

应用灰色模型GM(1,1)预测上海市副溶血性弧菌引起的食源性疾病发病率

陆冬磊, 吴春峰, 段胜刚, 罗宝章, 徐碧瑶, 刘弘

摘要: [目的] 建立灰色模型GM(1,1), 预测上海市副溶血性弧菌引起的食源性疾病的发病率。[方法] 利用2007—2010年上海市食源性疾病的监测数据, 运用Excel 2010, 建立GM(1,1)模型, 进行精度及适用性检验, 预测2011、2012年发病率并验证。[结果] GM(1,1)预测模型为 $X(t+1)=124.32e^{0.59t}-47.14$, 关联度 $r_t=0.614$, 后验差比值 $C=0.24$, 小误差概率 $P=1$, 发展系数 $-a=0.59$, 模型预测2011、2012年发病的平均相对误差分别为57.48%、3.90%。[结论] 建立的GM(1,1)模型精度好, 谨慎适用短期(2~3年)预测; 预测上海市副溶血性弧菌引起的食源性疾病的发病率呈上升趋势。

关键词: GM(1,1)模型; 副溶血性弧菌; 食源性疾病; 预测; 模型验证

Forecasting of Foodborne Disease Caused by *Vibrio parahaemolyticus* in Shanghai using Grey Model GM(1,1) LU Dong-lei, WU Chun-feng, DUAN Sheng-gang, LUO Bao-zhang, XU Bi-yao, LIU Hong (Food Hygiene and Foodborne Disease Control Department, Shanghai Institutes of Preventive Medicine, Shanghai 200336, China) · The authors declare they have no actual or potential competing financial interests.

Abstract: [Objective] To forecast the incidence of foodborne disease caused by *Vibrio parahaemolyticus* (VP) in Shanghai by establishing grey model GM(1,1). [Methods] The Shanghai foodborne disease surveillance data in 2007–2010 were processed by Excel 2010 and used to build GM(1,1) model. After tests of accuracy and applicability, the model was used to forecast the incidence rate of VP-related foodborne disease in 2011 and 2012 for verification. [Results] The GM(1,1) prediction model established was $X(t+1)=124.32e^{0.59t}-47.14$, r_t (relativity) = 0.614, C (posterior error ratio) = 0.24, P (small error probability) = 1, $-a$ (development coefficient) = 0.59. The average relative error of model forecast results versus actual data for 2011 and 2012 was 57.48% and 3.90% respectively. [Conclusion] The precision of the GM(1,1) model warrants application with caution in short-time (2–3 years) forecast. The predicted incidence rate of foodborne disease caused by *Vibrio parahaemolyticus* in Shanghai is increased according to the established model.

Key Words: GM(1,1) model; *Vibrio parahaemolyticus*; foodborne disease; forecast; model verification

副溶血性弧菌是常见的食源性疾病致病菌, 多污染水产品^[1]。有资料估计, 我国每年由副溶血性弧菌导致的食源性疾病约495.1万人次^[2]。副溶血性弧菌是上海市食物中毒的主要致病因素之一, 2000—2007年由该菌引起的食物中毒事件220起, 占事件总数的57.4%, 涉及病例6231人^[3]。上海市自2006年开始试点食源性疾病监测网络, 包括医院腹泻病例监测、社区居民急性胃肠炎监测等。本研究根据2007—2010年监测数据, 采用灰色模型GM(1,1)建立上海市副溶血性弧菌引起的食源性疾病发病率的估算预测模

型, 为评价上海市副溶血性弧菌引起的食源性疾病危害及采取相应控制措施提供数据参考。

1 材料与方法

1.1 获取原始数据序列

2007—2012年上海市副溶血性弧菌引起的食源性疾病发病率来源于上海市食源性疾病监测网络, 相应各年人口学数据来源于上海市统计年鉴。获取2007—2010年社区急性胃肠炎发病率、就诊率, 医院肠道门诊腹泻病例中副溶血性弧菌感染率, 估算2007—2010年上海市社区居民副溶血性弧菌引起食源性疾病的发病率 $X(t)$ ($t=1, 2, 3, \dots, n$), 设为原始数据序列。

1.2 数据处理与分析

采用Excel 2010进行数据录入与运算^[4]。

1.3 GM(1,1)预测模型建立方法和计算步骤

GM(1,1)预测模型的建立及计算参照文献^[5]。

DOI: 10.13213/j.cnki.jeom.2015.14607

[基金项目] 上海市卫生局青年科研项目(编号: 20114Y030)

[作者简介] 陆冬磊(1983—), 男, 学士, 主管医师; 研究方向: 食源性疾病预防控制; E-mail: ludonglei@scdc.sh.cn

[作者单位] 上海市预防医学研究院食品卫生食源性疾病预防科, 上海200336

1.3.1 计算累加数列 $Y(t)$ 根据灰色系统理论, 将无规律的原始数据按 $Y(t) = \sum X(t) (t=1, 2, 3, \dots, n)$ 累加生成成为较有规律的数列 $Y(t)$ 。

1.3.2 计算均值生成数列 $Z(t)$ 对累加生成数列按 $Z(t) = [Y(t) + Y(t-1)]/2 (t=2, 3, 4, \dots, n)$ 作移动平均生成均值数列 $Z(t)$ 。

1.3.3 建立 GM(1, 1) 模型 建立关于 $Y(t)$ 的一阶线性微分方程: $dY(t)/dt + aY(t) = \mu$, 此式即 GM(1, 1) 预测模型。对该微分方程求解得: $Y(t+1) = [X(t) - \mu/a]e^{-at} + \mu/a$ 。式中 a 为发展系数, μ 为灰作用量, 均为待求系数, 可由下式计算:

$$a = \{ (n-1) [-\sum_{t=2}^n X(t)Z(t)] + [\sum_{t=2}^n Z(t) \sum_{t=2}^n X(t)] \} / D,$$
$$\mu = \{ [\sum_{t=2}^n Z(t)] [-\sum_{t=2}^n X(t)Z(t)] + [\sum_{t=2}^n Z^2(t)] [\sum_{t=2}^n X(t)] \} / D,$$
$$D = (n-1) [\sum_{t=2}^n Z^2(t)] - [\sum_{t=2}^n Z(t)]^2.$$

1.4 模型检验

1.4.1 残差检验 对实际值与模型预测值的残差进行逐点检验。首先按模型计算 $\hat{X}(t+1)$, 将 $\hat{X}(t+1)$ 累减生成 $\hat{X}(t)$, 最后计算 $\hat{X}(t)$ 与原始数据序列 $X(t)$ 的绝对残差序列及相对残差序列并计算平均相对残差。

1.4.2 关联度检验 通过考察模型值曲线和建模序列曲线的相似程度进行检验。计算序列 $\hat{X}(t)$ 与 $X(t)$

的绝对残差序列 δt , 及 δt 的最小值和最大值 ($\min \delta t$ 和 $\max \delta t$)。计算关联系数 εt : $\varepsilon t = (\min \delta t + P \max \delta t) / (\delta t + P \max \delta t) (t=1, 2, 3, \dots, n, P=0.5)$; 计算关联度 r_i : $r_i = \frac{1}{n} \sum_{t=1}^n \varepsilon t$ 。计算出 $\hat{X}(t)$ 与原始序列 $X(t)$ 的关联系数, 然后算出关联度。根据经验, 关联度 >0.6 便是满意的。

1.4.3 后验差检验 S_1 为残差的标准差, S_2 为原始序列的标准差, $S_1 = \sqrt{\sum_{t=1}^n (\delta t - \bar{\delta})^2 / (n-1)}$; $S_2 = \sqrt{\sum_{t=1}^n [X(t) - \bar{X}]^2 / (n-1)}$ 。后验差比值 $C = S_1/S_2$, 小误差概率 $P = P\{|\delta(t) - \bar{\delta}| < 0.6745 S_2\}$ 。根据 C 值及 P 值的大小, 综合评价模型的拟合效果。若预测精度的等级达到一定的要求, 可应用预测模型进行外推预测。

1.5 预测效果评价

通过 $\hat{X}(t) = \hat{Y}(t) - Y(t-1) (t=2, 3, \dots, n)$ 得原始资料 $X(t)$ 的估计值, 按 $X(t)$ 和 $\hat{X}(t)$ 的残差大小对模型进行检验和效果评价。

2 结果

2.1 建立 GM(1, 1) 预测模型

对 2007—2010 年上海市副溶血性弧菌感染的食源性疾病病例数 $X(0)$ 建立 GM(1, 1) 模型, 结果见表 1。对参数进行最小二乘估计, 得参数计算值: $\alpha = -0.5908$, $\mu = 27.8095$ 。建立灰色预测模型: $Y(t+1) = 124.32e^{0.59t} - 47.14, t=1, 2, 3, \dots, n$ 。

表 1 2007—2010 年上海市副溶血性弧菌引起的食源性疾病发病率 ($/10^5$) GM(1, 1) 模型计算表

年份 Year	t	实际发病率 X_t Real incidence	发病率累加 Y_t Incidence accumulation	均值 Z_t Mean	均值平方 Z_t^2 Mean square	$X_t Z_t$	预测发病率累加 $\hat{Y}_t$ Forecast incidence accumulation	预测发病率 $\hat{X}_t$ Forecast incidence
2007	0	77.18	77.18	—	—	—	77.18	—
2008	1	133.78	210.96	144.0684	20755.70	19273.37	177.88	100.70
2009	2	170.26	381.22	296.0894	87668.96	50413.02	358.55	180.67
2010	3	375.59	756.81	569.0135	323776.40	213713.10	684.73	326.18

2.2 模型拟合优度检验

2.2.1 关联度检验 经计算, 关联度 $r_i = 0.614$, 满足分辨率 $P=0.5$ 时的检验准则关联度 >0.60 , 拟合效果为较满意。

2.2.2 后验差比值及小误差概率 $\delta t = \{9.05, 34.44, 25.38\}$, $S_1 = 25.24$, $S_2 = 106.44$, $0.6745 S_2 = 71.79$, 100% 的 $\delta t < 71.79$, 故小误差概率 $P=1$ 。后验差比值 $C = S_1/S_2 = 0.24$ 。参照模型精度等级表 (表 2), 可以判断模型的预测精度为 1 级, 可用于预测。

2.2.3 发展系数 发展系数 $-\alpha = 0.59$, 参照模型适用范围表 (表 3), 用 GM(1, 1) 模型作短期预测应十分谨慎。

表 2 GM(1, 1) 模型精确度判断等级

Table 2 Accuracy rank of GM(1, 1) model		
精度等级 (Accuracy rank)	C	P
1 级 (好) (Good)	≤ 0.35	≥ 0.95
2 级 (合格) (Qualified)	$(0.35, 0.5]$	$[0.80, 0.95)$
3 级 (勉强) (Possibly qualified)	$(0.5, 0.65]$	$[0.70, 0.80)$
4 级 (不合格) (Unqualified)	>0.65	<0.70

2.3 拟合效果检验

根据模型计算, 2008—2012 年副溶血性弧菌引起的食源性疾病发病率预测值分别为 $100.70/10^5$ 、 $180.67/10^5$ 、 $326.18/10^5$ 、 $588.88/10^5$ 、 $1063.16/10^5$, 平均相对误差为 21.07%。对 2011—2012 年发病率外推预测的平均相对误差为 30.69% (表 4)。

表3 GM(1,1)模型适用范围

Table 3 GM(1,1) model applicable range reference

发展系数- α Development coefficient	适用范围 Applicable range
≤ 0.3	GM(1,1)模型可用于长期预测 Suitable for long-term prediction
(0.3, 0.5]	GM(1,1)模型可用于短期预测, 中长期预测慎用 Suitable for short-term prediction, be cautiously used in long-term prediction
(0.5, 0.8]	用GM(1,1)模型作短期预测应十分谨慎 Be cautiously used in short-term prediction
(0.8, 1]	应采用残差修正GM(1,1)模型 Should use residue modified GM(1,1) model
>1	不宜用GM(1,1)模型 Unsuitable to use GM(1,1) model

表4 2011—2012年上海市副溶血性弧菌引起的食源性疾病
发病率预测值与实际值比较Table 4 Relative error between forecast value and actual
value of incidence rates of foodborne disease caused by *Vibrio*
parahaemolyticus in Shanghai in 2011–2012

年份(Year)	2011	2012
预测发病率(Forecast incidence, /10 ⁵)	588.88	1063.16
实际发病率(Actual incidence, /10 ⁵)	1385.11	1106.29
相对误差(Relative error, %)	57.48	3.90

3 讨论

副溶血性弧菌引起的食源性疾病在沿海地区多发, 各类水产品中副溶血性弧菌检出率较高, 部分菌株具有耐药性^[6]。在上海, 随着人们饮食结构的改变, 生食、半生食(水产品)的饮食方式逐渐增多, 副溶血性弧菌引起的食源性疾病发生的可能性同样增加。上海市副溶血性弧菌引起的食物中毒事件是上海市食物中毒事件主要的危害因素。了解上海市副溶血性弧菌引起的食源性疾病发病情况, 并开展预警预测具有重要的公共卫生意义。

灰色系统理论是由邓聚龙于20世纪80年代初期提出的, 国内外长期应用于自然资源勘探、影像压缩、疾病预测等领域。GM(1,1)预测模型灰色系统理论中的基础预测模型为单变量一阶线性模型, 该模型具有所需样本量小、无需典型概率分布、计算简便和预测效果好等优点。它允许少量数据的预测, 对影响因素不明确的事件可以开展预测。本研究中副溶血性弧菌导致食源性疾病发生的影响因素众多, 如气温、食物中细菌的污染情况、地区等, 均会对病例的感染造成影响, 应用常规的统计学方法, 难以考虑到所有的影响因素。GM(1,1)预测模型适用于对副溶血性弧菌引起的食源性疾病发病情况的预测, 避免了逐一考虑众多影响因素单一作用的繁琐步骤; 该模型应用预

测的缺点在于, 模型受数据本身影响较大, 若数据本身波动性大, 会影响预测效果。

本研究结合上海市2007—2010年的数据, 制定上海市副溶血性弧菌所致食源性疾病病例感染数量的GM(1,1)预测模型, 并应用2011、2012年实际监测结果进行验证。本研究中模型通过关联度检验、后验差检验、发展系数等方法, 对模型精度、适用范围等进行了检验, 确保了模型的适用性及预测精度。结果显示该预测模型精度为1级(好), 开展短期预测(2~3年)需谨慎。根据监测数据及该模型预测结果验证, 2011年发病率预测误差较大, 2012年发病率预测结果为1063.16/10⁵, 与实际发病率1106.29/10⁵接近, 上海市副溶血性弧菌引起的食源性疾病感染病例逐年增长, 提示有关部门需要引起重视, 开展控制措施, 减少副溶血性弧菌引起的食源性疾病的发生。

本研究以上海市食源性疾病监测为基础, 由于该项监测工作开展时间有限, 现有数据较少, 应用GM(1,1)预测模型对数据建模受到前期基础数据的影响较大, 进行外推预测时尚需谨慎, 可能存在波动性大的缺点, 随着监测工作的深入开展, 逐步增加数据后对模型进行调整, 可有效提高模型精度及预测准确性。

·作者声明本文无实际或潜在的利益冲突。

参考文献

- [1] Food and Agriculture Organization of the United Nations. Risk assessment of *Vibrio parahaemolyticus* in seafood: interpretative summary and technical report [M]. Washington DC: Food and Agriculture Organization of the United Nations, 2011.
- [2] 毛雪丹, 胡俊峰, 刘秀梅, 等. 用文献综述法估计我国食源性副溶血性弧菌病发病率[J]. 中华疾病控制杂志, 2013, 17(3): 265-267.
- [3] 田明胜, 郑雷军, 彭少杰, 等. 2000—2007年上海市副溶血性弧菌致集体性食物中毒分析及对策[J]. 中国食品卫生杂志, 2008, 20(6): 514-517.
- [4] 李秀央, 李振洪, 蔡雪霞. 用EXCEL实现灰色数列模型GM(1,1)的预测[J]. 数理医药学杂志, 2000, 13(4): 296-297.
- [5] 刘思峰, 党耀国, 方志耕, 等. 灰色系统理论及其应用[M]. 北京: 科学出版社, 2010: 146-149.
- [6] 李薇薇, 梅玲玲, 唐震, 等. 2007—2009年中国副溶血性弧菌临床分离株分子特征分析[J]. 中华预防医学杂志, 2014, 48(1): 44-52.

(收稿日期: 2014-09-10)

(英文编辑: 汪源; 编辑: 王晓宇; 校对: 张晶)