者 } \log(y_i/n_i) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \varepsilon$$

其中 $t_i$ 为各年度的月份, $i = 1, 2, \dots, 12$ , $X_5$ 、 $X_6$ 两变量的系数分别为 $\beta_5$ 、 $\beta_6$ 。

(3) 模型拟合效果评价

采用 AIC、BIC 指标评价,两个指标越小,模型拟合效果越好。

(4) 统计分析软件

采用 excel 2016 对数据进行预处理,采用 anaconda 平台的 Python 3.9 进行平稳性检验,采用 SAS 9.4 来建立分段回归模型;采用 Wald 检验进行回归系数的检验,检验水准  $\alpha = 0.05$ 。

结 果

1. KPN 感染率的时间序列图

由 KPN 感染率的时间序列图(图 2)可知,干预前 KPN 感染率处于波动状态,月均感染率为 9.47‰,干预前后 7 个月(2019 年 1 月至 7 月)KPN 感染率呈现下降趋势,每月感染率均低于干预前,而在 2019 年 8 月开始升高,至 2020 年 1 月的 6 个月间 KPN 感染率基本上升到干预前的水平。由于 2020 年 2 月、3 月为我国新冠疫情最严重的两个月,受疫情影响,轻、中度患者尽可能居家治疗,住院患者中重症比例较高,因此,出现了 KPN 感染率异常升高的现象。2020 年 4 月后国内疫情得以控制,山西没有出现新发病例,医院全面开诊,住院患者中轻、中症比例增大,KPN 感染率迅速下降,基本达到了疫情前的水平。

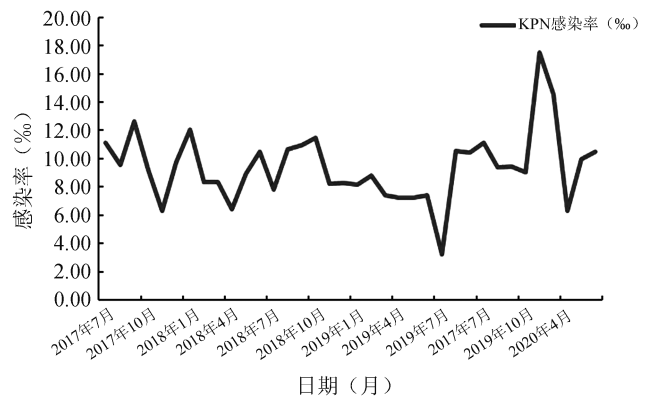


图 2 2017—2020 年 KPN 感染率时序图

2. 平稳性检验

本文对每月 KPN 感染率进行了 ADF 平稳性检验。由表 2 可知:感染率的 ADF 检验统计量为 -4.5945、P 值小于 0.05,自相关图(图 3)显示自相关系数在 1 阶后快速衰减,以上结果说明该序列为平稳序列,本文不需要针对季节自相关性进行正、余弦变化。

表 2 ADF 检验

检验方法	显著水平			检验统计量	P
	1%	5%	10%		
ADF	-3.6327	-2.9485	-2.6130	-4.5945	0.0001

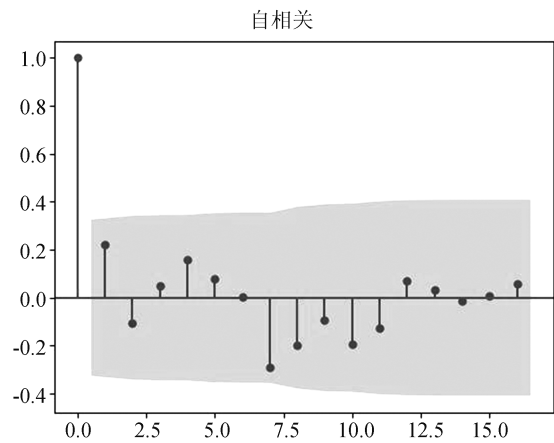


图 3 感染率自相关图

3. 模型异常值的调整及参数估计

(1) 异常值调整前后模型拟合效果比较

由于 KPN 感染率在 2020 年 2 月、3 月存在已知原因的异常值,因此需要引入控制异常值的虚拟变量,由表 3 可知,调整前模型的 AIC 为 301.10、BIC 为 307.44、MAPE 为 20.08%,调整后模型的 AIC 为 287.30、BIC 为 295.21、MAPE 为 18.39%,调整了异常值后模型的 AIC、BIC、MAPE 减小,模型拟合效果得到了改善。

表 3 不同模型的拟合效果

模型	AIC	BIC	MAPE (%)
调整前	301.10	307.44	20.08
调整后	287.30	295.21	18.39

(2) 泊松分段回归模型的参数估计

本研究构建的泊松分段回归模型为模型 4:  $\log$

$(y_i/n_i) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$ 。由表 4 分段回归模型参数估计结果可知:仅 $\hat{\beta}_1$ 没有显著性,其他 4 个参数估计值均有统计学差异( $P < 0.05$ )。结合干预前后实际感染率的对数值与估计感染率对数值的曲线图(图 4)可知:在 2019 年抗菌药物管理实施前 KPN 感染率随时间变化的趋势不存在统计学意义, $\hat{\beta}_1 = -0.0056, P = 0.4062, RR = 99.45\%$  (95% CI: 98.15% ~ 100.76%);从干预后的短期效果来看,水平改变量的参数估计值 $\hat{\beta}_2 = -0.2076, P = 0.0369$ ,即干预前后两段回归模型在干预前一刻 KPN 感染率的对数值之差,其差值为 $-0.2076$ ,说明 2019 年抗菌药物管理实施后的短时间内感染 KPN 的风险是下降的,是干预前的 81.25% (95% CI: 66.85% ~ 98.75%),比干预前下降了 18.75%,短期干预效果明显;从长期效果来看,干预后斜率的参数估计值 $\hat{\beta}_1 + \hat{\beta}_3 = 0.0154, RR = \exp 0.0154 = 1.0155$ ,即干预之后时间每增加一个月, KPN 感染率的对数值增加 0.0154,感染 KPN 的风险增加 1.55%,说明干预后感染 KPN 的风险随着时间的增加而增加。干预前的 KPN 感染率平均估计值为 9.48%,干预后的前 13 个月(2019 年 1 月至 2020 年 1 月) KPN 感染率估计值均低于干预前水平,而从干预后的第 14 个月(2020 年 2 月)开始 KPN 感染率估计值反弹回干预前的水平。 $\hat{\beta}_3 = 0.0210, P = 0.0323$ ,有显著性,说明干预后回归线斜率比干预前增大了 0.0210,其  $RR = 1.0212$  (95% CI: 1.0018 ~ 1.0410);新冠疫情期间由于住院患者重症比例的增加,引起的水平改变量的参数估计值 $\hat{\beta}_4 = 0.5236, P < 0.0001$ ,即在干预后,在出现异常值期间 KPN 感染率对数值的截距比未出现异常值时增加

表 4 调整后分段回归模型的参数估计结果

变量	$\hat{\beta}$	标准误	$c^2$ 值	P 值	RR 值	RR 95% CI	
						下限	上限
常数项	-4.6059	0.0740	3872.95	<0.0001	0.0100	0.0086	0.0116
$X_1$	-0.0056	0.0067	0.69	0.4062	0.9945	0.9815	1.0076
$X_2$	-0.2076	0.0995	4.35	0.0369	0.8125	0.6685	0.9875
$X_3$	0.0210	0.0098	4.58	0.0323	1.0212	1.0018	1.0410
$X_4$	0.5236	0.1244	17.71	<0.0001	1.6881	1.3228	2.1544

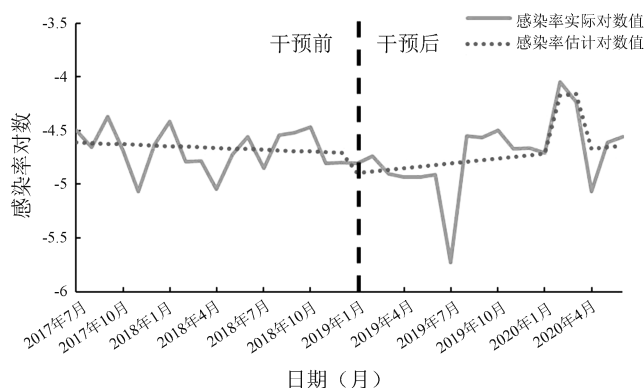


图 4 干预前后感染率实际对数值与估计对数值

了 0.5236,出现异常值期间感染 KPN 的风险是其他时间的 1.6881 倍(95% CI: 1.3228 ~ 2.1544),增加了 68.81%。

将实际感染率的对数值与估计感染率的对数值进行指数转换,得到实际感染率与估计感染率,时间序列图见图 5。

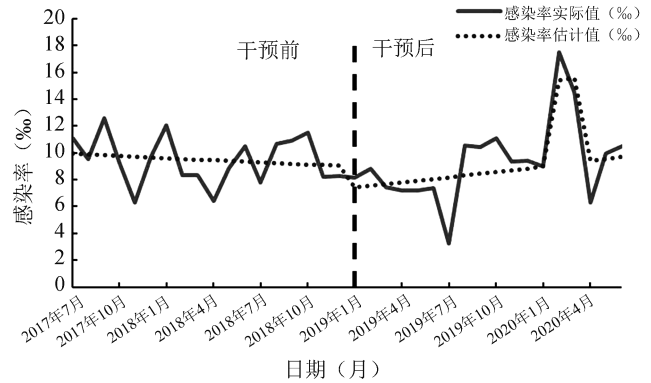


图 5 干预前后感染率的实际值与估计值

### 讨 论

本研究介绍了中断时间序列设计、分段回归模型,并以山西省某三甲医院实施抗菌药物管理项目为例,应用泊松回归来构建分段回归模型,探讨了长短期干预效果以及出现异常值的解决方法,为公共卫生干预措施的效果评价提供了有力的方法学工具。

中断时间序列设计可以通过对干预前后的两段时间分别拟合模型从而控制研究指标在干预前的变化趋势,通过斜率、水平改变量更加客观地反映了干预的长、短期效果<sup>[4, 14]</sup>。本研究中实施抗菌药物管理项目后短期效果较好,而长期效果较差,这可能是由于实施干预措施仅在短期内严格执行而没有长期坚持。中断时间序列设计要求序列平稳,当序列不平稳(存在季节自相关)时需要模型进行校正,当拟合线性分段回归模型时,可使用 Maximum Likelihood、差分等方法进行校正<sup>[3, 14]</sup>;当拟合泊松分段回归模型时,可使用正余弦变化、傅里叶变化等方法进行校正<sup>[13]</sup>。在中断时间序列设计中可能受到其他突发事件的影响而产生异常值,当这些异常值可以人为解释时,可以通过设置虚拟变量来进行处理,从而构建最优模型。

本研究采用中断时间序列虽然能有效的分析出干预的长、短期效果,但是也存在一定的局限性。本文通过设置虚拟变量来调整模型,从而减少了数据异常值对结果的干扰,然而这种调整发挥了多大的作用有待于进一步的研究。

### 参 考 文 献

[1] 陈喜瑞. 消毒供应中心集中管理模式对医院感染防控效果的影响[J]. 现代诊断与治疗, 2021, 32(16): 2650-2652.