

基于COVID-19的“大流行”病定量界定研究

刘元珂, 范馨月*

贵州大学数学与统计学院, 贵州 贵阳
Email: 785744277@qq.com, *fan.xinyue@163.com

收稿日期: 2021年7月4日; 录用日期: 2021年7月28日; 发布日期: 2021年8月5日

摘要

本文基于Logistic预测模型对新型冠状病毒肺炎(COVID-19)的传染率以及最大感染人数进行了预测, 计算出感染率与非大流行传染病的感染率, 并进行比较, 从而对大流行传染病进行定量界定。基于中国2020年1月22日~2020年4月15日疫情数据作为实验数据, 建立防疫政策函数、经济函数、医疗函数, 以及Logistic预测模型对COVID-19传染率及最大感染人数进行了预测, 并且基于预测结果计算出感染率与非大流行传染病的感染率并作了比较, 量化了大流行病的界定。运用Logistic预测模型对COVID-19的传染率进行预测, 其结果具有很好的拟合性, 对比分析可量化出“流行”与“大流行”病的界限。

关键词

“大流行”病, Logistic预测模型, COVID-19, 预测

Quantitative Definition of “Pandemic” Disease Based on COVID-19

Yuanke Liu, Xinyue Fan*

School of Mathematics and Statistics, Guizhou University, Guiyang Guizhou
Email: 785744277@qq.com, *fan.xinyue@163.com

Received: Jul. 4th, 2021; accepted: Jul. 28th, 2021; published: Aug. 5th, 2021

Abstract

This paper predicts the infection rate of new coronary pneumonia based on the Logistic prediction model, and compares the prediction result with the infection rate of non-pandemic infectious dis-

*通讯作者。

eases, so as to quantify the definition of pandemic infectious diseases. Based on China's epidemic data from January 22, 2020 to April 15, 2020, as experimental data, the epidemic prevention policy function, economic function, medical function, and Logistic prediction model are established to predict the infection rate of new coronary pneumonia, and the infection rate of pandemic infectious diseases is compared to quantify the definition of pandemic. The Logistic prediction model has a good fit for predicting the infection rate of new coronary pneumonia. Comparative analysis can quantify the boundary between "epidemic" and "pandemic".

Keywords

"Pandemic" Disease, Logistic Prediction Model, COVID-19, Prediction

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

随着新型冠状病毒肺炎(COVID-19)的突然爆发,世界卫生组织将这种症状类似于普通感冒的病症称之为大流行病,这是继2009年的H1N1流感后的首例大流行病。世界卫生组织曾表示,大流行是“新疾病的全球传播”也就是说“大流行”特征所指的并不是疾病的严重性(致死率),而是疾病传播的广泛程度(传染率)。对于达到大流行水平与否,当下是没有定量的严格标准,也没有触发该定义的病例或死亡数量阈值[1]。随着搜索引擎及社交媒体等网络技术的发展,早点发现传染病流行趋势然后快速反应,可以减少传染性疾病对人群的影响,由于空气传播是明显高于体液传播,传染途径可以作为传染病大流行的一个特征因素[2]。目前传染病的预测指标主要有暴发时间、暴发程度、波峰时间及波峰强度、持续时间、累及地区的预测[3],如果能给出大流行的一个定量界定标准,那么便可以辅助公共健康干预措施的选择和医疗资源的分配,能更有效地预防和控制大流行病的发生。

本文采用中国2020年1月22日~2020年4月15日共85天的COVID-19疫情数据作为实验数据,构建防疫政策函数、经济函数、医疗函数,以及Logistic预测模型对COVID-19的传染率进行预测,计算出感染率对比SARS、乙肝等12种传染病的感染率,给出“大流行”病的基于感染率的定量描述。

2. 模型构建

2.1. 特征分析

综合考虑影响COVID-19的各种因素,为了使所选取的指标具有全面性和代表性,采用Pearson相关系数筛选出确诊人数,治愈人数,死亡人数,国民生产总值,人口密度,疫情持续时间六个指标作为研究对象,由于COVID-19的来源及首例传染病人的日期未确定,暂时以2019年12月1日作为疫情开始时间。具体指标可以区分为正向指标和逆向指标:正向指标也称为效益型指标,指标原始数值越大,对控制COVID-19疫情的发展越有利;正向指标原始数值越小,对控制COVID-19疫情发展越不利。逆向指标也称为成本型指标,逆向指标原始数值越小,对控制COVID-19疫情越有利;逆向指标原始数值越大,对控制疫情越不利。所选取的指标中,治愈人数,国民生产总值,人口密度3个指标为正向指标;确诊人数,死亡人数,疫情持续时间3个指标为逆向指标。

令隶属函数[4]

$$u_h(x) = \begin{cases} u_h^+(x), & x \text{ 为正向指标} \\ u_h^-(x), & x \text{ 为逆向指标} \end{cases}$$

正向指标隶属度 $u_h^+(x)$ 为:

$$u_h^+(x) = \begin{cases} 1, & x = x_{\max} \\ \frac{x - x_{\min}}{x_{\max} - x_{\min}}, & x_{\min} < x < x_{\max} \\ 0, & x = x_{\min} \end{cases} \quad (1)$$

逆向指标隶属度 $u_h^-(x)$ 为:

$$u_h^-(x) = \begin{cases} 1, & x = x_{\min} \\ \frac{x_{\max} - x}{x_{\max} - x_{\min}}, & x_{\min} < x < x_{\max} \\ 0, & x = x_{\max} \end{cases} \quad (2)$$

其中 x_{\max}, x_{\min} 分别表示第 i 个指标的 m 个样本原始数据值中的最大、最小值, 式(1)式(2)中的 $u_h(x) \in [0, 1]$, 隶属函数值越接近于 1 表示疫情防控水平越好, 当 $u_h(x) = 1$ 时达到最优控制水平。

根据分析各项指标对 COVID-19 每日新增感染人数的影响显著水平, 各指标之间的关联程度及各指标的信息量决定指标的权重, 从而在一定程度上避免了主观因素带来的偏差。设 x_{ij} (经过无量纲化后, 数据不能为 0) 表示第 j 个样本的第 i 个指标的数值, ($j = 1, 2, \dots, m; i = 1, 2, \dots, p$), 其中 m 和 p 分别为样本个数和指标个数。采用熵值法确定各指标权重的大小, 进行权重变换, 计算得出各指标新的权重:

$$\omega_i = \frac{g_i}{\sum_{i=1}^p g_i} \quad (3)$$

其中 g_i 代表第 i 个指标的信息熵冗余度, 将治愈人数, 国民生产总值, 人口密度 3 个正向指标带入隶属函数式(1), 确诊人数, 死亡人数, 疫情持续时间 3 个逆向指标原始数据值带入隶属函数式(2)分别计算出其隶属度。再由熵值法计算出指标的权重向量集 ω , 结果如表 1 所示:

Table 1. The weight vector set of influencing factor index

表 1. 影响因素指标的权重向量集

治愈数(人)	GDP(十亿美元)	人口密度	确诊数(人)	死亡数(人)	持续时间(天)
0.18849755	0.111101362	0.09738699	0.29422281	0.23285441	0.07602461

由表 1 看出确诊人数和死亡人数权重较大, 是分析新型冠状病毒肺炎传染性的两个重要指标, 其中治愈数和死亡人数是反映此次疫情医疗条件的重要指标; GDP 是反映此次疫情的经济指标; 确诊人数和持续时间与政策有着密不可分的关系, 将其作为政策指标。

2.2. 函数的构建

根据中国与世界新型冠状病毒疫情确诊人数增长率对比图, 如图 1 所示: 可以清晰地看出, 2020 年 1 月 22 日至 2 月 20 日中国与世界的疫情增长趋势基本一致, 但是 2 月 20 日之后有了明显差异, 中国的疫情增长率一直逼近 X 轴, 显然是中国采取了有效的政策防疫措施, 成功抑制住了新型冠状病毒疫情的增长。

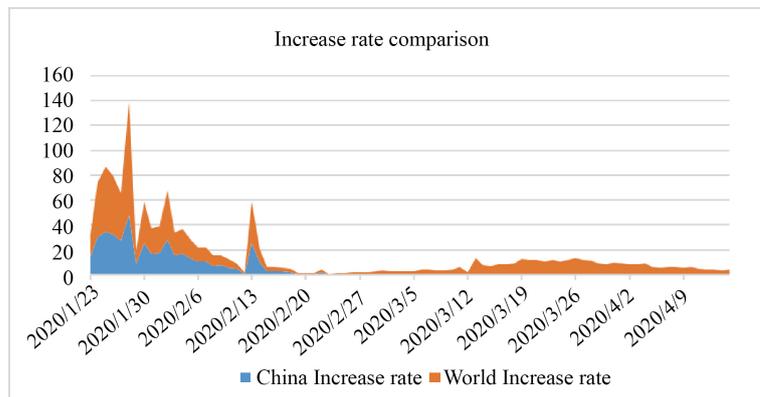


Figure 1. China’s epidemic growth rate from January to May 2020 compared with that of the world

图 1. 中国 2020 年 1 月至 5 月与世界疫情增长率对比

国内 COVID-19 疫情的发展大体上可以分为以下三个阶段:

第一阶段(2020 年 1 月 22 至 29 日): 爆发期, 1 月 29 日达到峰值, 共 8 天;

第二阶段(2020 年 1 月 30 日至 2 月 13 日): 持续期, 共 14 天;

第三阶段(2020 年 2 月 14 日至 4 月 15 日): 衰退期, 共 62 天。

COVID-19 疫情发展的各个阶段, 各因素取值和变化规律(函数)不同, 因此, 需要用分段函数来描述疫情各阶段实际情况。

2.2.1. 防疫政策函数

防疫政策函数 $p(t)$ 是一个由政府的防疫措施决定二分变量, 它的变化反映了政府措施的力度。防疫政策函数值由政府对于疫情关注程度决定。其中, 疫情刚爆发时, 因为政府还没有出台相应的政策, 所以对于疫情感染人数来说是没有影响记为 0, 然后随着疫情时间的持续, 政策一旦出台, 就会影响抑制传播人数的增长速率, 因此有影响就记为 1。政府对疫情的关注程度是随着疫情的严重性逐渐增加的, 并且在认识到 COVID-19 疫情危害之后会越来越受关注, 在关注度达到一定程度时, 其增长速度也会逐渐减缓, 最终在某一水平逐渐趋于稳定。因此, 对实验数据采用分段函数光滑化拟合得到如下防疫政策函数[5]:

$$p(t) = \begin{cases} 0, & t < 8, \\ \frac{1}{98}(t-8)^2, & 8 \leq t < 15, \\ -\frac{1}{98}(t-22)^2 + 1, & 15 \leq t < 22, \\ 1, & t \geq 22. \end{cases} \quad (4)$$

其中 t 为疫情爆发天数。为了直观地表现出防疫政策函数的变化趋势, 图 2 给出防疫政策函数趋势图:

函数的拐点出现在第八天、第十五天、第二十二天。疫情开始的前八天, 政府对于疫情认识不足, 政策滞后。在政府认识到疫情危害时, 迅速做出反应, 出台一系列防控政策。在达到一定防疫程度之后, 政府出台政策缓慢, 并在第二十二天关注度及采取的防控措施力度基本达到最大值。该防疫政策函数的建立主要是基于中国的疫情数据所得, 所以能很好的反应出我国疫情爆发的天数与政策力度之间的关系, 从而为后续进一步建立 Logistic 预测传染率模型做准备。

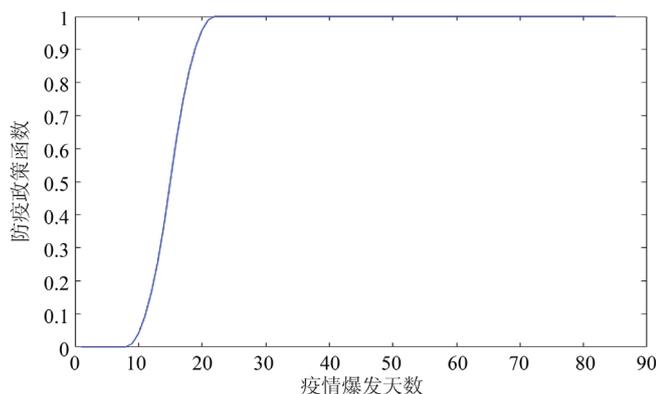


Figure 2. Epidemic prevention policy function
图 2. 防疫政策函数

2.2.2. 经济函数

经济函数 $g(u)$ 是由我国经济水平决定, 表示经济对 COVID-19 疫情的影响。国家经济决定了政府对于抗击疫情财力上的支持, 因此, 总体经济水平越高, 抗击疫情的力度就相对越大。因此, 由于国内总体经济水平对医疗政策等有一定的促进作用, 可定义:

$$g(u) = u \quad (5)$$

其中 u 为国民生产总值, 取 2019 年国民生产总值 1.4311 (单位: 十万亿元), 带入得 $g(u) = 1.4311$ 。

2.2.3. 医疗函数[6]

医疗函数 $h(t)$ 是由有效医疗决定的, 定义有效医疗为国家及社会对 COVID-19 防控及治疗投入的人力, 物力, 财力等。在疫情期间, 国家及社会对 COVID-19 的医疗投入随时间变化快速增大; 在疫情持续期时, 医疗投入随时间逐渐缓和, 达到某一顶峰; 在疫情衰退期, 医疗投入总和基本趋于平稳。因此, 考虑如下医疗函数:

$$h(t) = \beta\sqrt{t} \quad (6)$$

其中 β 为调节参数。图 3 给出医疗函数趋势, 其中四条曲线从上到下依次为 $\beta = 4, 3, 2, 1$ 时医疗函数的曲线。

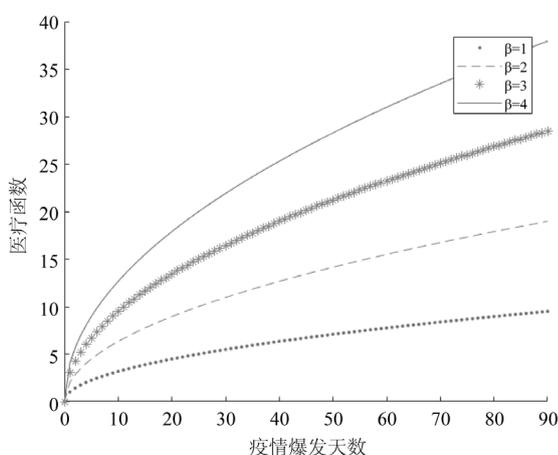


Figure 3. COVID-19 medical function
图 3. COVID-19 医疗函数

2.3. Logistic 预测模型

基于上述构建的防控政策函数、经济函数和医疗函数, 建立以下 Logistic 传染率预测模型:

$$y = \frac{kb}{1 + \alpha * q^{-(m_1g(u)+m_2p(t)+m_3h(t))t}} \tag{7}$$

其中 α, k 为参数; q 表示传染率; t 表示时间, 共 85 天; $m_i (i = 1, 2, 3)$ 表示函数对应的权重; b 表示人口密度, $p(t)$ 为防疫政策函数, $g(u)$ 为经济函数, $h(t)$ 为医疗函数。考虑到用最小二乘方法, 即通过求解下述模型获得未知参数的估计。

$$\min_{k,b,\alpha,q} \sum_{t=1}^{85} \left(y_t - \frac{kb}{1 + \alpha * q^{-(m_1g(u)+m_2p(t)+m_3h(t))t}} \right)^2 \tag{8}$$

注意到 k, b 都是常数, 因此上述模型可以等价求解

$$\min_{k,b,\alpha,q} \sum_{t=1}^{85} \left(y_t - \frac{K}{1 + \alpha * q^{-(m_1g(u)+m_2p(t)+m_3h(t))t}} \right)^2 \tag{9}$$

其中 $K = kb$ 。

采用 Armijo 步长的梯度下降法求解上述问题。然而, 上述问题是非凸的, 很难获得其全局最优解, 因此, 初始点的选择是至关重要的。通过我国传染人口数据图, 结合 Logistic 预测模型特征, 可以选择初始点为

$$K_0 = 100000, \alpha_0 = 1, q_0 = e.$$

此外, 选择 Armijo 步长初始步长 $a = 1$, 调节参数 $b = 0.8$ 。拟合结果如图 4 所示。

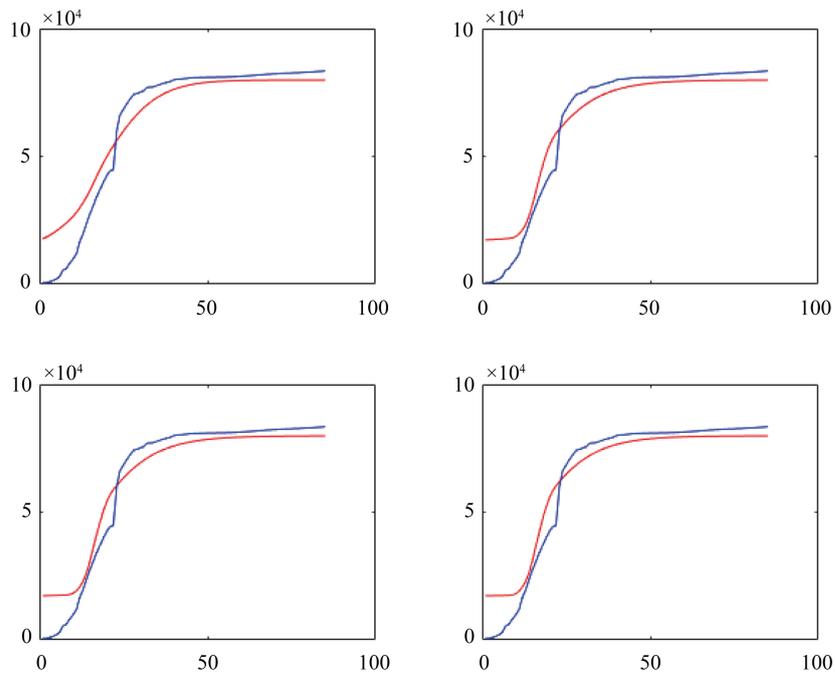


Figure 4. Adjust the parameter fitting diagram
图 4. 调参拟合图

图 4 中纵坐标表示传染人口总数, 横坐标为疫情爆发时间。从左至右, 从上到下, 依次表示医疗函数调节参数 $\beta = 1, 50, 100, 200$ 的拟合情况, 其中红色表示拟合结果, 蓝色表示真实数据。事实上, $\beta = 1$ 时, 更趋于真实情况, 这是因为疫情初期的传播速度是快速的, 不会出现平稳的情况, 上图中只有 $\beta = 1$ 是满足条件的。此时, 参数估计值 $K = 93835, q = 1.0005$ 。

根据上述建立的模型, K 表示为某地区感染该病的最大人数, 因此对于此次 COVID-19 疫情而言, 我国最大感染人数为 93,835 人, 从定性上分析我们知道: “大流行”特征所指的并不是疾病的严重性(致死率), 而是疾病传播的广泛程度(感染率), 其与流行病的传播广泛程度最大的区别在于其是在全球范围内进行的传播, 因此在考虑感染率的过程中还需加入全世界感染该病毒的国家数这一指标, 构建以下感染率模型: 其中 K 表示为某地区感染该病的最大人数, P 表示该地区人口总数, G 表示感染国家数占比, I (感染率/万人)。即

$$I = \frac{K * G}{P}$$

3. 结果分析

根据“2005~2020 年全国法定传染病报告”[7]汇总了 COVID-19 与 12 种传染病[3]的感染情况, 如表 2 所示:

Table 2. COVID-19 and 12 infectious diseases
表 2. COVID-19 与 12 种传染病感染情况

序号	传染病名称	传播途径	最大感染人数(K)
1	COVID-19	空气传播、飞沫传播	92,093
2	HINI	空气传播、飞沫传播	103,854
3	SARS	空气传播、飞沫传播	8422
4	H7N9	消化道、呼吸道	589
5	HIV/AIDS	性传播	57,194
6	风疹	空气传播、飞沫传播	120,354
7	鼠疫	空气传播、飞沫传播	22
8	麻疹	空气传播	123,136
9	流行性腮腺炎	空气传播、飞沫传播	78,233
10	霍乱	消化道传播	973
11	流行性感冒	空气传播、飞沫传播	1,141,642
12	百日咳	空气传播、飞沫传播	6658

由表 2 可知: 最大感染人数偏多的有: 流行性感冒、流行性腮腺炎、麻疹、风疹、HINI 和 COVID-19, 因此这些传染病都是被定义在流行病的范畴的。

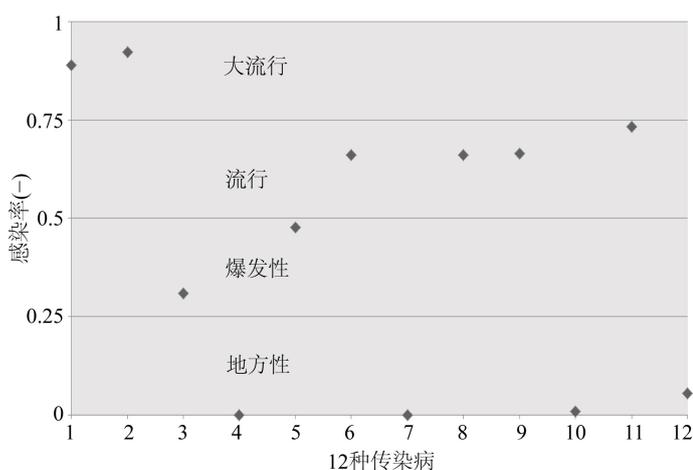
基于感染率模型, 本文主要选取了中国、美国、韩国、日本等 15 个国家, 计算出如下 12 种传染病的感染率(表 3)。

Table 3. 12 infectious diseases infection rate**表 3.** 12 种传染病感染率

序号	传染病名称	感染国家数	感染国家占比	感染率/万人(I)
1	COVID-19	15	1	0.890
2	HINI	15	1	0.923
3	SARS	15	1	0.309
4	H7N9	1	0.06	0.000
5	HIV/AIDS	15	1	0.477
6	风疹	10	0.66	0.660
7	鼠疫	15	1	0.000
8	麻疹	10	0.66	0.661
9	流行性腮腺炎	13	0.87	0.665
10	霍乱	15	1	0.008
11	流行性感冒	15	1	0.732
12	百日咳	15	1	0.055

4. 讨论

基于上述各类传染病的感染率, 参照世界卫生组织对传染病的阶段性进行划分, 获得如下结论: 当感染率(I)低于 0.3 时, 将 H7N9、鼠疫、霍乱、百日咳划分为地方性传染病; 当感染率(I)介于 0.3~0.6 之间时, 将 SARS、HIV/AIDS、流行性腮腺炎、流行性感冒划分为爆发性传染病; 当感染率(I)介于 0.6~0.9 之间时, 将风疹、麻疹、划分为流行性传染病, 最后将感染率(I)高于 0.9 的 COVID-19、HINI 划分为大流行性传染病。如图 5 所示。这一划分将传染病定性的分类赋予了定量的标准, 这对未来流行病传播风险的识别与预判提供了数学基础, 使得在今后当传染病出现时我们可以及时地将其进行归类, 从而更快地对疾病进行管控。

**Figure 5.** Classification of 12 infectious diseases**图 5.** 12 种传染病的分类

因此最终获得传染病大流行病界定的标准如下:

$$I = \frac{K * G}{P} = \begin{cases} \text{地方性传染病} & I < 0.25 \\ \text{爆发性传染病} & 0.25 \leq I < 0.5 \\ \text{流行性传染病} & 0.5 \leq I < 0.75 \\ \text{大流行传染病} & 0.75 \leq I \end{cases}$$

就中国目前的疫情发展情况来看, 国内的疫情基本已经结束, 此次疫情的感染人数与前文通过 Logistic 预测模型所预测的最大感染人数基本吻合。并且最终划分的大流行传染病与世界卫生组织所公布的名单之间也保持一致, 所以说明该传染病的量化界定是合理性的。但是, 由于在此次研究过程中, 数据有限, 各类传染病的最大感染人数均只选择国内的数据, 无法收集到全世界各个国家感染各类传染病的人数, 所以在后续研究过程中, 会不断地收集完善此类数据, 以达到一个更好的预测结果。

基金项目

本文受到贵州省科学技术基金(201942920301110098)和贵州大学线上线下混合式课程建设项目(XJG202060)资助。

参考文献

- [1] 邓小龙, 温颖. 基于传染病模型的 LPA 特征阈值社团划分方法[J]. 电子学报, 2016, 44(9): 2114-2120.
- [2] 王永炎, 王忠, 白卫国, 等. 新型冠状病毒肺炎全球大流行的因素分析与防控疫程的启示[J]. 北京中医药大学学报, 2020, 43(8): 621-622.
- [3] 潘海峰, 赵婵娜, 叶冬青. 信息流行病学研究进展[J]. 中华疾病控制杂志, 2019, 23(5): 497-500.
- [4] 于丽娜, 张燕, 王晓宇. 深圳居民健康水平评估与测控模型研究[J]. 计算机产品与流通, 2019(8): 91+164.
- [5] Liang, K. (2020) Mathematical Model of Infection Kinetics and Its Analysis for COVID-19, SARS and MERS. *Infection, Genetics and Evolution*, **82**, Article ID: 104306. <https://doi.org/10.1016/j.meegid.2020.104306>
- [6] 李立杰, 温景海, 王雅. 运用秩和比法综合评价我院医疗质量[J]. 统计与管理, 2010(3): 61.
- [7] 国家卫生健康委员会. 疫情通报[EB/OL]. <http://www.nhc.gov.cn/>, 2020-03-12.