

人群华支睾吸虫感染灰色预测模型的构建

臧炜¹ 卢光辉² 诸廷俊¹ 陈颖丹^{1*}

【摘要】目的 构建灰色 GM (1, 1) 模型用于人群华支睾吸虫感染情况预测。**方法** 利用“华支睾吸虫病病例/症状监测系统”收集的黑龙江省桦川县 2013 年 7 月—2014 年 6 月人群华支睾吸虫感染监测数据构建时间序列, 根据数列矩阵运算原理, 采用 EXCEL 软件中的 MINVERSE 等函数对该资料进行拟合, 确定人群华支睾吸虫感染 GM (1, 1) 预测模型并判断精度。**结果** 2013 年 7 月—2014 年 6 月间共收集了 271 例华支睾吸虫感染者信息, GM (1, 1) 模型的时间响应函数为 $X_{t+1}=16.0245e^{-0.0787t}$, 后验差比值 $C=0.5884$, 精度达到了基本合格水平。**结论** 初步构建了华支睾吸虫感染预测的灰色 GM (1, 1) 模型, 为今后预测人群华支睾吸虫感染情况提供了基础。

【关键词】 华支睾吸虫病; 感染; 灰色模型

Construction of grey model for prediction of *Clonorchis sinensis* infection Zang Wei¹, Lu Guanghui², Zhu Tingjun¹, Chen Yingdan^{1*}. National Institute of Parasitic Diseases, Chinese Center for Disease Control and Prevention, Key Laboratory of Parasite and Vector Biology, Ministry of Health, WHO Collaborating Center for Malaria, Schistosomiasis and Filariasis, Shanghai 200025, China

*Corresponding author: Chen Yingdan, Email: cyingdan@126.com

【Abstract】Objective To construct the grey model (1,1) for the prediction of *Clonorchis sinensis* infection. **Methods** The time related sequence was set up based on the data of *Clonorchis sinensis* infection in Huachuan County Heilongjiang Province from July 2013 to June 2014 collected from the monitoring system for case/symptom of clonorchiasis. According to the principle of matrix operation, the data were analyzed with series of function like MINVERSE in EXCEL software to establish the grey prediction model (1,1) which was evaluated for the accuracy subsequently. **Results** Total information of 271 people with *Clonorchis sinensis* infection was collected from July 2013 to June 2014, and the time response function of the model was $X_{t+1}=16.0245e^{-0.0787t}$ while C value was 0.5884, so the accuracy was also qualified. **Conclusion** The grey model (1,1) for the prediction of *Clonorchis sinensis* infection was preliminarily constructed, which provides the basis for prediction of *Clonorchis sinensis* infection in future.

【Key words】 Clonorchiasis; Infection; Grey model

华支睾吸虫病目前是我国危害严重的一种食源性寄生虫病, 主要由生食或半生食含华支睾吸虫囊蚴的鱼虾引起。2004 年第二次全国人体寄生虫病现状调查结果显示, 我国有 27 个省 (自治区、直辖市) 有华支睾吸虫病流行, 其中以广东、广西和黑龙江等省 (自治区) 人群感染较普遍, 据估算我国大约有 1 249 万人口感染华支睾吸虫, 疾病负担研究显示, 感染者伤残权重达 0.75%^[1-2]。近年来, 在开展华支睾吸虫病综合防治试点工作的同时, 一

些华支睾吸虫病重点流行地区也建立了华支睾吸虫病人群和中间宿主监测报告体系。

本研究以从华支睾吸虫病病例/症状监测系统收集的黑龙江桦川县 2013 年 7 月—2014 年 6 月人群华支睾吸虫感染数据为基础, 以周为单位构建时间序列, 根据数列矩阵运算原理对该资料进行拟合分析, 建立华支睾吸虫感染 GM (1, 1) 预测模型并判断其精度, 为今后预测人群华支睾吸虫感染情况奠定了基础。

1 资料来源及方法

1.1 资料来源

从华支睾吸虫病病例/症状监测系统收集黑龙江省桦川县 2013 年 7 月—2014 年 6 月间该县疾病预防控制中心华支睾吸虫病门诊粪检确诊的病

DOI: 10.3760/cma.j.issn.1673-4122.2015.03.011

作者单位: ¹200025 上海, 中国疾病预防控制中心寄生虫病预防控制所, 卫生部寄生虫病原与媒介生物学重点实验室, 世界卫生组织疟疾、血吸虫病和丝虫病合作中心; ²154300 佳木斯, 桦川县疾病预防控制中心

*通信作者: 陈颖丹, Email: cyingdan@126.com

例数据。

1.2 方法

1.2.1 灰色预测模型 GM (1, 1) 的建立

设由 (n+1) 个原始数据组成的时间数列 (以周为单位) 为 $X_t (t=0, 1, 2, 3, \dots, n)$, 按 $Y_t = \sum_{i=0}^t X_i (t=0, 1, 2, \dots, n)$ 累加生成数列 Y_t , 以累加生成数列 Y_t 为基础, 按 $Z_t = (Y_t + Y_{t-1}) / 2 (t=1, 2, \dots, n)$ 生成移动平均数列。GM (1, 1) 模型 Y_t 估计值 Y_t' 的一阶线性微分方程为:

$$Y_t' = (X_0 - \mu e^{-\alpha t} / \alpha) + \mu / \alpha (t=0, 1, 2, \dots, n)$$

式中 X_0 为初始时刻 ($t=0$) 的原始数据, μ 、 α 为待定的系数, 根据最小二乘法估计参数向量, 按矩阵运算法则得到其表达式为:

$$\mu = [(\sum_{t=1}^n Z_t^2)(\sum_{t=1}^n X_t) - (\sum_{t=1}^n Z_t)(\sum_{t=1}^n X_t Z_t)] / D,$$

$$\alpha = [(\sum_{t=1}^n X_t)(\sum_{t=1}^n Z_t) - n(\sum_{t=1}^n X_t Z_t)] / D$$

$$\text{其中 } D = n(\sum_{t=1}^n Z_t^2) - (\sum_{t=1}^n Z_t)^2$$

由于灰色模型实际上是由生成数列为基础构建的模型, 因此需要对累加生成数列累减还原, 即 GM (1, 1) 模型计算的结果是预测值的累加和, 故由 $X_t' = Y_t' - Y_{t-1}' (t=1, 2, \dots, n)$ 来计算预测值。

1.2.2 模型精度检验

通过计算后验差比值 C 和小误差概率 P 完成。

S_1 和 S_2 分别为原始数列 X_t 和绝对残差数列 δ_t 的标准差,

$$\delta_t = X_t - X_t' (t=0, 1, 2, \dots, n)$$

$$C = S_2 / S_1, P = P \{|\delta_t - \delta| < 0.6475 S_1\}$$

根据 C 、 P 值的计算结果, 按表 1 的标准进行模型预测精度的等级判定 (如果 C 、 P 不在一个判别等级上, 以所处的较高级别为准)。

1.2.3 计算各数列及参数值

按上述步骤, 应用 EXCEL 中的 MINVERSE,

表 1 灰色模型预测精度等级判定

Table 1 Evaluation of prediction precision grade of grey model

预测精度 Prediction precision		P	C
1 级 Grade 1	优秀 Excellent	>0.95	<0.35
2 级 Grade 2	合格 Qualified	>0.80	<0.50
3 级 Grade 3	基本合格 Basically qualified	>0.70	<0.65
4 级 Grade 4	不合格 Unqualified	≤0.70	≥0.65

MMULT, ROUND, POWER 和 STDEVP 等函数编写模型计算公式, 得出各数值。

2 结果

2.1 感染人数时间分布

从 2013 年 7 月起, 以周为单位记录系统收集的华支睾吸虫感染人数, 至 2014 年 6 月底共收集了 271 例华支睾吸虫感染者信息。其中 2013 年第 33 周报告的病例数最多, 达 17 例, 第 40 周开始报告病例数保持在较低水平 (图 1)。

2.2 模型函数建立

按上述方法建立桦川县监测点周人群华支睾吸虫感染预测 GM (1, 1) 模型, 其时间响应函数为 $Y_t' = -204e^{-0.0787t} + 244$, 求导还原得 $X_{t+1}' = 16.0245e^{-0.0787t}$ 。

2.3 模型精度检验

根据公式计算 C 值和 P 值分别为 0.5884 和 0.7537, 均属于表 1 标准中的 3 级, 综合 C 、 P 结果判定: 本模型预测精度基本合格, 可尝试用于外推预测。

3 讨论

华支睾吸虫病的传播过程复杂, 是生物环境因素和社会因素共同作用的结果, 因此要开展精确的预测比较困难。由于系统记录到的病例数是一组按

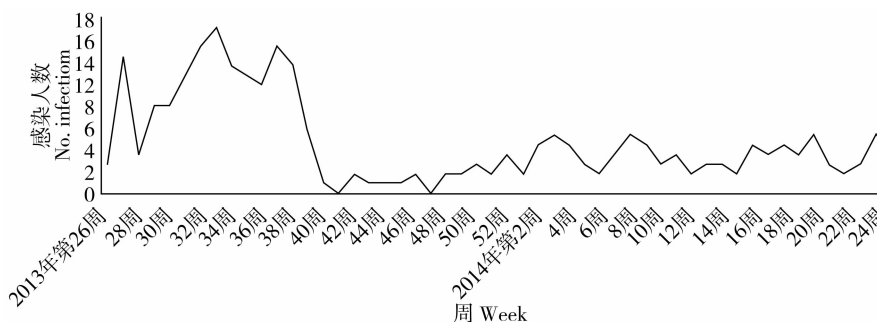


图 1 华支睾吸虫感染人数的时间分布

Fig. 1 Trend of the number of Clonorchis sinensis infection

时间顺序排列的离散的、随机的原始数据,理论上可通过构建单变量一阶线性模型如灰色模型 (grey model, GM) 或 ARIMA 模型或对发病情况进行预测。灰色模型利用的是时间数据序列,对原始数据的概率分布和样本含量没有严格要求,而且只需要 4 个以上的等时空距的观测数据^[3]。在建模过程中,首先通过累加把一组离散的、随机的原始数据转换为较有规律的生成数列,以提高原始序列的平滑程度,随后对一次累加生成的数列进行矩阵运算得到参数值,可用于小样本系统下疾病某一特征值的拟合和预测,以揭示该特征值的变化规律和未来发展态势。国内已有应用灰色模型对疟疾和血吸虫病流行进行预测的研究,并得到了较好的验证^[4-5]。

在开展华支睾吸虫病监测的基础上,本研究初步探索了华支睾吸虫感染情况灰色预测模型的构建。但由于华支睾吸虫病监测工作刚启动,该病例监测系统收集的信息全部来自于县级疾控中心门诊,并未覆盖到下辖的乡镇卫生院(所)乃至村卫生室,因此总体样本量偏少,而且在当地华支睾吸虫病综合防治示范区工作的带动下,2013 年 8—9 月主动前往县级疾控中心门诊咨询受检的人数较平时有一过性增多,这些都是影响模型精度的潜在因素。因此,随着华支睾吸虫病监测工作覆盖面扩大,数据来源将更加全面,所以需加强对日常收集信息的观察,特别关注病例数异常变动的情况并寻找其原因,以期在构建模型前剔除一些受外界干扰因素影响的数据,进一步提升预测模型的精度。

灰色预测模型的应用存在局限性,多应用于数据序列呈单调变化的情况,没有纳入社会、环境等因素,目前多应用于疾病短期发展趋势预测^[6]。因此,华支睾吸虫病作为一种和自然、社会因素密切相关的疾病,有必要根据一段时间华支睾吸虫病发病情况结合中间宿主感染、人群行为等因素建立时间序列参数分析模型,从而为华支睾吸虫病的科学防控提供可靠的信息依据和技术支撑,为今后其他食源性寄生虫病流行暴发的监测预警工作提供借鉴意义。

志谢 感谢黑龙江省疾病预防控制中心病媒寄生虫病防治所和桦川县疾病预防控制中心对本研究的大力支持!

参 考 文 献

- [1] 王陇德. 全国人体重要寄生虫病现状调查[M]. 北京: 人民卫生出版社, 2008: 10.
- [2] 钱门宝, 周晓农, 方悦怡, 等. 加强中国华支睾吸虫病研究[J]. 中国寄生虫学与寄生虫病杂志, 2011, 29(3): 211-214.
- [3] 朱恩学, 耿兴斌. 灰色系统在传染病预测中的应用[J]. 现代预防医学, 1994, 21(4): 229.
- [4] 李武剑, 易显明, 辜伟伟. 用灰色模型实现某市疟疾流行趋势的预测[J]. 长江大学学报, 2006, 3(2): 282-283.
- [5] 张姝, 张强, 尹治成. 灰色模型在血吸虫病感染率预测中的应用[J]. 现代预防医学, 2007, 34(7): 1292-1293.
- [6] Chaves LF, Pascual M. Comparing models for early warning systems of neglected tropical diseases[J]. PLoS Negl Trop Dis, 2007, 1(1): e33.

(收稿日期: 2015-02-20)

(本文编辑: 孙雅雯, 陈勤)

· 读者 · 作者 · 编者 ·

本刊对题名和摘要的要求

题名应以准确、简明的词语反映文章中最重要、最核心的特定内容,一般用短语,不用具有主、谓、宾结构的完整句。中文题名一般不超过 20 个汉字。如设副标题,可用冒号将副题名与主题名分开。题名中尽量避免使用非公知公认的缩略语、字符、代号等,应使用原形词。

论著类论文的摘要采用结构式,包括目的(Objective)、方法(Method)、结果(Results)和结论(Conclusions)4 部分。目的必须明确;方法描述要具体、详细、清楚,如果有随机分组必须交代随机的方法;结果中要给出关键性或主要的数据,百分率后要在括号中给出具体的数值比;结论要与目的呼应,必须是从该文的结果中推导出,不能把与结果无关的内容写到结论中。摘要采用第三人称撰写,不列图、表,不引用文献,不加评论和解释。英文题目和摘要一般与中文摘要内容相对应。

本刊编辑部