

传染病传播趋势及医疗费用动态预测研究

——基于灰色系统理论与Lee-Carter模型

孟祥波¹, 史博文², 向欣³, 曹瑞泽²

¹天津科技大学理学院, 天津

²天津科技大学机械工程学院, 天津

³天津科技大学经济与管理学院, 天津

收稿日期: 2025年8月4日; 录用日期: 2025年8月28日; 发布日期: 2025年9月5日

摘要

本文综合运用灰色系统理论与人口统计学模型, 构建多维度预测框架, 研究了病毒性肝炎、肺结核、梅毒三类重点传染病的流行趋势及其医疗费用增长规律。通过融合GM (1, 1)模型与Lee-Carter模型, 跳出单一方法下数据假设的局限, 构建一阶线性微分方程, 捕捉医疗费用的动态演化规律和传染病的长期趋势以及对时间因子进行外推预测。基于此, 本文尝试解析三种传染病的数据预测及预防指标, 初步分析了疾病与经济的相关性, 为卫生规划提供一种潜在的量化参考工具。

关键词

GM (1, 1)模型, Lee-Carter模型, 动态演化, 传染病

Research on Transmission Trends of Infectious Diseases and Dynamic Prediction of Medical Costs

—Based on Grey System Theory and Lee-Carter Model

Xiangbo Meng¹, Bowen Shi², Xin Xiang³, Ruize Cao²

¹College of Sciences, Tianjin University of Science & Technology, Tianjin

²College of Mechanical Engineering, Tianjin University of Science & Technology, Tianjin

³College of Economics and Management, Tianjin University of Science & Technology, Tianjin

Received: Aug. 4th, 2025; accepted: Aug. 28th, 2025; published: Sep. 5th, 2025

文章引用: 孟祥波, 史博文, 向欣, 曹瑞泽. 传染病传播趋势及医疗费用动态预测研究[J]. 建模与仿真, 2025, 14(9): 67-81. DOI: 10.12677/mos.2025.149585

Abstract

This paper comprehensively applies the grey system theory and demographic models, constructs a multi-dimensional prediction framework, and studies the epidemic trends of three key infectious diseases—viral hepatitis, tuberculosis, and syphilis—as well as the laws governing the growth of their medical expenses. By integrating the GM (1, 1) model and the Lee-Carter model, we transcend the constraints imposed by data assumptions under a single method, construct a first-order linear differential equation, capture the dynamic evolution patterns of medical expenses and the long-term trends of infectious diseases, and perform extrapolative forecasting for time factors. Building upon this, this paper attempts to analyze the data prediction and prevention indicators of three infectious diseases, preliminarily examines the correlation between diseases and the economy, and thereby aims to serve as a potential quantitative reference tool for health planning.

Keywords

GM (1, 1) Model, Lee-Carter Model, Dynamic Evolution, Infectious Diseases

Copyright © 2025 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

卫生经济负担与传染病防控是全球公共卫生领域长期关注的两大核心议题。近年来，中国卫生总费用从 2015 年的约 4×10^4 亿元增长至 2023 年的 9×10^4 亿元，这一趋势不仅反映出随着社会经济水平的提高，也揭示了政府在医疗卫生领域资源投入的持续增加。

与此同时，传染病发病率在中国乃至全球范围内均呈现复杂多变的趋势。一方面，病毒性肝炎、结核病等传统传染病由于其顽固性和传播性，仍然对公共卫生构成持续威胁[1]，这些传染病的防控不仅需要高效的医疗救治手段，更需要科学的资源配置和精准的防控策略。然而，现有研究多聚焦于卫生经济或传染病学单一领域，缺乏对二者之间动态关联的系统性分析[2]。

近年来梅毒在我国增长迅速，2014~2019 年中国梅毒总报告发病率逐年增长[3]，已成为报告病例数最多的性病。所报告的梅毒中，潜伏梅毒占多数，一、二期梅毒也较为常见，先天梅毒报告病例数也在增加。梅毒患者多为无症状、隐性感染，因其高危人群的隐蔽性，通过现有传染病报告系统收集的报告病例数并不能反映出全部状况[4]。

2. 研究对象与模型选取

传染病发病率和死亡率趋势的分析与预测是公共卫生领域的核心议题。病毒性肝炎、结核病、梅毒作为我国重点防控传染病，其发病规律的精准剖析对资源配置与防控策略制定至关重要。我们以这三种传染病的发病率和死亡率为研究对象，选取我国 2014~2023 年传染病发病率数据进行挖掘与分析。传染病的发病率和死亡率模型主要将时间因素考虑进来，符合实际情况。但是我国人口基数大，即数据处理均是一年一次，所以难做实时记录发病率和死亡率，因此本文先选择了 ARIMA 后选择了由 Lee 和 Carter 提出的离散时间随机模型，其在寿险精算中起着很重要的作用。Lee-Carter 动态死亡率评估模型被认为随机预测方法中最具代表性的同时也是死亡率预测应用最多的动态模型之一。

卫生总费用是衡量一个国家或地区卫生资源投入和卫生健康事业发展水平的主要指标，为此本研究核算并分析 2015~2023 年我国卫生总费用基本情况，通过数据反映出我国在解决卫生经济负担方面所取得的成效，并分析当下的问题以及预测未来将面临的挑战，完善我国关于卫生筹资方面的政策[5]。即选用适用于研究中的短期和中长期预测场景，尤其当面临数据样本具有时间的变化以及量有限、数据分布规律性不明显的情况时具有计算效率高、工作量小的特点，能够确保定量分析结果与定性分析结论保持高度一致，同时保持较高的预测准确度的灰色模型进而建立微分方程模型，得到在离散点处的解经过累减生成的原始数据的近似估计值，从而预测原始数据的后续发展[6]。

3. ARIMA 与 Lee-Carter 模型的建立与拟合

3.1. 数据来源

表 1、表 2 中病毒性肝炎、结核病、梅毒 2014~2023 年发病率和死亡率数据均取自数据来源：国家数据 <https://data.stats.gov.cn/>经标准化审核，确保数据准确完整。

Table 1. Overview of the incidence rates of infectious diseases in the past decade

表 1. 近十年传染病发病率的数据总览

发病率	病毒性肝炎	肺结核	梅毒
2014	90.25	65.63	30.93
2015	89.47	63.42	31.83
2016	89.11	61.00	31.97
2017	93.02	60.53	34.49
2018	92.15	59.27	35.63
2019	92.13	55.55	38.37
2020	81.12	47.76	33.08
2021	86.98	45.37	34.05
2022	78.40	39.76	31.27
2023	90.69	31.27	37.60

Table 2. Overview of the mortality rates of infectious diseases in the past decade

表 2. 近十年传染病死亡率的数据总览

死亡率	病毒性肝炎	肺结核	梅毒
2014	0.0380	0.1653	0.0051
2015	0.0348	0.1673	0.0043
2016	0.0388	0.1783	0.0038
2017	0.0415	0.2046	0.0033
2018	0.0381	0.2257	0.0028
2019	0.0412	0.2141	0.0030
2020	0.0419	0.1367	0.0038
2021	0.0369	0.1251	0.0021
2022	0.0385	0.1563	0.0016
2023	0.1700	0.1537	0.0011

3.2. ARIMA 和 Lee-Carter 模型构建

3.2.1. ARIMA (p, d, q)模型

ARIMA (p, d, q)模型结合了 AR 模型的历史性与 MA 模型创新性的优点, 可以处理更复杂的时间序列问题[7]。

$$Y_t = c + \varphi_1 Y_{t-1} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \quad (1)$$

上面式(1)中 $\varphi_1, \varphi_2, \dots, \varphi_p$ 和 $\theta_1, \theta_2, \dots, \theta_q$ 均为该模型的参数, c 是常数, ε_t 是当前时期的白噪声, $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}$ 是过去时期的白噪声。

假设有一时间序列 y_t , 定义滞后运算 B , 它将一个时刻的观测值转化为前一时刻的观测值, 扩展这个运算符的概念, 使之滞后 n 个时间步长:

$$B^n y_t = Y_{t-n} \quad (2)$$

另外, 通过滞后运算来实现对一阶差分和 n 阶差分的表征。因此, n 阶差分可定义为相邻标签值的差值:

$$\Delta^n y_t = y_t - y_{t-n} = y_t - B^n y_t = (1 - B^n) y_t \quad (3)$$

基于上述各种理论模型, 假设对一组数据进行二阶差分, 则可以直接对原始数据应用滞后运算符, 以此类推, d 阶差分可以被表示为:

$$\Delta^d y = (1 - B)^d y_t \quad (4)$$

自相关函数(Auto-Correlation Function, 即 ACF)和偏自相关函数(Partial Auto-Correlation Function, 即 PACF)是时间序列分析中常用的两个关键统计量, 其作用在于判断时间序列的平稳性, 并通过计算确定 ARIMA 模型的参数[8]。

定义一个时间序列 $\{X_k\}$, 那么对于任意的滞后(lag) k , 自相关函数 $\rho(k)$ 可以表示为:

$$\rho(k) = \text{Cov}(X_t, X_{t+k}) / \text{Var}(X_t) \quad (5)$$

在给定中间滞后项 $X_{t-1}, X_{t-2}, \dots, X_{t-k+1}$ 的条件下, X_t 与 X_{t-k} 的线性相关程度可定义为偏自相关函数(PACF), 一般记作 φ_{kk} 。偏自相关函数的计算过程是先排除其它更早期滞后观测值的影响, 然后再计算二者的相关性[9]。

3.2.2. Lee-Carter 模型

Lee-Carter 模型是动态死亡率评估模型, 该模型建立在 ARIMA 模型的基础上进行求解。若用 $m_x(t)$ 表示时间且年龄为 x 岁的人群的死亡率, 则 $m_x(t)$ 满足以下函数关系:

$$\ln m_x(t) = \alpha_x + \beta_x \times k(t) + \mu_z(t) \quad (6)$$

其中 α_x 表示与年龄因素相关的系数, 是死亡率取自然对数后再取平均值; $k(t)$ 通常称为死亡指数, 表示与时间因素相关的系数即疾病死亡率随时间变化的速率; β_x 表示年龄因子对其敏感程度; $\mu_z(t)$ 为随机扰动项并且满足 $\mu_z(t) \sim N(0, \sigma_\mu^2)$ 。

为确保模型系数求解时获得唯一确定的数值解, 对参数进行了 Lee 和 Carter 标准化处理:

$$\sum_{x=0}^Z \beta_x = 1, \sum_{t=1}^T k(t) = 0, \mu_x(t) \sim N(0, \sigma_\mu^2) \quad (7)$$

其中 T 为待估计死亡率数据中包含的日历年总数。

我们的目标是总误差尽可能小, 为了避免误差正负抵消, 我们用总误差平方和构造损失函数:

$$\min_{\beta_x} \frac{1}{2} \times \sum_x \left[d_{x,t} \times \mu_x(t) \right]^2 = \min_{\beta_x} \left\{ \sum_{x,t} d_{x,t} \times \left[\ln m_x(t) - \hat{a}_x - \beta_x \times \hat{k}(t) \right]^2 \right\} \quad (8)$$

右边关于 β_x 求导并令其为零，整理得 β_x 的估计量：

$$\hat{\beta}_x = \frac{\sum_{t=1}^T d_{x,t} \times \hat{k}(t) \times [\ln m_x(t) - \hat{a}_x]}{\sum_{t=1}^T d_{x,t} \times \hat{k}(t)} \quad (9)$$

死亡率取对数后的方差近似等于死亡人数 $d_{x,t}$ 的倒数，即

$$d_{x,t} = \frac{1}{\text{Var}[\ln m_x(t)]} \quad (10)$$

加权最小二乘法不需要对参数 $k(t)$ 的估计结果进行调整。

3.3. 数据处理与预测

3.3.1. 发病率的 ARIMA 模型

首先对三种疾病进行曲线拟合，如图 1 所示。然后用 KPSS 进行平稳性的检验，以病毒性肝炎为例用 ARIMA 模型进行模型拟合和结果的预测，并且绘制出置信区间，如图 2 所示。

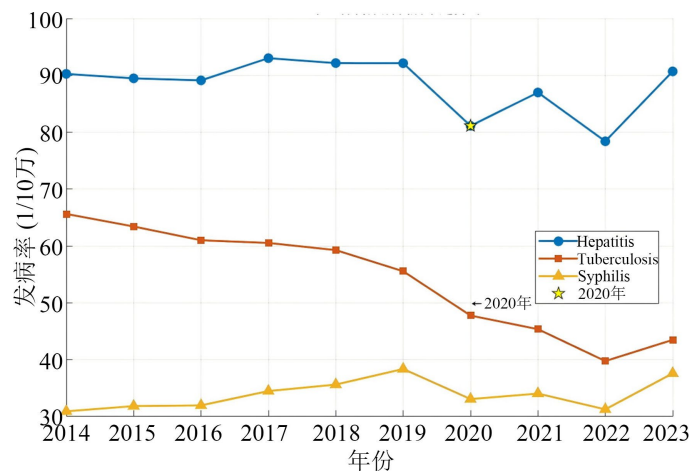


Figure 1. Comparison chart of incidence trends of three infectious diseases from 2014 to 2023

图 1. 2014-2023 年三种传染病发病率趋势对比图

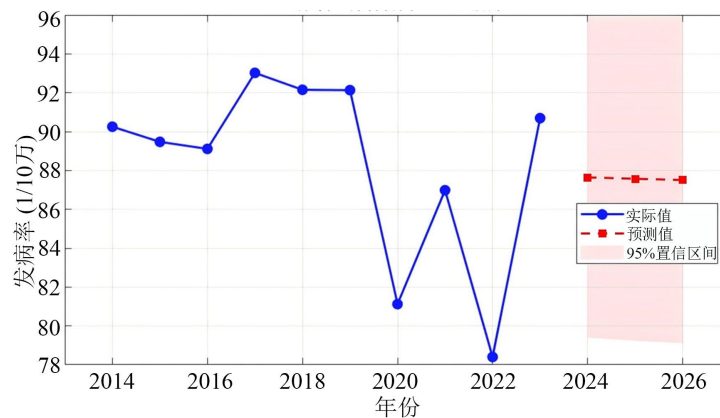


Figure 2. ARIMA prediction of viral hepatitis incidence rate

图 2. 病毒性肝炎发病率 ARIMA 预测

由图 2 可知置信区间为 95%，对应的年份在 2024~2026 年病毒性肝炎发病率落在粉色区间。同时也给出预测值，但是由于抽样误差等不确定性因素，真实的发病率不一定恰好等于预测值，因此置信区间波动比较范围比较大。

模型的预测结果以及参数的计算结果。

Table 3. Prediction results of the ARIMA model for incidence rate
表 3. 发病率 ARIMA 模型预测结果

指标	Value	Standard Error	T-Statistic	P-Value
Constant	0038649	0.046159	0.83731	0.40242
AR{1}	-0.91062	0.96012	-0.94844	0.3429
MA{1}	0.62828	1.1505	0.54607	0.58502
Variance	0.0058746	0.0031419	1.8698	0.061517

由表 3 中各参数的 P 值大于常见显著性水平 0.05，说明 AR{1}、MA{1} 项在模型中对病毒性肝炎发病率预测的影响不是特别显著。

由模型的计算得出：观测值数目：10，误差自由度：8；均方根误差：4.62， $R_2 = 0.203$ ，调整 $R_2 = 0.103$ ，F 统计量(常量模型)：2.04，p 值 = 0.191。结合 $R_2 = 0.203$ ，该值比较低，说明 ARIMA 模型对预测数据的拟合度程度不是很好，仅能解释 20.3% 的发病率数据。下面用 Lee-Carter 模型进行拟合预测。

3.3.2. 发病率的 Lee-Carter 模型

下面用 Lee-Carter 模型进行拟合预测。

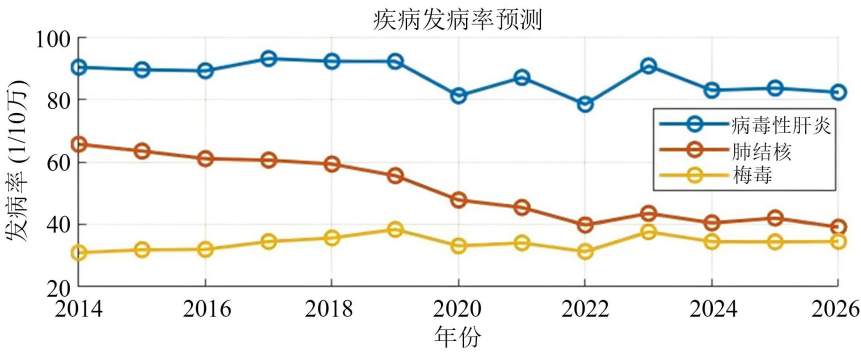


Figure 3. Prediction trend chart of the incidence rates of three diseases
图 3. 三种疾病发病率预测趋势图

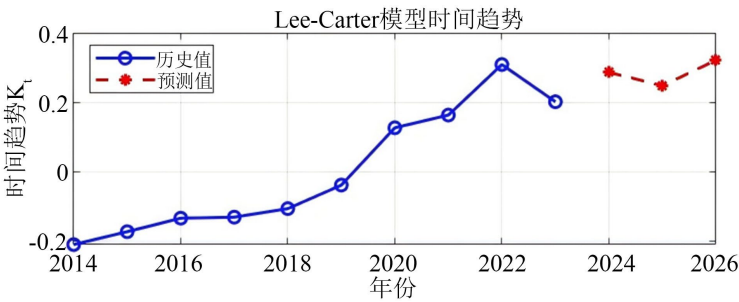


Figure 4. Prediction chart of total disease incidence based on the Lee-Carter model
图 4. Lee-Carter 模型疾病的总发病率预测图

根据图 4 的模型预测 K_t (时间趋势 K_t 与防控干预对死亡风险的调控逻辑——正值代表“时间 - 死亡”协同升高，负值体现反向抑制，近零暗示干预已锚定死亡风险。) 将会波动下降，但总的发病率仍受基准 a_i 起主导因素。在图 3 中，病毒性肝炎：尽管 K_t 下降可能减缓发病率的上升，但是其较高的基准发病率仍会导致总体居高不下；肺结核： K_t 下降可能加速发病率的降低，但是需要警惕下降趋缓后的潜在反弹风险；梅毒：因为敏感度系数非常小，可能保持目前的趋势，维持稳定。

Table 4. Evaluation results of Lee-Carter model parameters for incidence rates
表 4. 发病率率 Lee-Carter 模型参数评价结果

指标	基准死亡率 a_i	敏感性系数 b_i
病毒性肝炎	4.4797	-0.2180
肺结核	3.9784	-0.9739
梅毒	3.5215	0.0636

基准参数 a_i 反映疾病固有流行基线(由病原体传播能力、人群易感基础决定)，时间敏感系数 b_i 反映了各疾病发病率对时间趋势(K_t)的敏感程度。

根据表 4 可知：病毒性肝炎($b_i = -0.2180$)，负向关联，则因疫苗免疫屏障(如新生儿乙肝接种)与母婴阻断等长期干预，使“ K_t 上升(防控强化)”反而压缩传播空间，虽基线 a_i 仍较高，但已形成对短期趋势的“缓冲效应”。肺结核($b_i = -0.9739$)：敏感性极强，源于其空气传播的高流动性——国家《结核病防治规划》构建的分级诊疗与主动筛查体系，使“时间趋势 K_t 波动”直接作用于传播链切断效率 K_t 下降时筛查加强，发病率加速降低；反之潜伏感染复燃风险凸显，呼应图 4 中 K_t 反弹预警；梅毒($b_i = 0.0636$)：敏感性较弱，其性传播的隐蔽性(隐性感染占比高)及防控逻辑有关：哨点监测与性健康宣传的效果依赖长期行为养成，短期 K_t 波动难以快速改变流行轨迹。

3.3.3. 死亡率的 ARIMA 模型

首先对三种疾病进行曲线拟合，如图 5 所示。然后用 KPSS 进行平稳性的检验，以病毒性肝炎为例用 ARIMA 模型进行模型拟合和结果的预测，并且绘制出置信区间，如图 6 所示。

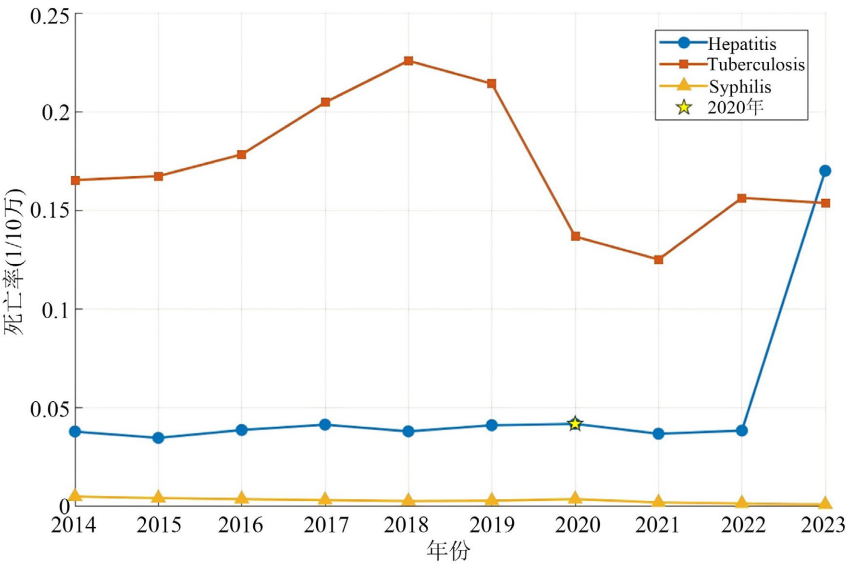


Figure 5. Comparative chart of mortality rate trends for three infectious diseases from 2014 to 2023
图 5. 2014~2023 年三种传染病死亡率趋势对比图

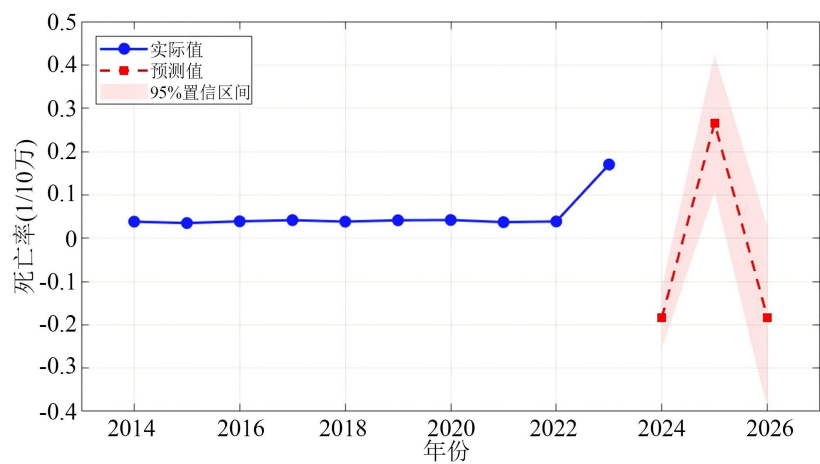


Figure 6. ARIMA prediction chart of viral hepatitis mortality rate
图 6. 病毒性肝炎死亡率 ARIMA 预测图

由图 6 可知置信区间为 95%，基于 ARIMA 模型在 2024~2026 年病毒性肝炎发病率落在粉色区间，预测值紧贴置信区间，显示波动范围不大。由于抽样误差等不确定性因素，真实的死亡率不一定恰好等于预测值。

Table 5. Prediction results of the ARIMA model form mortality rate
表 5. 死亡率 ARIMA 模型预测结果

指标	Value	Standard Error	T-Statistic	P-Value
Constant	0.11472	0.65586	0.17491	0.86115
AR{1}	1	2.7595	0.36238	0.71707
MA{1}	-0.39227	2.5229	-0.15548	0.87644
Variance	0.1193	0.12744	0.93619	0.34917

由表 5 中各参数的 P 值大于常见显著性水平 0.05，说明 AR{1}、MA{1}项在模型中对病毒性肝炎发病率预测的影响不是特别显著。

由模型的计算得出：观测值数目：10，误差自由度：8；均方根误差：0.0372， $R_2=0.285$ ，调整 $R_2=0.196$ ，F 统计量(常量模型)：3.19，p 值 = 0.191。结合 $R_2=0.285$ ，该值比较低，说明 ARIMA 模型对预测数据的拟合度程度不是很好，仅能解释 28.5%的发病率数据。以此用 Lee-Carter 模型进行拟合预测。

3.3.4. 死亡率的 Lee-Carter 模型

根据图 8 的模型预测，若 2023 年后的 K_t 会持续上升，结合图 7 数据预测病毒性肝炎死亡率将进一步上升，梅毒死亡率继续下降，肺结核死亡率保持稳定，但总的发病率仍受基准 a_i 起主导因素。

由表 6 可知：病毒性肝炎($a_i=-3.1018$, $b_i=0.6996$)的 a_i 的负值小(慢性化低基线致死，急性感染多自愈)，但 b_i 正值显著，呼应图中 K_t 上升时死亡率陡升——若慢乙肝“分级诊疗-抗病毒治疗-肝硬化监测”体系未有效落地，长期感染向肝衰竭/肝癌的进展将随时间加速，放大死亡风险；肺结核($a_i=-1.7726$, $b_i=-0.0743$)的 a_i 负值大但 b_i 趋近于 0，因国家结核病定点诊疗体系(免费抗结核药、全程督导)已稳定死亡风险，故图中死亡率保持平稳；梅毒($a_i=-5.8690$, $b_i=-0.7106$)的 a_i 负值最大(早期可经青霉素根治，基线致死极低)， b_i 负向显著，对应 K_t 上升时死亡率快速下降——哨点筛查扩围、性健康服务强化推动早期病例及时治愈，晚期心血管/神经梅毒等致死性并发症被遏制，与模型预测一致。

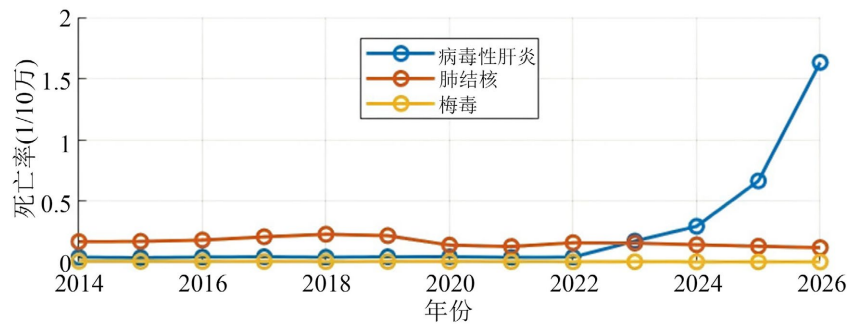


Figure 7. Prediction trend chart of mortality rates of three diseases

图 7. 三种疾病死亡率预测趋势图

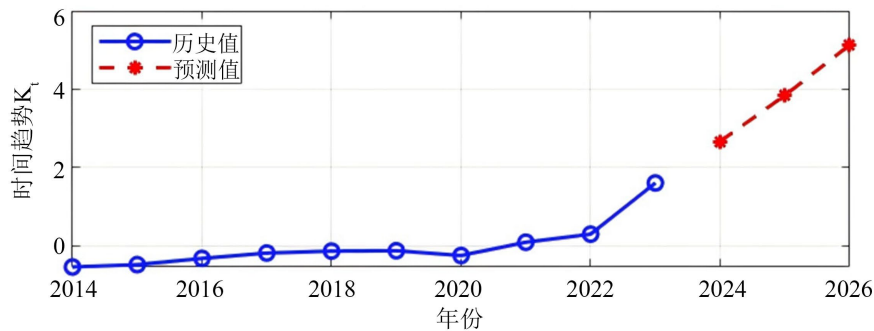


Figure 8. Prediction chart of total disease mortality rate based on the Lee-Cater model

图 8. Lee-Cater 模型疾病的总死亡率预测图

Table 6. Evaluation results of Lee-Carter model parameters for mortality rates

表 6. 死亡率 Lee-Carter 模型参数评价结果

指标	基准死亡率 a_i	敏感性系数 b_i
病毒性肝炎	-3.1018	0.6996
肺结核	-1.7726	-0.0743
梅毒	-5.8690	-0.7106

4. GM (1, 1)模型的建立与拟合

4.1. GM (1, 1)模型的构建

灰色系统理论是基于未经处理的非负离散数据列，经过一次累加处理，生成一个有序且降低随机性的新离散序列。然后通过构建相应的差分方程，来建立微分方程模型，求得到离散点处的解后，再通过累减生成对初始数据的逼近估算，最终基于该估计结果对原始数据的后续发展趋势进行预测[10]。

设 $x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))$ 是初始非负数据，对其进行一次累加生成新的数据 $x^{(1)}$ ：

$$x^{(1)}(m) = \sum_{i=1}^m x^{(0)}(i), m = 1, 2, \dots, n \quad (11)$$

令 $z^{(1)}$ 为数列 $x^{(1)}$ 的紧邻生成序列，即：

$$z^{(1)}(m) = \delta x^{(1)}(m) + (1 - \delta)x^{(1)}(m - 1) \quad (12)$$

$m = 2, 3, \dots, n$ 且 $\delta = 0.5$

下面引入矩阵：

$$u = (a, b)^T, Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}, B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix} \quad (13)$$

引入最小二乘法(OLS), 定义:

$$\hat{y}_i = kx_i + b \quad (14)$$

$$\hat{k}, \hat{b} = \arg \min_{k, b} \left(\sum_{i=1}^n (y_i - \hat{y}_i)^2 \right) = \arg \min_{k, b} \sum_{i=1}^n (y_i - kx_i - b)^2 \quad (15)$$

令:

$$L = \sum_{i=1}^n (y_i - kx_i - b)^2 = [y_1 - kx_1 - b, y_2 - kx_2 - b, \dots, y_n - kx_n - b] \begin{bmatrix} y_1 - kx_1 - b \\ y_2 - kx_2 - b \\ \vdots \\ y_n - kx_n - b \end{bmatrix} \quad (16)$$

则:

$$x^{(0)}(k) + \hat{a}z^{(1)}(k) = b \quad (17)$$

式(17)称为 GM (1, 1)的灰色微分方程, 其中 b 为灰作用量, $\hat{a}$ 为发展系数。

对于原始序列 $x^{(0)}$ 累加 r 次的序列为 $x^{(r)} = (x^{(r)}(1), x^{(r)}(2), \dots, x^{(r)}(n))$, 定义级比 $\sigma(k)$ 为累加序列中相邻两项的比值:

$$\sigma(k) = \frac{x^{(r)}(k)}{x^{(r)}(k-1)}, k = 2, 3, \dots, n \quad (18)$$

如果 $\forall k, \sigma(k) \in [a, b]$ 且区间长度 $\delta = b - a < 0.5$, 称累加 r 次的序列 $x^{(r)}$ 具有准指数规律
定义原始序列 $x^{(0)}$ 的光滑比即:

$$\rho(k) = \frac{x^{(0)}(k)}{x^{(0)}(1) + x^{(0)}(2) + \dots + x^{(0)}(k-1)} \quad (19)$$

假设 $x^{(0)}$ 为非负序列, 那么随着 k 的增加, $\rho(k)$ 最终会逐渐接近于 0, 因此要使 $x^{(1)}$ 具有准指数规律, 即 $\forall k, \delta < 0.5$, 只需确保 $\rho(k) \in (0, 0.5)$ 即可, 此时有序列 $x^{(1)}$ 的级比 $\sigma \in (1, 1.5)$ 。

4.2. 数据来源及一次拟合

卫生总费用、政府卫生支出、社会卫生支出、个人现金卫生支出及人均卫生费用数据均来自国家卫生健康统计年鉴, 时间跨度为 2015~2023 年, 数据经标准化校验确保准确性数据来源:

<https://data.stats.gov.cn/>。

Table 7. Graph of total health expenditure from 2015 to 2023
表 7. 2015~2023 年卫生总费用图

指标	2015 年	2016 年	2017 年	2018 年	2019 年
卫生总费用(亿元)	40974.64	46344.88	52598.28	59121.91	65841.39
指标	2020 年	2021 年	2022 年	2023 年	
卫生总费用(亿元)	72175	76844.99	85327.49	90575.81	

表 7 的卫生总费用在 2015~2023 年间线性增长, 呈持续上升趋势, 从 2015 年约 4×10^4 增长至 2023 年的 9×10^4 , 复合年均增长率达 9.8%。反映医疗需求扩大, 体现了社会对医疗卫生资源投入的持续增加。

Table 8. Graph of various health expenditures from 2015 to 2023

表 8. 2015-2023 年各卫生支出图

指标	2015 年	2016 年	2017 年	2018 年	2019 年
政府(亿元)	12475.28	13910.31	15205.87	16399.13	18016.95
社会(亿元)	16506.71	19096.68	22258.81	25810.78	29150.57
个人(亿元)	11992.65	13337.9	15133.6	16911.99	18673.87
人均(元)	2962.18	3328.61	3756.72	4206.74	4669.34
指标	2020 年	2021 年	2022 年	2023 年	
政府(亿元)	21941.9	20676.06	24040.89	24147.89	
社会(亿元)	30273.67	34963.26	38345.67	41676.8	
个人(亿元)	19959.43	21205.67	22940.94	24751.13	
人均(元)	5112.34	5439.97	6044.09	6425.32	

表 8 在医疗卫生总费用构成中, 政府、社会、个人支出呈现差异化增长特征: 政府支出经历“增长-回调-回升”波动, 社会支出持续加速(年均增 12.3%), 个人支出稳步递增(年均增 9.5%), 人均卫生费用从 2962.18 元升至 6425.32 元。各主体增长规律折射卫生筹资结构动态优化, 为后续投入效率与公平性研究提供时间序列支撑。

根据上述数据进行一次拟合曲线。

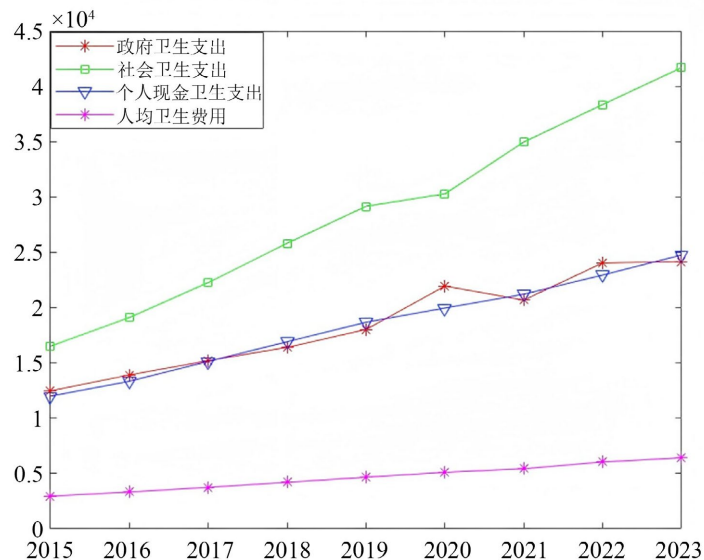


Figure 9. Comparison chart of fitted curves for each branch of expenditures

图 9. 各分支费用拟合曲线对比图

由图 9 分析得出: 政府支出在短期波动显著, 但是长期趋势向上; 社会支出拟合优度高, 能较好地捕捉其稳定的增长趋势; 个人支出费用拟合曲线斜率较低, 趋势较缓。总体趋势可以简单预测在接下来五年后社会支出可能突破 5 万亿元, 即政府支出是基层医疗和公共卫生的主要资金来源。

4.3. 对于医疗总费用利用改进的灰色模型进行数据处理及预测

利用灰色预测模型对原始数据进行数据处理。

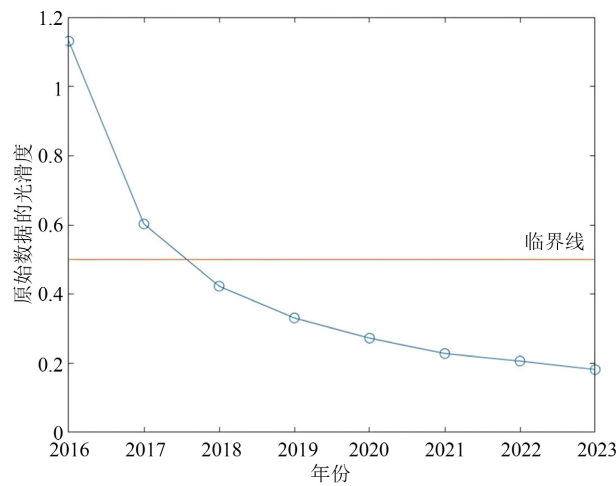


Figure 10. Smoothed ratio curve of total medical expenditure
图 10. 医疗总费用光滑比曲线

根据图 10 光滑比该数据具有准指数规律，下面进行 GM (1, 1)模型的数据处理。
选取前六组为训练数据，后三组为实验数据，如表 9 所示。

Table 9. Training data and experimental data table
表 9. 训练数据和实验数据表

训练数据		实验数据	
时间	医疗总费用(亿元)	时间	医疗总费用(亿元)
2015	40974.64	2021	76844.99
...	...	2022	85327.49
2020	72175.00	2023	90575.81

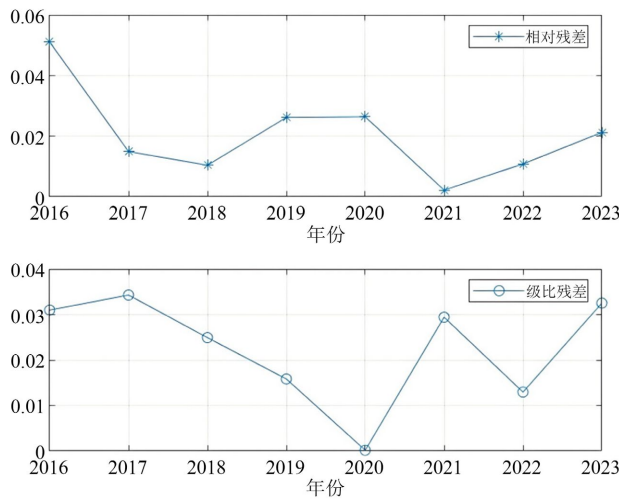


Figure 11. Visualization of residuals of fitted data
图 11. 拟合数据残差直观图

在进行统计模型的验证和分析时，来评估模型对数据的拟合程度常用的两种方法是残差检验和级比偏差检验。如图 11 所示：通过计算平均相对残差和平均级比残差，可以量化模型的拟合效果。

Table 10. Data of grey model fitted curve
表 10. 灰色模型拟合曲线数据

序号	年份	原始数据	拟合数据	相对残差	级比残差
1	2015	40974.64	40974.64	0	0
2	2016	46344.88	48709.851	0.0510298	0.0309829
3	2017	52598.28	53383.4909	0.0149285	0.0342866
4	2018	59121.91	58505.5598	0.0104251	0.0249180
5	2019	65841.39	64119.0838	0.0261584	0.0158359
6	2020	72175.00	70271.2174	0.0263773	0.0001606
7	2021	76844.99	77013.6393	0.0021947	0.0294121
8	2022	85327.49	84402.9868	0.0108348	0.0129379
9	2023	90575.81	92501.3316	0.0212587	0.0325112

进行残差检验和级比偏差检验，计算结果由表 10 可知：平均相对残差为 0.020401，表明数据的离散程度相对较小，模型能够较好地捕捉数据的总体趋势；平均级比残差为 0.022631，同样表明残差的分布比较集中，模型的拟合效果较好。结果都表明该模型对数据的拟合程度不错。这意味着模型能够合理地解释数据的变化，并且可以作为预测未来趋势的可靠工具。

利用数学 MATLAB 软件，使用 GM (1, 1)模型，利用最小二乘法得到发展系数 $\hat{a} = 0.11015$ ，灰作用量 $b = 42758.4087$ 求解模型预测结果如图 12 所示。

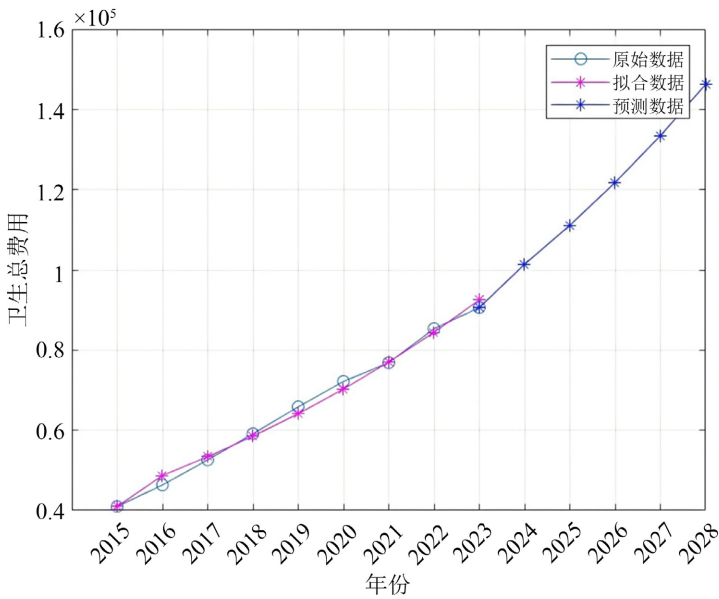


Figure 12. Fitted curve chart of grey model prediction data
图 12. 灰色模型预测数据拟合曲线图

根据以上数据列出拟合预测结果。

Table 11. Prediction results of the grey model
表 11. 灰色模型预测结果

2024 年	2025 年	2026 年	2027 年	2028 年
101376.7009	111103.649	121763.8838	133446.9527	146250.9952

根据灰色模型预测，2024 到 2028 年的我国卫生总费用持续上升，由表 11 可以看出预计从 10.1 万亿元增至 14.6 万亿元，这一增长主要受人口老龄化的加剧、同时还有第一、二类疾病的发病率等有关，同时还与各种医疗方面的新技术以及攻克各种疾病的难点有关。

5. 总结

本研究将灰色系统理论中的 GM (1, 1)模型与 Lee-Carter 模型有机地结合起来。GM (1, 1)模型能够把握数据的整体趋势，而 Lee-Carter 模型对传染病数据分析方面具有显著优势。探索突破单一模型数据假设约束的可能性，捕捉医疗费用动态规律与传染病长期趋势。另一方面，构建传染病多维预测框架，从多维度分析传染病的发病率、死亡率和医疗费用，初步分析了疾病与经济的相关性并且建立“模型参数 - 疾病机制 - 政策工具”的关联逻辑。

结合卫生经济与传染病数据，量化分析疾病负担，为医保政策的调整、医疗资源的分配等提供科学依据。更加全面地制定相关政策、并且深入地了解疾病造成经济负担的实际情况，从而制定出更加符合实际需求、更加科学合理的政策。结合传染病发病趋势的分析，可以更加合理地分配医疗资源，缓解医疗资源供需矛盾，提高资源配置效率。构建传染病预测模型，通过动态监测发病趋势，提升对突发公共卫生事件的应急响应能力。这有助于在传染病爆发初期及时发现并采取措施，防止病毒扩散蔓延，降低对社会经济的影响。同时，预警能力的提升也将为政府决策提供有力支持，确保在关键时刻能够迅速做出正确决策。但研究存在显著局限：数据维度仅覆盖三类传染病，未纳入耐药性、社会经济等混杂因素；参数估计依赖历史数据稳态假设，对突发公共卫生事件的适应性未验证。

未来可拓展方向包括：拓展病种至 HIV、流感等复杂传染病，整合病原体变异、政策干预强度等多源数据；通过交叉验证优化模型，探索极端场景下的预测能力。本研究为卫生规划提供一种可验证的量化参考工具——如病毒性肝炎的参数特征提示需强化慢病程管理投入，肺结核的稳定趋势支持现有诊疗政策延续，以期为准配置卫生资源提供初步数据支撑。

基金项目

中国轻工业联合会教育工作分会 2024 年度课题(QGJY2024159)；天津市教委科研计划项目(2018KJ113)。

参考文献

- [1] 李明. 科技发展引领未来[N]. 光明日报, 2024-10-01(1).
- [2] 周宇辉. 我国传染病流行现状与防控体系建设研究[J]. 中国卫生政策研究, 2023, 16(4): 74-78.
- [3] 郭子予, 汪文新, 刘冰. 新发重大传染病频发背景下疾病预防控制工作的卫生经济学再评价[J]. 中国卫生经济, 2020, 39(10): 47-50.
- [4] 岳晓丽, 龚向东, 李婧, 等. 2014-2019 年中国梅毒流行趋势与特征分析[J]. 中华皮肤科杂志, 2021, 54(8): 668-672.
- [5] 李莉, 杨建军, 王晓蕊, 等. 应用 Spectrum-STI 模型描述 2010-2020 年甘肃省 15-49 岁人群梅毒疫情变化[J]. 中国艾滋病性病, 2025, 31(3): 312-314.
- [6] 代玉巧, 严运楼, 刘政. 基于 GM(1, 1)模型的北京市卫生总费用及构成变化趋势预测分析[J]. 现代预防医学, 2021, 48(11): 1996-2000.

-
- [7] 高威, 文韬, 王铮, 等. 基于灰色 GM(1, 1), ARIMA 及 Holt-Winters 模型的延吉市肺结核发病率短期预测研究[J]. 中国卫生统计, 2025, 42(2): 263-266.
 - [8] 郑明贵, 李丽, 于明. 中国 2025-2035 年稀土需求预测——基于灰色关联分析和 ARIMA-GM-BP 神经网络的组合模型[J]. 稀土, 2025, 46(2): 145-158.
 - [9] 马峥, 李行, 张典. 基于 MSTP 协议和 5G 双域技术的校园融合组网方法研究[J]. 信息技术与信息化, 2024(4): 37-44.
 - [10] 龙宇. 基于组合模型的气象干旱预测研究[D]: [硕士学位论文]. 荆州: 长江大学, 2024.