

我国病毒性肝炎感染和死亡人数预测分析

李海兰, 张涛*

广西科技大学, 广西 柳州

收稿日期: 2021年8月30日; 录用日期: 2021年10月25日; 发布日期: 2021年11月1日

摘要

病毒性肝炎是一种危害肝脏的极具传染性的疾病, 可有多种肝炎相关病毒诱发。在我国每年感染病毒性肝炎的人数都在1百万以上, 且病毒性肝炎的传染性较强, 对于我国这样一个世界人口大国而言, 对病毒性肝炎的预防工作尤为重要。于是本研究以2011~2019年我国病毒性肝炎感染人数和死亡人数为基础, 分别采用传统的GM(1,1)预测模型、BP神经网络预测模型和一元线性回归预测模型对2020~2029年病毒性肝炎感染人数和死亡人数进行预测, 并在考虑到模型本身产生的系统性误差的基础上, 利用各模型预测平均完全误差二次方根占比为权重将3种预测模型的预测结果进行加权平均作为最终的预测结果。最终得出结论病毒性肝炎感染人数在未来10年呈现显著的下降趋势, 但病毒性肝炎死亡人数却呈现上涨趋势。

关键词

病毒性肝炎, 灰色GM(1,1)模型, BP神经网络, 线性回归

Prediction of Viral Hepatitis Infection and Death in China

Hailan Li, Tao Zhang*

Guangxi University of Science and Technology, Liuzhou Guangxi

Received: Aug. 30th, 2021; accepted: Oct. 25th, 2021; published: Nov. 1st, 2021

Abstract

Viral hepatitis is a very infectious disease that endangers the liver and can be induced by a variety of hepatitis related viruses. The number that infects virus hepatitis in our country every year is in

*通讯作者。

1 million above, and the infectivity of virus hepatitis is stronger, to our country such a world's most populous country, the preventive work to virus hepatitis is particularly important. Therefore, based on the number of viral hepatitis infections and deaths in China from 2011 to 2019, this study used the traditional GM(1,1) prediction model, BP neural network prediction model and univariate linear regression prediction model respectively to predict the number of viral hepatitis infections and deaths from 2020 to 2029. On the basis of considering the systematic error of the model itself, the quadratic root ratio of the average complete error of each model is used as the weight and the weighted average of the prediction results of the three models is taken as the final prediction result. It is concluded that the number of viral hepatitis infection will decrease significantly in the next 10 years, but the number of viral hepatitis deaths will increase.

Keywords

Viral Hepatitis, Grey GM(1,1) Model, BP Neural Network, Linear Regression

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

病毒性肝炎主要由肝炎病毒诱发,其作用结果为肝细胞损伤,病毒性肝炎主要的特点是传播途径多、传染速度快、流行性强、患病负担高,正是基于这样的特点,结合我国人口基数大的现实条件,我们有必要对其进行分析研究。考虑到我国病毒性肝炎数据是从2011年开始收录的,病毒性肝炎自进行数据收录开始,一开始感染和死亡人数是逐年下降的,但在2015年后感染人数和死亡人数,随着时间的推移均呈现上涨趋势但上涨幅度差别较大且具有波动性。本研究针对我国2011~2019年病毒性肝炎感染人数和死亡人数为基础进行数据分析和预测,以便于对我国的病毒性肝炎防控提供科学的依据,也提醒广大的市民注意自身防护。

2. 研究现状

2.1. 发病率预测技术研究

当前我国所做的有关病毒性肝炎发病率的分析预测是十分有意义的,发病率的预测不仅能在很大程度上展示出该病毒的感染面,并且可以对未感染人群的行为产生一定的引导作用,因此很多专家及学者利用不同的预测模型对近年来我国病毒性肝炎发病率进行研究分析。吴小清等人利用季节指数分析不同类别的肝炎发病的季节性,并使用差分自回归移动平均模型(ARIMA)对2020年病毒性肝炎的发病率进行了预测[1];邓炯采用SPSS软件对绵阳市涪城区病毒性肝炎的发病特征进行了 χ^2 检验、趋势 χ^2 检验,得出病毒性肝炎的发病率呈现下降趋势[2];陈晓娥等人对各类型的病毒性肝炎发生率采用卡方检验的方法进行比较,最终得出结论甘肃省2010~2020年病毒性肝炎发病率呈现明显下降的趋势[3],由上述的研究我们可以得出国内针对病毒性肝炎的预测大部分都是利用时间序列和检验的方式进行的,并且均得出发病率下降的结论,但没有研究从感染人数和死亡人数入手进行研究。

2.2. 人数预测技术研究

人数总量预测研究比较常用的方法是利用单一的预测模型对某一人口规模进行预测,但是考虑到单

一的预测模型, 其预测值与真实值之间的偏差较大, 因此一部分的学者选用两种及两种以上的预测模型结合, 依次来降低预测产生的误差。李金伟等人采用 GM(1,1)预测模型对 2020~2030 年信阳市总人口数目和老龄化人口数目进行了预测, 预测了老龄化占比的变化趋势[4]; 徐丽丽等人考虑到单一的预测模型具有一定的局限性, 于是利用灰色预测模型与径向网络模型的组合对山东省人口总量进行了短期预测[5]; 李自鹏等人利用灰色预测模型、曲线增长模型与线性回归模型分别对徐州市的常住人口和户籍人口数目做出预测, 最后考虑到不同的预测模型对人口的预测存在不同程度的偏差, 于是采用将三组预测结果进行平均的方法得出三种预测结果的平均值, 作为最终的预测结果, 以达到降低预测误差的目的[6]; 钟恒恺采用灰色预测模型和回归预测模型, 通过获取以往南京市的旅游人数, 分别用以上两种预测模型未来中短期内南京市旅游人数和收入进行预测[7]; 鉴于先前学者对于病毒性肝炎的发病率分析和预测, 结合部分学者采用的两种及两种以上模型结合预测的方式, 本文主要从病毒性肝炎感染人数和死亡人数入手, 采用不同的预测模型对未来病毒性肝炎感染人数和死亡人数进行预测, 考虑到不同的预测模型预测的精度不同, 故为了将误差降低采用了将三种预测结果进行加权平均的方法, 以此达到对未来几年内我国病毒性肝炎感染和死亡人数预测的目的。

3. 数据来源与描述性分析

本研究所使用的 2011~2019 年的病毒性肝炎感染人数和死亡人数数据, 来源于国家统计局统计年鉴。

图 1 为病毒性肝炎感染人数条形图, 由图可以看出我国病毒性肝炎感染人数在 2011~2015 年呈逐年减少趋势, 但在 2016 年后突然猛增到 128 万, 并在接下来的几年内感染人数持续稳定在 128 万左右。

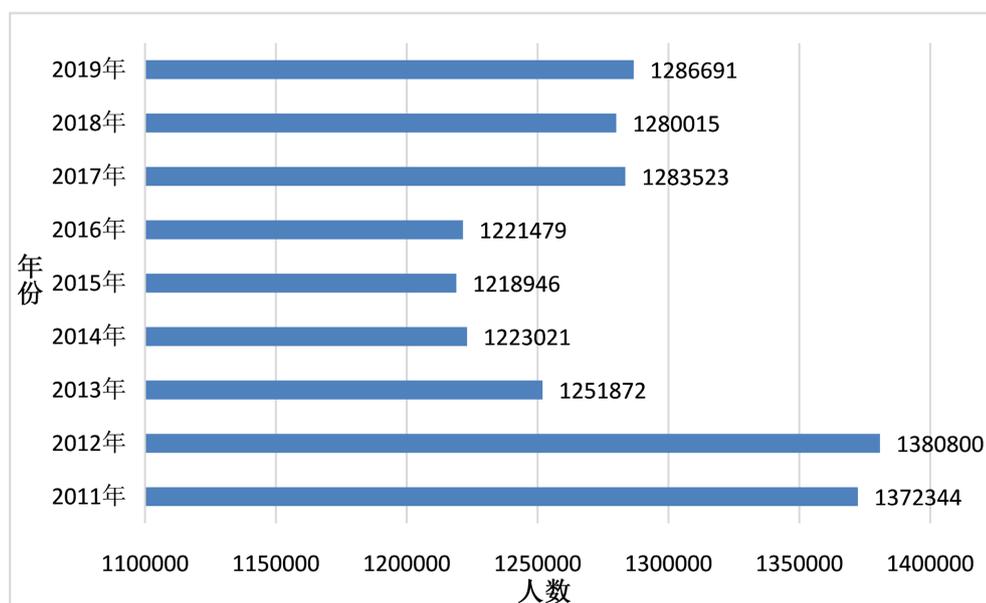


Figure 1. Bar chart of viral hepatitis infection

图 1. 病毒性肝炎感染人数柱状图

由图 2 可以看出 2011~2015 年病毒性肝炎死亡人数在不断的减少, 由 830 万减少到了 474 万, 在 2015 年病毒性肝炎死亡人数达到了最少, 但在 2015 后, 病毒性肝炎的死亡人数又出现了增长的趋势, 但是具有波动。

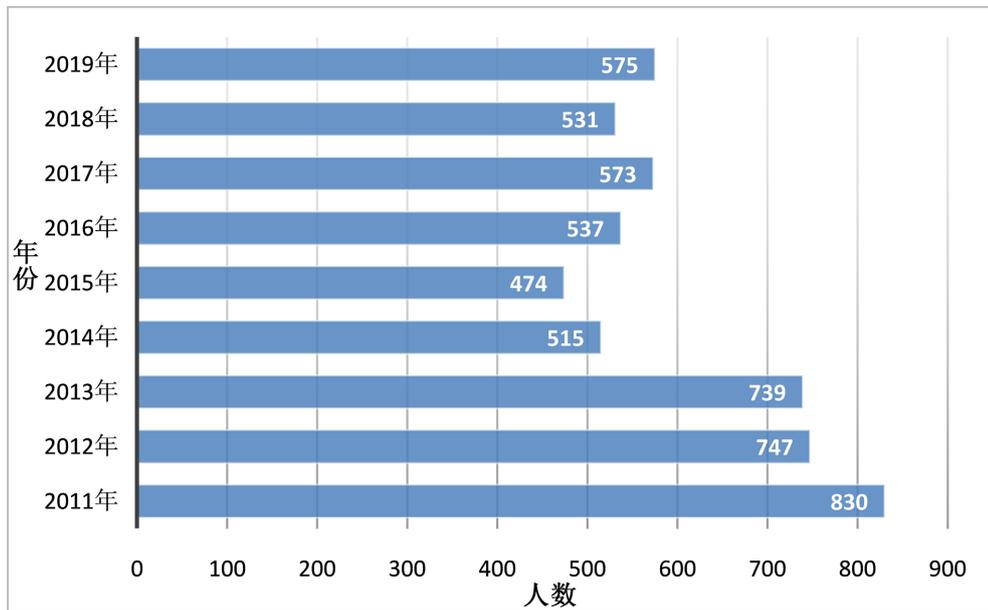


Figure 2. Bar chart of deaths from viral hepatitis
图 2. 病毒性肝炎死亡人数柱状图

4. 数据建模

4.1. 传统的 GM(1,1)灰色预测模型

传统的 GM(1,1)灰色预测模型是使用已有数据序列 $X^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)\}$, 经过累加生成 $X^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)\}$, 将原始数据的随机性降低, 累加后的数据往往会出现相对明显的规律。利用累加后的数据序列进行一阶微分方程构建, 通过求解利用累加序列构建的一阶微分方程, 再将求得的结果依次累减, 经过还原得到的便是灰色预测的结果, 从而利用其对未来的数据进行预测。具体的步骤如下:

步骤 1: 设 $X^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)\}$ 为 GM(1,1)模型的原始序列, 对其一次累加得公式:

$$X^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)\}$$

其中, $x^{(1)}(t) = \sum_{i=1}^t x^{(0)}(i)$, $t = 1, 2, \dots, n$ 。

令 $z^{(1)}$ 为 $x^{(1)}$ 的均值序列:

$$z^{(1)} = \{z^{(1)}(2), z^{(1)}(3), \dots, z^{(1)}(n)\}$$

其中, $z^{(1)}(t) = \frac{1}{2}[x^{(1)}(t) + x^{(1)}(t-1)]$, $t = 2, 3, \dots, n$ 。

步骤 2: 建立 GM(1,1)模型的灰微分方程:

$$x^{(0)}(t) + \alpha z^{(1)}(t) = u$$

相应的白化微分方程为:

$$\frac{dx^{(1)}(t)}{dt} + \alpha x^{(1)}(t) = u$$

步骤 3: 在对微分方程的参数进行估计时采用的方法为最小二乘法, 得到发展系数 α 与灰色作用量 u , 如下所示:

$$A = \begin{bmatrix} \alpha \\ u \end{bmatrix} = (B^T B)^{-1} B^T Y$$

其中:

$$B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix}, \quad Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}$$

步骤 4: 求解方程得:

$$\hat{x}^{(1)}(t+1) = \left[x^{(0)}(1) - \frac{u}{\alpha} \right] e^{-\alpha t} + \frac{u}{\alpha}$$

步骤 5: 由于 GM(1,1)模型得到的经过一次累加后的量, 于是需要将所得到的结果进行相反的计算使得其还原, 根据还原公式:

$$\hat{x}^{(0)}(t+1) = \hat{x}^{(1)}(t+1) - \hat{x}^{(1)}(t)$$

可对原始数列进行拟合和预测。

评估模型的指标有:

$$re = \frac{\hat{x}^{(0)}(t) - x^{(0)}(t)}{x^{(0)}(t)} \times 100\%$$

$$are = \frac{1}{n-1} \sum_{t=2}^n \frac{|\hat{x}^{(0)}(t) - x^{(0)}(t)|}{x^{(0)}(t)}$$

$$aae = \frac{1}{n-1} \sum_{t=2}^n |\hat{x}^{(0)}(t) - x^{(0)}(t)|$$

aae 表示 GM(1,1)模型的预测平均完全误差(表 1)。

在对 2020~2029 年病毒性肝炎感染人数进行预测时, 传统的 GM(1,1)模型预测平均误差 p 值为 0.034, 显示该模型适合于对病毒性肝炎的感染人数进行预测。

利用传统的 GM(1,1)模型对 2020~2029 年病毒性肝炎的死亡人数进行预测时, 传统的 GM(1,1)预测模型的预测平均误差 p 值为 $0.027 < 0.05$, 显示该模型可以用来对我国病毒性肝炎死亡人数进行预测。

在利用 GM(1,1)模型进行感染和死亡人数预测时, 利用 aae (预测平均完全误差)来衡量预测的准确度, 预测模型 p 值显示均小于 0.05, 表明此次预测可行。

4.2. BP 神经网络预测模型

BP 神经网络是以生物的神经网络为原型出发, 包括输入层、隐藏层和输出层。

本文中采用的激活函数, 具体形式如下:

$$f(x) = \frac{1}{1 + e^{-x}}$$

在本次研究中共设定输入层存在 3 个神经元, 隐含层神经元个数为 10 个, 输出层有 1 个神经元。

输入变量:

$$x_i = (x_1, x_2, x_3), i = 1, 2, 3$$

隐含层输出变量:

$$h_j = (h_1, h_2, \dots, h_{10}) = f\left(\sum_{i=1}^n w_{ij}x_i + a_j\right), j = 1, 2, \dots, 10$$

输出层变量为:

$$y_1 = \sum_{j=1}^{10} h_j w_{ij} + b$$

依次对数据进行逐年预测, 预测出 2014~2029 年病毒性肝炎感染人数和死亡人数的数据, 利用 2014~2019 年的预测值与真实值之间的离差平方和作为该预测模型拟合好坏的评价标准。在对我国病毒性肝炎感染和死亡人数进行预测时, 两次预测预测值与真实值之间的离差平方和均较小, 显示可以用该模型对我国病毒性肝炎的感染人数和死亡人数进行预测。

4.3. 线性回归模型

1) 病毒性肝炎感染人数预测

利用 2011~2019 年我国病毒性肝炎感染人数为原始数据, 应用 matlab 软件, 可以得到其关于时间 t_1 的一元线性回归预测模型

$$y_1 = 1328455 - 9720.12t_1$$

其中: y_1 表示相应年份的病毒性肝炎感染人数预测值; t_1 表示相应的年份序数, 将不同的年份序数代入预测模型中, 即可得到对应年份的病毒性肝炎感染人数预测值。

由 F 检验可知 $F = 10.6353 > F_{0.05}(1, 7) = 5.59$ 说明回归方程在 $\alpha = 0.05$ 水平上是高度显著的, 表明可以建立线性回归模型。

2) 病毒性肝炎死亡人数预测

利用 2011~2019 年我国病毒性肝炎死亡人数为原始数据, 应用 matlab 软件, 可以得到其关于时间 t_2 的一元线性回归预测模型

$$y_2 = 778.2778 - 32.9667t_2$$

其中: y_2 表示相应年份的病毒性肝炎死亡人数预测值; t_2 表示相应的年份序数, 将不同的年份序数代入预测模型中, 即可得到对应年份的病毒性肝炎死亡人数预测值。

由 F 检验可知 $F = 7.5848 > F_{0.05}(1, 7) = 5.59$ 说明回归方程在 $\alpha = 0.05$ 水平上是高度显著的, 表明可以建立线性回归模型。

5. 结论与建议

5.1. 结论

利用 3 种预测模型对我国病毒性肝炎感染人数进行预测, 具体结果见表 2。结果显示 3 种预测模型的预测误差都较小, 根据 3 种预测模型预测得到的 2020~2029 年我国病毒性肝炎感染人数预测值可以看出, 一元线性回归预测模型的预测结果相对较小, 而 BP 神经网络预测模型的预测结果相对较大, 相比之下 GM(1,1) 预测模型的预测结果较为居中, 为了消除所使用模型本身产生的系统性误差, 于是本研究利用预测平均完全误差为媒介, 计算出三种预测模型的预测平方完全误差开二次方, 利用预测平方完全

误差二次方根占比为权重计算预测值加权算数平均值, 以此加权平均值作为最终的预测值, 进一步降低预测的误差, 具体计算步骤与结果如下:

$$aae = \frac{1}{n-1} \sum_{t=2}^n |\hat{x}^{(0)}(t) - x^{(0)}(t)|$$

其中 aae: 预测平均完全误差; $\hat{x}^{(0)}(t)$ 为预测值; $x^{(0)}(t)$ 为原始数据; n 为个数。

Table 1. The average total error of the three prediction models

表 1. 三种预测模型的预测平均完全误差数值

	GM(1,1)		BP 神经网络		一元线性回归	
	感染人数	死亡人数	感染人数	死亡人数	感染人数	死亡人数
aae	43,532.95	66.38	17,208.13	49	49,184.00	40.88
\sqrt{aae}	208.65	8.15	131.18	7	221.77	6.39

最终得到 2020~2019 年我国病毒性肝炎感染人数分别为: 1,252,573、1,247,969、1,242,701、1,237,324、1,231,982、1,226,668、1,221,361、1,216,057、1,210,757、1,205,462 人。

Table 2. Prediction results of viral hepatitis infection in different prediction models

表 2. 不同预测模型病毒性肝炎感染人数预测结果

年份	GM(1,1)	BP 神经网络	一元线性回归	加权均值
2020	1,249,989	1,292,726	1,231,254	1,252,573
2021	1,245,966	1,295,847	1,221,534	1,247,969
2022	1,241,955	1,296,104	1,211,814	1,242,701
2023	1,237,957	1,295,878	1,202,093	1,237,324
2024	1,233,973	1,295,776	1,192,373	1,231,982
2025	1,230,001	1,295,776	1,182,653	1,226,668
2026	1,226,042	1,295,787	1,172,933	1,221,361
2027	1,222,095	1,295,790	1,163,213	1,216,057
2028	1,218,162	1,295,790	1,153,493	1,210,757
2029	1,214,241	1,295,790	1,143,773	1,205,462

从 3 种预测模型对 2020~2029 年病毒性肝炎死亡人数预测的结果来看, 3 种模型的预测误差均较小, 预测结果的精度也相对较高, 使用 3 种不同的预测模型对我国 2020~2029 年因病毒性肝炎而死亡人数进行预测, 预测的结果展示在表 3 中。由表 3 的预测结果可以看出传统的 GM(1,1)模型预测值偏小, 而一元线性回归模型预测值偏大, 相比较而言 BP 神经网络预测模型的预测值居中, 为了消除所使用模型本身产生的系统性误差, 于是采用上述的加权平均算法将 3 种预测结果的加权均值作为最终的预测结果, 最终得到 2020~2019 年我国病毒性肝炎死亡人数分别为: 537、552、581、614、653、701、755、816、885、960 人。

Table 3. Prediction results of viral hepatitis deaths in different prediction models**表 3.** 不同预测模型病毒性肝炎死亡人数预测结果

年份	GM(1,1)	BP 神经网络	一元非线性回归	加权均值
2020	477	499	655	537
2021	456	487	747	552
2022	436	494	860	581
2023	417	495	997	614
2024	399	492	1155	653
2025	382	492	1337	701
2026	365	493	1541	755
2027	349	493	1767	816
2028	334	493	2017	885
2029	319	493	2288	960

图 3 为病毒性肝炎感染人数未来 10 年预测值, 由图 3 我们可以看出我国病毒性肝炎感染人数在未来 10 年将会不断地减少, 并且呈现稳定的减少趋势, 但是减少的幅度较小, 预测结果可以看出到 2029 年我国仍然有 1,205,462 人感染病毒性肝炎, 这就意味着我国应加大力度对病毒性肝炎进行预防, 政府也应该出台相关的政策从源头上对其进行控制, 使其感染人数减少幅度加大。

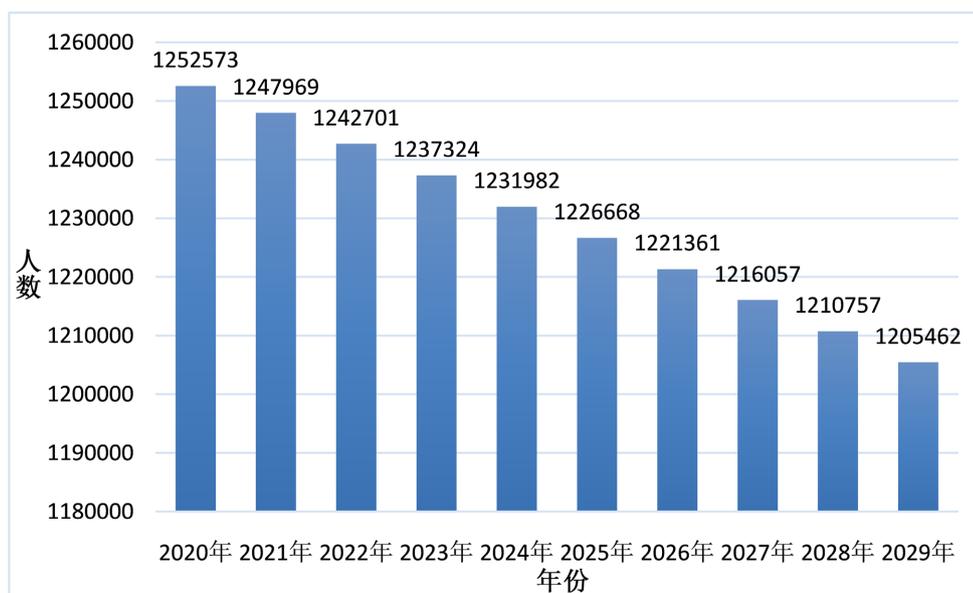
**Figure 3.** Estimated number of viral hepatitis infections over the next 10 years**图 3.** 病毒性肝炎感染人数未来 10 年预测值

图 4 为我国病毒性肝炎死亡人数未来 10 年预测值, 由图 4 我们可以看出我国病毒性肝炎死亡人数在未来 10 年呈现出不断增长的趋势, 并且呈现出增长速率增大的趋势, 从预测结果可以看出, 到 2029 年

我国病毒性肝炎死亡人数达到 960 人, 虽然从图 3 我们可以看出我国病毒性肝炎感染人数是不断减少的, 但病毒性肝炎死亡人数的预测值显示死亡人数在不断地增加, 因此我国应强化病毒性肝炎的治疗手段, 争取能将我国病毒性肝炎的死亡率降到最低, 为我国人民的健康做出更加坚实的保障。

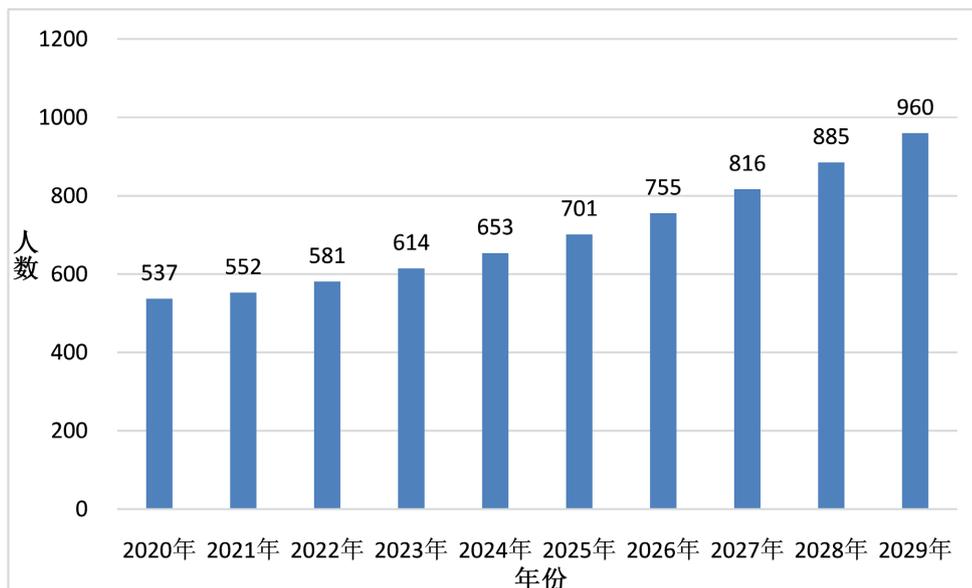


Figure 4. Projected number of deaths from viral hepatitis over the next 10 years

图 4. 病毒性肝炎死亡人数未来 10 年预测值

5.2. 建议

1) 病毒性肝炎的预防工作仍然要采取严格管控, 政府应加大病毒性肝炎知识普及力度, 加强教育宣传工作, 提供病毒性肝炎防治学习资料和平台, 提高病毒性肝炎疫苗接种率, 尤其应该注意学生、医护人员、流动人口等重点人群的疫苗接种工作。

2) 涉及到病毒性肝炎传染方式多的特点, 相关部门应注意严格对血液和输血设备进行管控, 提倡高质量的义务献血, 对于消毒政策应该严格执行。

3) 要防止病从口入, 应加大监督力度、注意自我防护, 去往人流量大或者有高风险传染的区域应佩戴口罩, 少接触传染源, 时刻做好消毒工作。加大力度宣传有关病毒性肝炎预防相关的知识, 不断提高人民群众自身的防护意识, 从源头上阻断病毒性肝炎的传播, 有效保护人民的生命安全和身体健康。

参考文献

- [1] 吴小清, 许阳婷, 苏晶晶, 徐庆, 王炜翔. 1989~2020 年南京市病毒性肝炎流行趋势分析[J]. 预防医学, 2021, 33(3): 236-240+245.
- [2] 邓炯. 2005~2018 年绵阳市涪城区病毒性肝炎流行特征统计分析[J]. 预防医学情报杂志, 2020, 36(8): 1049-1054.
- [3] 陈晓娥, 查成喜, 赵子莹. 2010~2020 年某三甲医院患者病毒性肝炎实验室结果特征分析[J]. 甘肃科技, 2021, 37(8): 80-82+62.
- [4] 李金伟, 王瑞瑞. 基于灰色模型的信阳市老龄化人口趋势预测[J]. 现代商贸工业, 2021, 42(8): 46-47.
- [5] 徐丽丽, 李洪, 李劲. 基于灰色预测和径向基网络的人口预测研究[J]. 计算机科学, 2019, 46(S1): 431-435.
- [6] 李自鹏, 欧向军, 周蓓蓓, 钱嘉琳, 欧亚根. 徐州市人口发展特征及其预测[J]. 江苏师范大学学报(自然科学版), 2021, 39(1): 21-25.
- [7] 钟恒恺. 基于数学模型的南京旅游人数预测与方案优化[J]. 通讯世界, 2018, 25(12): 262-264.