

DOI:10.12138/j.issn.1671-9638.20252077

· 论著 ·

青海省 2014—2023 年细菌性痢疾流行特征及预测模型初探

姜雨淇^{1,2},赵金华^{1,2},龙江³,张阳¹,邓萍¹,覃胜林¹,张华一⁴

(1. 青海大学医学院公共卫生系,青海 西宁 810001; 2. 青海省疾病预防控制中心业务部,青海 西宁 810007; 3. 重庆市疾病预防控制中心传染病预防控制所,重庆 北碚 400700; 4. 青海省疾病预防控制中心传染病预防控制所,青海 西宁 810007)

[摘要] 目的 比较五种时间序列模型,预测 2024 年青海省细菌性痢疾月发病率,为防控提供参考依据。方法 分析 2014—2023 年青海省细菌性痢疾流行特征,应用 R4.3.1 软件建立季节性自回归移动平均(SARIMA)模型、三次指数平滑法(Holt-Winters)模型、指数平滑(ETS)模型、神经网络自回归(NNAR)模型、指数平滑空间状态(TBATS)模型,分析模型拟合效果并比较其准确度。结果 2014—2023 年青海省累计报告细菌性痢疾 5 833 例,无死亡病例,男女性别比为 1.23:1。2016 年报告发病率最高(15.45/10 万),2023 年报告发病率最低(3.68/10 万),2014—2016 年上升,之后下降,总体下降趋势明显。5 岁以下年龄组病例数最多,占总数的 29.76%(1 736 例)。人群分布中构成比居前 3 位的分别为幼托机构及散居儿童(35.56%)、农民(24.65%)、学生(12.62%)。除 Holt-Winters 相加模型外,其余四种模型预测趋势均与实际一致;其中拟合效果最好的是 ETS 模型,整体表现较为均衡(训练集:MAE = 0.13、RMSE = 0.21、MAPE = 19.55%;测试集:MAE = 0.11、RMSE = 0.16、MAPE = 28.66%),建议基于 ETS 模型对青海省细菌性痢疾发病率进行预测。结论 2014—2023 年青海省细菌性痢疾总体呈下降趋势,6—8 月为流行高峰,托幼及散居儿童为高危人群。五种预测模型中 ETS 模型的拟合效果最好,后续可基于 ETS 模型对细菌性痢疾发病率进行预测。

[关键词] 细菌性痢疾;时间序列模型;ETS 模型;流行特征;预测

[中图分类号] R181.3⁺2

Epidemiological characteristics and prediction model of bacillary dysentery in Qinghai Province, 2014—2023

JIANG Yuqi^{1,2}, ZHAO Jinhua^{1,2}, LONG Jiang³, ZHANG Yang¹, DENG Ping¹, QIN Shenglin¹, ZHANG Huayi⁴(1. Department of Public Health, Medical College of Qinghai University, Xining 810001, China; 2. Department of Business, Qinghai Provincial Center for Disease Prevention and Control, Xining 810007, China; 3. Institute for Infectious Disease Control and Prevention, Chongqing Provincial Center for Disease Control and Prevention, Beibei 400700, China; 4. Institute for Infectious Disease Control and Prevention, Qinghai Provincial Center for Disease Prevention and Control, Xining 810007, China)

[Abstract] **Objective** To compare five time series models and predict the monthly incidence of bacillary dysentery in Qinghai Province in 2024, and provide reference for the prevention and control. **Methods** The epidemic characteristics of bacterial dysentery in Qinghai Province from 2014 to 2023 were analyzed. R4.3.1 software was used for establishing seasonal autoregressive integrated moving average (SARIMA) model, Holt-Winters triple exponential smoothing (Holt Winters) model, exponential smoothing (ETS) model, neural network autoregression (NNAR)

[收稿日期] 2025-01-21

[基金项目] “昆仑英才·高原名医”项目(青卫健办[2021]104 号)省市级;2024 年度公共卫生人才培养支持项目

[作者简介] 姜雨淇(2001-),女(汉族),浙江省杭州市人,硕士研究生在读,主要从事传染病监测预警研究。赵金华共同第一作者。

[通信作者] 龙江 E-mail: 68803648@163.com

model, and trigonometric seasonality, Box-Cox transformation, ARMA errors, trend and seasonal components (TBATS) model. Fitting effect of the models was analyzed and accuracy was compared. **Results** From 2014 to 2023, a total of 5 833 cases of bacterial dysentery were reported in Qinghai Province, without deaths, male to female ratio being 1.23:1. The highest incidence was reported in 2016 (15.45 per 100 000 people), and the lowest incidence was reported in 2023 (3.68 per 100 000 people). Incidence increased from 2014 to 2016, then decreased, showing an obvious overall downward trend. Case number in <5 years age group was the highest, accounting for 29.76% of the total cases ($n=1736$). Regarding population distribution, the top three were children in childcare institutions and scattered children (35.56%), farmers (24.65%), and students (12.62%). Except the additive Holt-Winters model, the predicted trends of the other four models were consistent with actuality. The ETS model had the best fitting effect, with a relatively balanced overall performance (training set: MAE = 0.13, RMSE = 0.21, MAPE = 19.55%; testing set: MAE = 0.11, RMSE = 0.16, MAPE = 28.66%). It is recommended to predict the incidence of bacillary dysentery in Qinghai Province based on ETS model. **Conclusion** From 2014 to 2023, bacterial dysentery in Qinghai Province showed a downward trend, with the peak of the epidemic from June to August. Preschool and scattered children were high-risk groups. Among the five prediction models, ETS model has the best fitting effect, and can be used to predict the incidence of bacillary dysentery.

[Key words] bacterial dysentery; time series model; ETS model; epidemic characteristics; predict

细菌性痢疾(bacillary dysentery)是由志贺菌属引起的肠道传染病,经消化道感染后导致发热或腹泻,严重者可出现全身毒血症或中毒性脑病^[1]。随着致病菌耐药性增强,全球每年约有 2.7 亿病例^[2]。2005—2018 年全国约 333 万例细菌性痢疾病例,造成重大社会经济负担^[3-4]。目前我国细菌性痢疾多为散发,尚无有效疫苗预防,且常年位居法定传染病报告数量前 5 位,防控形势严峻。

季节性自回归移动平均模型(seasonal autoregressive integrated moving average, SARIMA)通过分解时间序列季节性,使模型更好地预测周期性疾病^[5];三次指数平滑法(Holt-Winters)模型包括累加和累乘两种形式,特别适用于中短期预测;指数平滑(exponential smoothing, ETS)模型、神经网络自回归(neural network autoregression, NNAR)模型和指数平滑空间状态(trigonometric seasonality, Box-Cox transformation, ARMA errors, trend and seasonal components, TBATS)模型更依赖于指数加权,赋予近期观测数据更高的权重,兼具全期平均和移动平均优势,在细菌性痢疾等周期性疾病的预测中发挥重要作用^[6]。

青海省位于我国西部,平均海拔为 4 058.40 m,其高海拔低氧环境可能影响人体免疫系统功能。为系统分析青海省细菌性痢疾的流行特征及预测趋势,本研究采用 SARIMA、Holt-Winters、ETS、NNAR、TBATS 五种时间序列模型,通过比较均方根误差(root mean square error, RMSE)、平均绝对误差(mean absolute error, MAE)和平均绝对百分

比误差(mean absolute percentage error, MAPE)指标,为青海省细菌性痢疾防控提供科学依据^[7-8]。

1 资料与方法

1.1 资料来源 2014—2023 年青海省细菌性痢疾疫情资料来源于中国疾病监测管理信息系统,经各级疾病预防控制中心专业人员逐级审核,以“发病日期”为统计标准导出个案信息及发病率数据,筛选现住址为青海省的细菌性痢疾病例;人口数据来自青海省统计局。

1.2 模型建立 SARIMA 模型参数表示为 SARIMA(p, d, q)(P, D, Q)m,通过时序图可视化和单位根检验评估序列平稳性,若序列非平稳则需进行常规差分(阶数为 d),此过程迭代直至差分后序列平稳以确定最终 d 值;若存在季节性波动则额外进行 D 阶季节性差分。Holt-Winters 模型参数包括水平平滑、趋势平滑、季节平滑三个平滑参数,需估计各分量初始值。NNAR 模型基于前馈神经网络,核心参数为输入层节点数(即用于预测的滞后观测值数量,通常表示为 p 或包含季节性滞后的 p, P)和隐藏层节点数(k)。ETS 模型参数包括基本分量结构及相应的平滑参数和初始状态值。TBATS 模型采用傅里叶级数建模复杂季节性,通过 Box-Cox 变换处理异方差性。所有模型主要使用 R 语言中 tidyverse、forecast 和 forecastHybrid 等程序包构建。

1.3 统计学方法 采用 EXCEL 2019 软件整理青海省 2014—2023 年月发病数据,以 2014 年 1 月—

2022 年 12 月为训练集,2023 年 1—12 月为测试集;采用 Origin 2024 软件对性别和职业分布进行可视化,运用描述性流行病学方法分析流行特征;所有预测模型通过 R4.3.1 构建,检验水准 $\alpha = 0.05$ 。

2 结果

2.1 流行概况 2014—2023 年青海省累计报告细菌性痢疾 5 833 例,无死亡病例。年报告发病率从 2014 年的 14.57/10 万缓慢上升至 2016 年 15.45/10 万,增幅 6.04%;2016 年为转折点,从最高点(2016 年)下降至 2023 年的 3.68/10 万,降幅达 76.18%。总体下降趋势较明显,2023 年发病率(3.68/10 万)较 2014 年(14.57/10 万)下降 74.74%。病例中男性 3 216 例,女性 2 617 例,性别比 1.23:1,男女发病率整体均呈下降趋势。见图 1。

年龄分布显示,各年龄组均有病例发生。5 岁以下组发病最多(1 736 例,占 29.76%),其次为 5—9 岁组(595 例,占 10.20%), ≥ 85 岁组最少(46 例,占 0.79%)。发病率随年龄增长呈现先下降后上升的趋势。

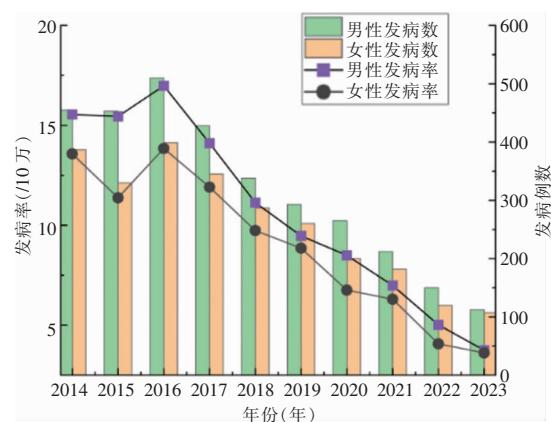


图 1 2014—2023 年青海省细菌性痢疾性别分布图

Figure 1 Gender distribution of bacterial dysentery in Qinghai Province from 2014 to 2023

人群分布显示,幼托及散居儿童占比最高(2 074 例,35.56%),农民次之(1 438 例,24.65%),学生居第三位(736 例,12.62%)。2016 年幼托及散居儿童占比达峰值(382 例,42.02%),2021 年降至最低(106 例,26.90%)。各年份职业构成比虽有波动,但幼托及散居儿童始终居首位。见图 2。

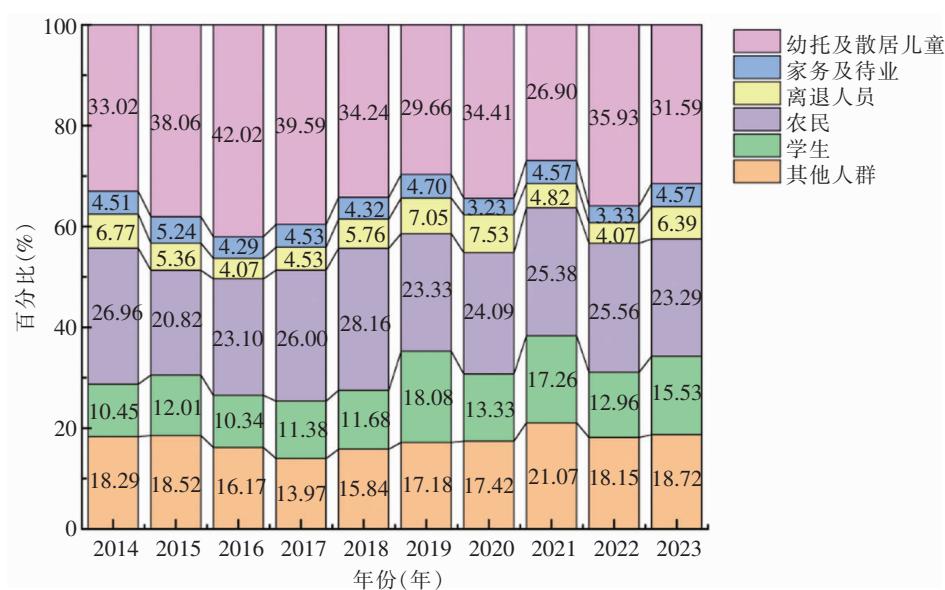


图 2 2014—2023 年青海省细菌性痢疾职业人群百分比条图

Figure 2 Percentage bar chart of population with bacterial dysentery in Qinghai Province from 2014 to 2023

2.2 模型建立

2.2.1 序列平稳化判断 2014—2023 年细菌性痢疾发病呈现明显季节性波动,每年 7—8 月为流行高峰;2016 年 8 月达峰值(3.26/10 万),2022 年 12 月为谷值(0.08/10 万)。使用 R4.3.1 软件绘制的时

间序列图显示明显的短期趋势和季节性趋势,见图 3。时间序列分解显示,发病数随季节成分波动,总体呈下降趋势见图 4。经扩展迪基-富勒检验(augmented Dickey-Fuller test, ADF)显示,时间序列具有平稳性($t = -8.26, P = 0.01$)。

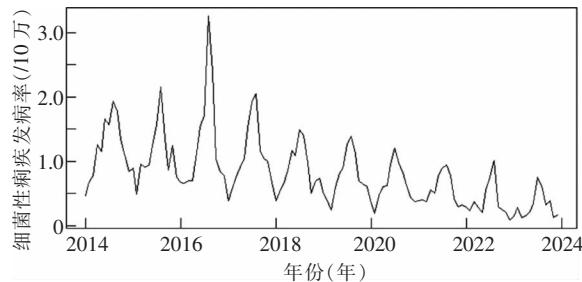


图 3 2014—2023 年青海省细菌性痢疾时间序列图

Figure 3 Time series chart of bacterial dysentery in Qinghai Province from 2014 to 2023

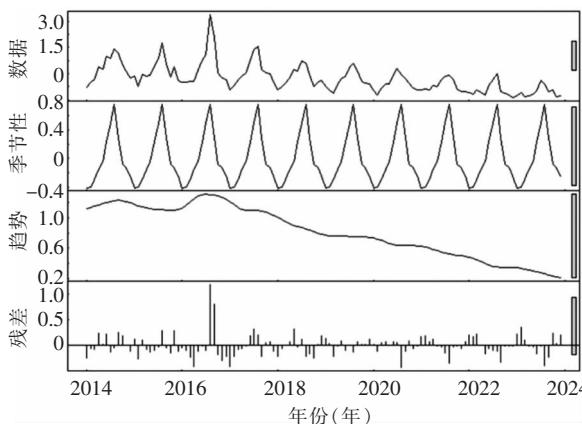


图 4 2014—2023 年青海省细菌性痢疾季节趋势分解图

Figure 4 Decomposition map of seasonal trend of bacterial dysentery in Qinghai Province from 2014 to 2023

2.2.2 模型效果评价 以 2014 年 1 月—2022 年 12 月作为训练集,2023 年 1 月—2023 年 12 月为测试集,结果显示,除 Holt-Winters 相加模型之外,其余五种模型拟合趋势与 2023 年实际情况基本一致。其中 ETS 模型整体表现最好。在训练集中平衡了各项指标,测试集中除 RMSE 值 $>$ NNAR 和 SARIMA 模型外,其余各种指标均是最小(训练集 MAE = 0.13、RMSE = 0.21、MAPE = 19.55%; 测试集 MAE = 0.11、RMSE = 0.16、MAPE = 28.66%),为最佳模型,具体指标结果见表 1。五种时间序列模型的总体预测比较汇总见图 5。

2.3 基于 ETS 模型的预测 根据五种预测模型拟合结果显示,ETS 模型不需要固定的时间序列长度,能更好地适应数据长度不同的时间序列,效果最好,因此基于 ETS 模型预测 2024 年青海省细菌性痢疾。结果显示,2024 年青海省细菌性痢疾相较于 2023 年有所下降,6—8 月为发病高峰,其余月份仍有散在发生,见图 6,应有针对性的制定策略,有效应对细菌性痢疾的流行。

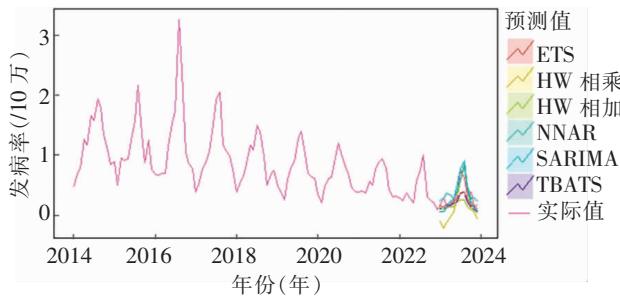


图 5 2014—2023 年青海省细菌性痢疾五种模型预测结果图

Figure 5 Prediction results of five models for bacterial dysentery in Qinghai Province from 2014 to 2023

表 1 2014—2023 年青海省细菌性痢疾五种模型准确度比较

Table 1 Comparison of accuracy of five models for bacterial dysentery in Qinghai Province from 2014 to 2023

数据集	模型	MAE	RMSE	MAPE(%)
训练集	SARIMA	0.16	0.23	22.09
	Holt-Winters 相加模型	0.18	0.26	27.83
	Holt-Winters 相乘模型	0.17	0.24	24.74
	ETS	0.13	0.21	19.55
	NNAR	0.14	0.21	23.38
	TBATS	0.14	0.21	19.94
测试集	SARIMA	0.11	0.13	40.40
	Holt-Winters 相加模型	0.20	0.23	91.72
	Holt-Winters 相乘模型	0.17	0.22	46.01
	ETS	0.11	0.16	28.66
	NNAR	0.12	0.14	55.34
	TBATS	0.12	0.17	35.51

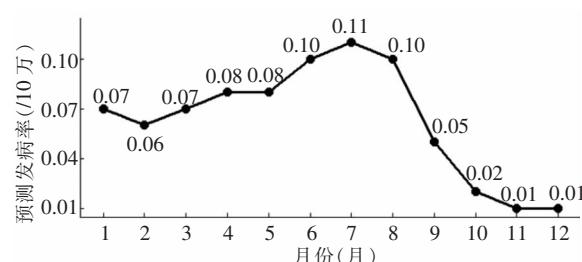


图 6 基于 ETS 模型预测 2024 年青海省细菌性痢疾结果图

Figure 6 Prediction results of bacterial dysentery in Qinghai Province in 2024 based on ETS model

3 讨论

2014—2023 年青海省细菌性痢疾累计报告 5 833 例。2016 年为转折点,此前呈缓慢上升趋势(2016 年发病率较 2014 年上升了 6.04%),之后转

为下降(2023 年较 2016 年下降了 76.18%)。总体呈现明显下降趋势,2023 年发病率较 2014 年总体下降 74.74%,由 14.57/10 万降至 3.68/10 万。这反映了政府及疾控部门防控措施有效,人群防病能力增强,细菌性痢疾防控工作取得明显成效^[9]。然而,青海省细菌性痢疾发病率仍高于湖北^[10]、云南^[11]等省份,这可能与其地处高原,部分偏远地区饮水设施落后有关,亟需完善供水设施建设,以减少细菌性痢疾等水源传播疾病的发生。同时,主要发病人群为幼托机构儿童及散居儿童,这与国内其他报道^[12-13]一致。低年龄儿童因免疫系统发育尚不完善、尚未养成良好卫生饮食习惯,发病风险较高。因此,仍需加强患者及带菌者的及时发现与隔离。

国内在细菌性痢疾预警研究中应用时间序列模型相对较少,多数研究仅基于描述性分析或单一模型的预测^[14]。新疆自治区^[15]运用 ARIMA 模型发现其拟合效果较好;深圳市福田区^[16]通过描述性分析及耐药性监测,显示细菌性痢疾仍具较强致病性。本研究采用五种预测模型进行相互验证,在相当程度上降低了模型过度拟合的风险,提升了预测结果的稳健性与准确性,为青海省细菌性痢疾的监测预警与防控工作提供了坚实的学科基础和参考依据。

本研究中,除 Holt-Winters 相加模型拟合效果相对较差外,其他模型均能较好反映青海省 2014—2023 年细菌性痢疾的流行特征及规律,其中,ETS 模型整体表现最优(训练集:MAE = 0.13、RMSE = 0.21、MAPE = 19.55%)。这可能与 ETS 模型擅长短期预测并能有效分解时间序列数据有关^[17]。同时,不同模型指标结果的差异说明不能局限于单一模型预测,而应综合多个模型结果进行判断。预测结果显示,2024 年青海省细菌性痢疾发病率将比 2023 年有所下降,但仍呈现出 6—8 月为高峰的典型季节性特征。

由于细菌性痢疾感染后获得的免疫力时间短,且病原菌可通过物体表面间接传播,人群一旦感染,常需投入大量人力物力进行防控^[11]。本研究运用多种预测模型对青海省细菌性痢疾进行预测,具有一定的创新性,但也存在局限性。如未将人口密度、地理因素等影响因素作为协变量纳入模型,可能限制了预测结果的准确性和外推性。未来研究可尝试纳入时空地理加权回归等综合模型,以进一步提升预测精度。

利益冲突:所有作者均声明不存在利益冲突。

参 考 文 献

- Spiga L, Jimenez AG, Santos RL, et al. How microbiological tests reflect bacterial pathogenesis and host adaptation[J]. Braz J Microbiol, 2021, 52(4): 1745—1753.
- Lu T, Das S, Howlader DR, et al. *Shigella* vaccines: the continuing unmet challenge[J]. Int J Mol Sci, 2024, 25(8): 4329.
- 黄兴凤. 环境气象因素对细菌性痢疾发病的影响机制及其归因风险[D]. 桂林: 桂林电子科技大学, 2023.
Huang XF. Influence mechanism and attribution risk of environmental meteorological factors on the incidence of bacillary dysentery[D]. Guilin: Guilin University of Electronic Technology, 2023.
- Zhu YL, Lin SH, Dong SB, et al. Incidence and trends of 17 notifiable bacterial infectious diseases in China, 2004—2019 [J]. BMC Infect Dis, 2023, 23(1): 273.
- Yang PY, Cheng PF, Zhang N, et al. Statistical machine learning models for prediction of China's maritime emergency patients in dynamic: ARIMA model, SARIMA model, and dynamic Bayesian network model[J]. Front Public Health, 2024, 12: 1401161.
- Al Mobin M, Kamrujjaman M. Downscaling epidemiological time series data for improving forecasting accuracy: an algorithmic approach[J]. PLoS One, 2023, 18(12): e0295803.
- Tesfaye S, Regassa F, Beyene G, et al. Spatiotemporal analysis and forecasting of lumpy skin disease outbreaks in Ethiopia based on retrospective outbreak reports[J]. Front Vet Sci, 2024, 11: 1277007.
- Zenia S, L'Hadj M, Selmane S. A hybrid approach based on seasonal autoregressive integrated moving average and neural network autoregressive models to predict scorpion sting incidence in El Oued province, Algeria, from 2005 to 2020[J]. J Res Health Sci, 2023, 23(3): e00586.
- Alphonse N, Odendall C. Animal models of shigellosis: a historical overview[J]. Curr Opin Immunol, 2023, 85: 102399.
- 蔡晶, 黄淑琼, 张鹏, 等. 2005—2021 年湖北省细菌性痢疾流行特征分析及短期发病预测[J]. 中华疾病控制杂志, 2022, 26(11): 1303—1308.
Cai J, Huang SQ, Zhang P, et al. Epidemiological analysis and short-term prediction of bacillary dysentery in Hubei Province from 2005 to 2021[J]. Chinese Journal of Disease Control & Prevention, 2022, 26(11): 1303—1308.
- 张荣兵, 田荣, 何继波, 等. 2014—2022 年云南省细菌性痢疾流行病学特征分析[J]. 中华疾病控制杂志, 2024, 28(6): 641—645.
Zhang RB, Tian R, He JB, et al. Epidemiological characteristics of bacterial dysentery in Yunnan Province from 2014 to 2022[J]. Chinese Journal of Disease Control & Prevention, 2024, 28(6): 641—645.
- 田祎, 黄瑛, 陈艳伟, 等. 2016—2019 年北京市细菌性痢疾流行病学特征分析[J]. 首都公共卫生, 2022, 16(4): 244—

248.

Tian Y, Huang Y, Chen YW, et al. Epidemiological characteristics analysis of bacillary dysenteric cases in Beijing, 2016 – 2019[J]. Capital Journal of Public Health, 2022, 16(4): 244 – 248.

[13] 崔伟红, 吕燕灵, 刘海韵, 等. 2007—2021 年山东省烟台市戊型肝炎发病特征及预测分析[J]. 疾病监测, 2023, 38(8): 929 – 933.

Cui WH, Lv YL, Liu HY, et al. Incidence characteristics and prediction of hepatitis E in Yantai, Shandong, 2007 – 2021 [J]. Disease Surveillance, 2023, 38(8): 929 – 933.

[14] 王婷, 贺湘焱. 细菌性痢疾流行特征及发病预测的研究进展 [J]. 职业与健康, 2024, 40(10): 1422 – 1425.

Wang T, He XY. Research progress on epidemiological characteristics and incidence prediction of bacillary dysentery[J]. Occupation and Health, 2024, 40(10): 1422 – 1425.

[15] 王婷, 贺湘焱. 基于季节自回归移动平均模型的新疆细菌性痢疾发病趋势预测[J]. 公共卫生与预防医学, 2023, 34(5): 30 – 34.

Wang T, He XY. Seasonal autoregressive moving average model-based prediction of bacteriophage dysentery incidence trends in Xinjiang[J]. Journal of Public Health and Preventive Medicine, 2023, 34(5): 30 – 34.

[16] 张桂欣, 许玉成, 周杰, 等. 2013—2022 年深圳市福田区细菌

性痢疾流行特征及病原学特性分析[J]. 应用预防医学, 2024, 30(2): 91 – 94.

Zhang GX, Xu YC, Zhou J, et al. Analysis on the epidemiological characteristics and etiological characteristics of bacillary dysentery in Futian district, Shenzhen from 2013 to 2022[J]. Applied Preventive Medicine, 2024, 30(2): 91 – 94.

[17] Maipan-Uku JY, Cavus N. Forecasting tuberculosis incidence: a review of time series and machine learning models for prediction and eradication strategies[J]. Int J Environ Health Res, 2025, 35(3): 645 – 660.

(本文编辑:左双燕)

本文引用格式: 姜雨淇, 赵金华, 龙江, 等. 青海省 2014—2023 年细菌性痢疾流行特征及预测模型初探[J]. 中国感染控制杂志, 2025, 24(10): 1389 – 1394. DOI: 10.12138/j.issn.1671 – 9638.20252077.

Cite this article as: JIANG Yuqi, ZHAO Jinhua, LONG Jiang, et al. Epidemiological characteristics and prediction model of bacillary dysentery in Qinghai Province, 2014 – 2023[J]. Chin J Infect Control, 2025, 24(10): 1389 – 1394. DOI: 10.12138/j.issn.1671 – 9638.20252077.