

引用格式: 朱仁杰, 唐仕浩, 刘彤彤, 等. 基于改进 SIR 模型的新型冠状病毒肺炎疫情预测及防控对疫情发展的影响[J]. 陕西师范大学学报(自然科学版), 2020, 48(3): 33-38. [ZHU R J, TANG S H, LIU T T, et al. COVID-19 epidemic prediction based on improved SIR model and the impact of prevention and control on epidemic development[J]. Journal of Shaanxi Normal University (Natural Science Edition), 2020, 48(3): 33-38.] DOI: 10.15983/j.cnki.jsnu.2020.03.018

基于改进 SIR 模型的新型冠状病毒肺炎疫情预测及 防控对疫情发展的影响

朱仁杰, 唐仕浩, 刘彤彤, 郭 燕, 董珊珊, 成 瀛*, 杨铁林*
(西安交通大学 生命科学与技术学院 生物医学信息工程教育部重点实验室, 陕西 西安 710049)

摘 要: 基于约翰·霍普金斯大学系统科学与工程中心获取的 2020 年 3 月 4 日至 4 月 4 日期间意大利、韩国、英国、美国、法国、西班牙、德国等 7 个疫情比较严重国家的新型冠状病毒肺炎(COVID-19)确诊、死亡、治愈病例数据, 以 SIR 模型为基础, 通过添加新的参数表征感染系数随时间的变化, 并引进机器学习的方法对参数进行估计, 对 7 个国家疫情的发展情况进行模拟和预测, 并分析各国控制接触率等防治措施对于疫情发展的影响。结果表明: 改进的 SIR 模型对 COVID-19 疫情发展趋势的分析基本可靠; 除英国和美国以外, 其他 5 个国家的疫情现阶段已经得到初步控制, 而英国和美国还需要加强防控, 以减少疫情带来的损失。研究结果证实了在 COVID-19 疫情防控中, 采取减少集会、避免接触、集中收治等降低人群接触率措施的重要性, 以期后续疫情防控提供参考。

关键词: SARS-CoV-2; COVID-19; SIR 模型; 疫情分析; 接触率; 防控措施

中图分类号: R181.8; O212 **文献标志码:** A

文章编号: 1672-4291(2020)03-0033-06

开放科学(资源服务)标识码(OSID):

COVID-19 epidemic prediction based on improved SIR model and the impact of prevention and control on epidemic development

ZHU Renjie, TANG Shihao, LIU Tongtong, GUO Yan,
DONG Shanshan, CHENG Ying*, YANG Tielin*

(Key Laboratory of Biomedical Information Engineering of Ministry of Education, College of Life Science and Technology, Xi'an Jiaotong University, Xi'an 710049, Shaanxi, China)

Abstract: The data of confirmed cases, death cases and cured cases of corona virus disease 2019 (COVID-19) in seven epidemic seriously affected countries (Italy, South Korea, the United Kingdom, the United States, France, Spain, Germany) were obtained from the System Science and Engineering Center at Johns Hopkins University. Data were collected between March 4, 2020 and April 4, 2020. Traditional SIR model was used and improved by adding a new parameter to characterize the infection coefficient changes over time, and machine learning was introduced to estimate the parameters. Based on the improved SIR model, the development of COVID-19 in 7 countries were simulated and predicted, and the effects of contact rate control measures of different countries on epidemic development were analyzed. Results showed that the improved

收稿日期: 2020-04-07

基金项目: 中央高校基础科研业务费新型冠状病毒肺炎应急科研项目(xzy032020039, xzy032020023)

* 通信作者: 杨铁林, 男, 教授, 博士生导师, 从事生物信息与遗传学研究。E-mail: yangtielin@xjtu.edu.cn;

成瀛, 男, 副教授, 博士生导师, 从事分子生物学与疾病机制研究。E-mail: yingcheng@xjtu.edu.cn

SIR model was reliable for analyzing the epidemic trend of COVID-19. Apart from the United Kingdom and the United States, the other five countries have had their pandemic under initial control, while the United Kingdom and the United States need to do more to reduce the pandemic damage. Our finding confirmed the importance of reducing population contact rate in COVID-19 prevention and control, such as reducing assembly, avoiding contact and centralized admission, so as to provide reference for the subsequent epidemic prevention and control.

Keywords: SARS-CoV-2; COVID-19; SIR model; epidemic analysis; contact rate; prevention and control measures

多年来,传染病一直是世界各国医疗卫生系统面临的重大考验。目前世界主要传染病共有 64 种,涉及 82 个国家和地区。除人们熟知的流感外,麻疹和登革热涉及国家和地区均超过 20 个^[1]。2019 年,一种具有“人传人”特性的病毒性肺炎在世界多国相继出现^[2],其病原体被国际病毒分类学委员会(international committee on taxonomy of viruses, ICTV)的冠状病毒研究小组(coronaviridae study group, CSG)正式命名为严重急性呼吸综合征冠状病毒 2(severe acute respiratory syndrome coronavirus 2, SARS-CoV-2),该病毒导致的肺炎被命名为 COVID-19(corona virus disease 2019)。

根据世界卫生组织(world health organization, WHO)官网数据(<https://www.who.int/emergencies/diseases/novel-coronavirus-2019>),截至 2020 年 4 月 4 日,已有 212 个国家发现 COVID-19 确诊病例,其中 17 个国家的确诊人数超过 10 000 例。中国通过采取积极的医疗救治与预防控制措施,四月份境内疫情已经得到极为有效的控制,正常经济社会秩序正在有序恢复。但是,疫情在中国以外的国家和地区仍在快速蔓延,疫情重灾区从亚洲转变为欧洲和北美洲。截至 2020 年 4 月 4 日,美国为世界确诊人数最多的国家,COVID-19 感染病例达到 308 850 人。

自 COVID-19 疫情暴发以来,多项研究基于传染病动力学模型和相关数据对疫情发展进行分析,解读我国不同地区的疫情趋势并提出疫情防控的相关建议^[3-4]。此外,Zhan 等^[5]利用中国城市的 COVID-19 数据建立病毒传播谱,对意大利及韩国等城市疫情发展进行评估;Zhuang 等^[6]利用最大似然法估算了 COVID-19 在意大利及韩国的基本传染率并在早期对疫情进行了预测;Li 等^[7]对韩国采取防控措施前后的基本传染数分别进行了估算,并对意大利疫情做出分析;Zheng 等^[8]利用状态转移矩阵模型分析了韩国和意大利的疫情发展。也有研究对 COVID-19 的基本感染系数进行估算^[9-12],但这些研究多开展于 2020 年 3 月 5 日之前,早于疫情出现

全球性、大范围暴发的时间。

对于疫情的防控,不同国家采取了不同的措施,表现出不同的疫情发展趋势。例如,韩国在 2 月中下旬因大规模集会等原因,COVID-19 确诊人数出现大幅增加^[13],后续由于采取积极的防控措施,疫情得到有效控制;截至 2020 年 4 月 4 日,韩国累计确诊病例 10 237 例。意大利在 2020 年 3 月初 COVID-19 患者数量少于韩国,但后续每日新增确诊病例居高不下,截至 2020 年 4 月 4 日,累计确诊病例已经达到 124 632 例,远高于韩国。此外,美国、西班牙、法国、德国、英国等多个国家的病例确诊人数相继呈现出激增态势,疫情防控形势严峻。

易感-感染-移出(susceptible infected recovered, SIR)模型是研究传染病动力学的经典模型。该模型由 Kermack 与 McKendrick 在 1927 年用动力学的方法建立^[14]。本研究以 SIR 模型为基础,通过添加新的参数表征感染系数随时间的变化,同时引进机器学习的方法对参数进行估计,对多国疫情的发展趋势进行模拟和预测,并分析控制接触率对于疫情发展的影响,以期后续疫情防控提供参考。

1 研究方法

1.1 数据来源

研究数据直接来源于约翰·霍普金斯大学系统科学与工程中心(<https://systems.jhu.edu/>),并与央视新闻每日公布数据进行校核。收集 2020 年 3 月 4 日至 4 月 4 日意大利、韩国、英国、美国、法国、西班牙、德国 7 个 COVID-19 疫情严重国家的当日报告累计确诊人数、累计死亡人数和累计治愈人数,作为本研究的基础数据。

1.2 模型建立

1.2.1 SIR 模型的建立 在 SIR 模型中,人群由三部分组成:第一部分为易感者人数(susceptible),表示人群中易于被病毒感染的人数,用 S 表示;第二部分为感染者人数(infectives),表示人群中患病的人数,用 I 表示;第三部分为恢复者人数(recovered),表示患病后恢复的人数,用 R 表示;即 $N=S+I+R$ (N 为

人群总人数)。结合此次疫情特点,本研究对模型做如下假设:(1)由于各国 COVID-19 的患者数量占人群总数的比例较小,因此本研究假设易感人数始终等于人群总数,即 $S=N$ 。(2)感染系数为患者平均每天可以感染的易感者的数量,是流行病传染模型中的重要参数。感染系数=感染者每天平均接触人数×接触感染者后易感者患病的概率。通常,在疫情得到有效控制之前,各国政府会不断地加强防控措施,人群自我隔离和保护意识也会逐渐增强,造成感染者每天平均接触人数减少,感染系数也会相应减小。基于此,本研究引入一个关于时间的函数来表征接触率减小引起的人群感染系数的降低。

基于 SIR 模型的特点及本研究做出的基本假设,构建动力学函数

$$\frac{dS}{dT}=-\beta\frac{SI}{N},$$
$$\frac{dI}{dT}=\beta\frac{SI}{N}-\gamma I,$$
$$\frac{dR}{dT}=\gamma I,$$
$$\beta=at+b.$$

(1)(2)(3)(4)

其中: β 为感染系数,代表易感人群与传染源接触被感染的概率; $\beta=at+b$ 为本研究引入的关于时间的函数,用来表征感染系数随时间的变化; γ 为恢复系数,在这里取平均感染时间(14 d)的倒数。

基于 SIR 模型的基本假设,本研究在建立 COVID-19 动力学模型时,假设前期感染者和恢复者占

人口的比例可以忽略,因此认为易感人群总数始终等于总人口数,即 $S=N$ 。可以将(2)式化简为

$$\frac{dI}{dT}=\beta I-\gamma I.$$

(5)

以上微分方程求解为

$$I=Ce^{(\beta-\gamma)t}.$$

(6)

将 $t=0$ 带入(5)式中可知 $C=I_0$,即为开始时刻的感染者数量。

1.2.2 回归模型的建立 本研究利用机器学习的方法建立线性回归模型,为了便于模拟,将(5)式左右两边取自然对数得

$$\ln I=(\beta-\gamma)t+\ln I_0.$$

(7)

将公式(4)代入公式(7)中得

$$at^2+(b-\gamma)t=\ln I-\ln I_0.$$

(8)

1.3 基于机器学习的参数估计

公式(8)右边可以根据已知参数及本研究收集的数据进行计算,因此利用机器学习对公式(8)左边 $at^2+(b-\gamma)t$ 进行回归模拟,在 Python(版本 3.7.3)环境下,实现对其参数 a 和 b 的估计。

2 结果与讨论

2.1 疫情发展情况

根据收集到的连续 32 d 内(2020 年 3 月 4 日至 4 月 4 日)意大利、韩国、英国、美国、法国、西班牙、德国 7 个国家的 COVID-19 累计确诊人数、累计死亡人数和累计治愈人数,计算出各国每日现存患者的数量,绘制疫情发展数据图,结果如图 1 所示。

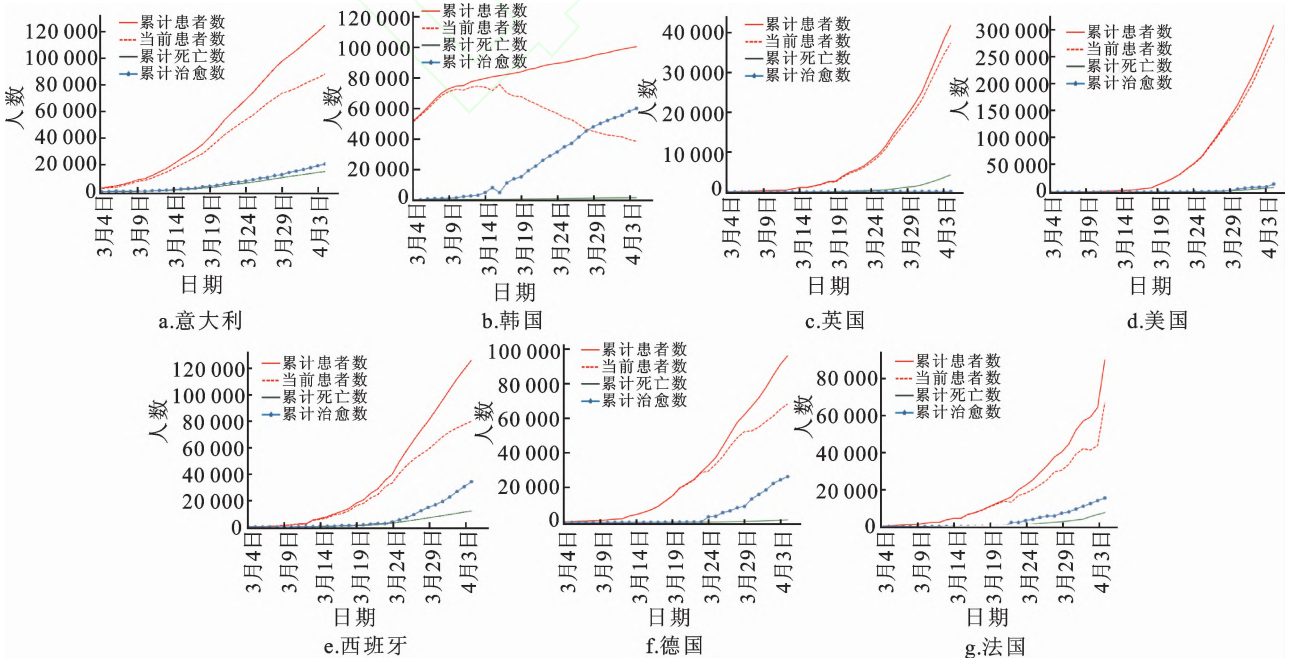


图 1 7 个国家 2020 年 3 月 4 日至 2020 年 4 月 4 日的 COVID-19 疫情发展数据

Fig.1 The COVID-19 epidemic data from March 4, 2020 to April 4, 2020 in seven countries

注:网络版为彩图。

2.2 参数估计

基于收集的 7 个疫情严重国家在 32 d 内的 COVID-19 疫情数据(图 1),利用机器学习方法进行回归分析,对参数 a 和 b 进行估计,结果如表 1 所示。可以看出,7 个国家估计的 a 值均小于 0,表示 7 个国家的感染系数均在减小,反映了防控措施不断加强以及人们自我隔离意识增强对抑制疫情传播具有积极作用。 b 值为各国在 2020 年 3 月 4 日当天的感染系数估值。

表 1 基于机器学习的参数训练值

Tab. 1 Parameters training values based on machine learning

国家	a	b
韩国	-2.016×10^{-3}	0.120 2
意大利	-3.348×10^{-3}	0.287 4
美国	-3.124×10^{-3}	0.423 6
英国	-2.298×10^{-3}	0.345 2
法国	-4.531×10^{-3}	0.378 1
西班牙	-5.395×10^{-3}	0.431 0
德国	-5.226×10^{-3}	0.415 5

2.3 疫情发展预测

基于估算的各参数值,利用改进的 SIR 模型拟合并预测意大利、韩国、英国、美国、法国、西班牙、德国 7 个国家从 2020 年 3 月 4 日到 2020 年 6 月 12 日 100 d 内的患病人数变化趋势,结果如图 2 所示。

可以看出,7 个国家利用改进的 SIR 模型预测得到的 COVID-19 患者人数与前期已知阶段的实际患病人数基本相符,可以对疫情发展进行合理预测。

表 2 展示了利用改进的 SIR 模型预测得到的 6 个国家的 COVID-19 患者数峰值及其出现的日期,由于韩国患者数已经开始下降,在此不列出。预测结果显示,韩国确诊病例数将在 5 月初降至 100 以下,综合分析,这与韩国的积极防控措施密切相关。相比于其他 6 个国家,韩国在 3 月初期的疫情形势最为严峻,2020 年 2 月下旬平均每日新增确诊人数在 500 人以上,2020 年 3 月 1 日患者数已经超过 3 700 例,但韩国政府采取的有力措施及时遏制了疫情扩散。事实上,在韩国疫情还未大范围暴发时,已有多家韩国公司根据中国公布的 SARS-CoV-2 基因序列开始研究相应诊断试剂盒;2020 年 2 月 23 日韩国将疫情警报调至最高级,并在随后一周内相继通过了传染病预防管理法、检疫法、医疗法修正案等法案,为疫情期间各项处置办法提供支撑和保障。此外,韩国在 3 月初借鉴了中国方舱医院与雷神山医院的宝贵经验,及时修改救助指南,将轻、重症患者分别安置,集中医疗资源治疗重症患者。结合本文分析结果可知,在多重措施防控下,韩国的疫情得到有效控制。

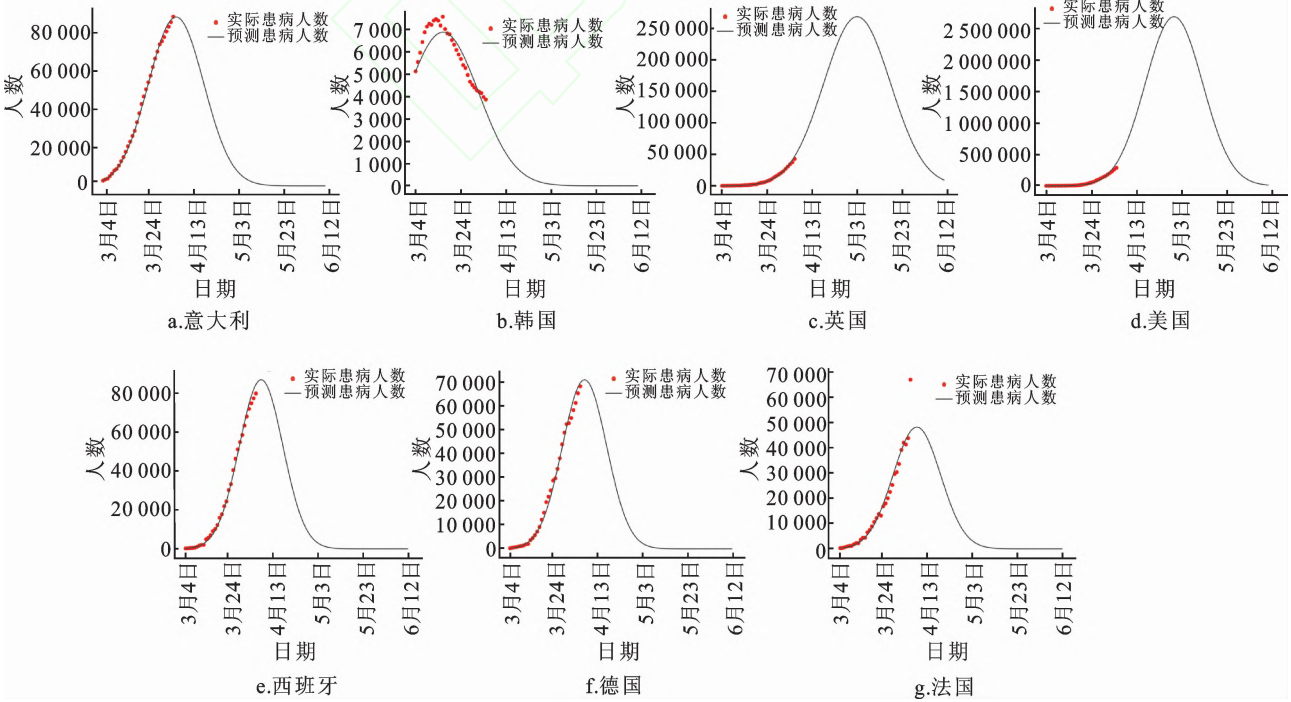


图 2 基于 SIR 模型的患病人数预测

Fig. 2 Prediction of the infection number based on improved SIR model

注:网络版为彩图。

表 2 6 国患者峰值数及峰值出现日期预测

Tab.2 Prediction of the peak number and date of the COVID-19 epidemic in six countries

国家	患者数峰值	峰值出现日期
美国	2 685 430	2020-04-29
英国	341 056	2020-05-05
法国	48 218	2020-04-07
意大利	88 121	2020-04-05
西班牙	87 103	2020-04-06
德国	71 000	2020-04-06

根据本研究的预测,意大利、西班牙、德国的疫情在 2020 年 4 月 10 日之前会得到有效控制,新增病例将逐渐呈下降趋势。2020 年 3 月 1 日,这 3 个国家的 COVID-19 患者数均不到 1 500 例,但未采取及时防治措施。虽然这 3 个国家中初期疫情最严重的意大利自 2020 年 3 月 4 日开始全国停课并在 3 月 8 日封锁全国交通,但民众未进行有效隔离,患者数持续保持较高增长,直至 3 月底每天新增病例仍然在 6 000 例以上。西班牙、德国疫情暴发较晚,但疫情发展速度快,截至 2020 年 4 月 4 日,西班牙累计确诊病例数超过意大利,达到 126 168 例,德国也达到 96 092 例,目前,3 个国家防控措施效果逐渐显现。由于法国在 2020 年 4 月 4 日集中统计了养老院患者,单日新增病例 23 100 例,导致无法对其疫情进行较为合理的预测。4 月 4 日法国累计确诊病例已经达到 89 953 例,疫情形势依然不容乐观。

当前数据显示,美国为世界上患者最多的国家,截至 2020 年 4 月 4 日,累计确诊病例达到 308 850 例。模型预测结果表明,美国的患者数将在 2020 年 4 月 29 日达到最高峰,患病人数达 2 600 000 例以上。英国的患者数会在 2020 年 5 月 5 日到达最高峰,大约为 340 000 例。基于此,两国有必要采取更为严格的措施,以防止疫情进一步蔓延。

2.4 防控措施对疫情的影响

参数 a 和 b 实际反映了感染者每天平均接触人数的大小, a 表示了 β (感染系数)随时间变化的趋势, b 则代表 β 的初值,二者主要取决于防控措施和人们的自我隔离意识。从表 1 可以看出,所有国家 a 值均小于 0,这说明各国都在加大防控力度。韩国和意大利 b 值(初始感染系数)较小,主要因为两国疫情暴发早,并较早采取了防控措施。西班牙和德国虽然初始感染系数较大,但是 a 绝对值较大,意味着两国及时加大了防控措施,疫情得到初步控制。而英国和美国,虽然初始感染系数并不是很大,但是

a 值较大,导致目前感染人数仍在高速增长。按照目前的情况,英、美两国在 4 月疫情不会明显好转。

图 3 为 2020 年 3 月 4 日起各国的感染系数在 60 d 内(截至 5 月 3 日)的变化。从图中可以看出,到 30 d 之后(即 4 月 3 日之后)英国和美国的感染系数超过其他国家,而且减少得更慢,表明此阶段两国的疫情发展仍不乐观。

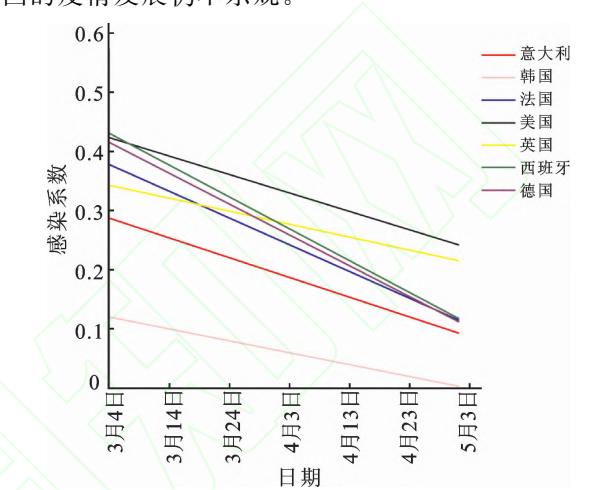


图 3 7 个国家感染系数随时间的变化

Fig.3 Infection coefficients of the seven countries

注:网络版为彩图。

2.5 中国疫情防控的借鉴意义

尽管 COVID-19 疫情最早暴发在中国,但目前来看中国在世界范围内对疫情的控制却是领先的。《Science》在 2020 年 3 月 31 日发表了中、美、英 15 家全球顶级机构、22 位科学家联合完成的研究,分析了中国在疫情最初 50 d 采取的防控措施。研究表明,仅 50 d 内,中国武汉“封城”就已经使武汉以外的全国 71 万余人免于 COVID-19 感染,使潜在感染人数减少 96%^[15],充分说明采取防控措施、减少人员接触的重要性。中国采取的防控措施主要有以下几个方面:(1)加强中央与地方的统一与协调,各级各部门及时、客观发布疫情信息和防控信息,及时发现并隔离疑似和确诊患者。实施精准管控,以现有确诊患者为依托,聚焦病例发生和可能传播的场所,对密切接触者 and 周围人群进行发散式排查,防止病毒在未知情况下迅速传播。(2)动员和选派专家及医护人员进驻疫情重灾区,集中调度防护与救治资源,通过在医疗救护、科研攻关、基础预防等方面的支持,控制疫情在重点地区的发展。(3)制定差异化的县域防控措施,强化社区防控网格化管理,科学划分疫情风险等级,明确分级分类的防控策略,防止疫情蔓延。(4)在本土疫情传播基本阻断后,实施“外

防输入、内防反弹”的总体防控策略。严格控制境外输入病例,对入境人员一律实施 14 d 医学隔离观察等措施。(5)国家卫健委根据疫情形势和临床救治工作研究进展,发布《新型冠状病毒肺炎诊疗方案》并及时更新,指导 COVID-19 救治与预防工作。

由于人群对 SARS-CoV-2 普遍易感,主要感染途径为侵入人体呼吸道上皮细胞,可通过呼吸道飞沫、气溶胶及带病毒分泌物等形式进行传播,中国境内采取的疫情监控、接触追踪、限制出行和隔离等措施,有效降低了人群的接触率,极为有效地控制了病毒的传播、遏制了疫情的发展。相较于中国为代表的东亚各国,欧美各国应对 COVID-19 疫情的优势在于有先例可循,中国遏制疫情的成功干预措施对于全世界防治 COVID-19,甚至是其他呼吸传播性疾病都具借鉴意义。

3 结论

基于 2020 年 3 月 4 日至 4 月 4 日期间意大利、韩国、英国、美国、法国、西班牙、德国 7 个国家 COVID-19 确诊、死亡、治愈病例数据,以 SIR 模型为基础,通过添加新的参数表征感染系数随时间的变化,并引进机器学习的方法对参数进行估计,对 7 国疫情的发展情况进行模拟和预测。改进的 SIR 模型对 COVID-19 疫情发展趋势的分析基本可靠;除英国和美国以外,其他 5 个国家的疫情在当前阶段已经得到初步控制,而英国和美国仍需加大防控,以减少疫情带来的损失。研究结果证实了减少集会、避免接触、集中收治等降低人群接触率的防疫措施,可以有效控制 SARS-CoV-2 病毒的传播、遏制 COVID-19 疫情的发展。研究结果将为后续疫情防控和未来类似事件应对提供参考。

参考文献:

[1] 韩辉,伍波,宋亚京,等. 2019 年 12 月全球传染病疫情概要[J]. 疾病监测, 2020, 35(1): 3-5.

[2] CHAN J F, YUAN S, KOK K H, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster[J]. The Lancet, 2020, 395 (10223): 514-523.

[3] 李承倬,武文韬,潘振宇,等. 基于 SIR 模型和基本再生数的浙江省新型冠状病毒肺炎防控效果分析[J]. 浙江医学, 2020, 42(4): 311-314.

[4] 杨雨琦,孙琦,王悦欣,等. 重庆市新型冠状病毒肺炎(NCP)疫情分析与趋势预测[J/OL]. 重庆师范大学(自

然科学版). <https://kns.cnki.net/kcms/detail/50.1165.n.20200218.0746.002.html>.

[5] ZHAN C, TSE C K, LAI Z, et al. Prediction of COVID-19 spreading profiles in South Korea, Italy and Iran by data-driven coding [J/OL]. medRxiv. DOI: 10.1101/2020.03.08.20032847.

[6] ZHAUNG Z, ZHAO S, LIN Q, et al. Preliminary estimating the reproduction number of the coronavirus disease (COVID-19) outbreak in Republic of Korea and Italy by 5 March 2020[J/OL]. medRxiv. DOI: 10.1101/2020.03.02.20030312.

[7] LI L, YANG Z, DANG Z, et al. Propagation analysis and prediction of the COVID-19 [J/OL]. medRxiv. DOI: 10.1101/2020.03.14.20036202.

[8] ZHENG Z, WU K, YAO Z, et al. The prediction for development of COVID-19 in global major epidemic areas through empirical trends in China by utilizing state transition matrix model[J/OL]. medRxiv. DOI: 10.1101/2020.03.10.20033670.

[9] LI Q, GUAN X, WU P, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia[J]. The New England Journal of Medicine, 2020, 382(13): 1199-1207.

[10] TANG B, WANG X, LI Q, et al. Estimation of the transmission risk of the 2019-nCoV and its implication for public health interventions[J]. Journal of Clinical Medicine, 2020, 9(2): 462.

[11] WU J T, LEUNG K, LEUNG G M. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study[J]. The Lancet, 2020, 395(10225): 689-697.

[12] ZHOU W, XIAO Y, HEFFERNAN J M. Optimal media reporting intensity on mitigating spread of an emerging infectious disease[J]. PLoS One, 2019, 14 (3): e213898.

[13] 李旻. 新冠肺炎疫情对韩国造成全方位冲击[J]. 世界知识, 2020(6): 24-25.

[14] KERMACK W O, MCKENDRICK A G. A contribution to the mathematical theory of epidemics[J]. Proceedings of the Royal Society of London (Series A), 2012, 115: 700-721.

[15] TIAN H, LIU Y, LI Y, et al. An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China[J]. Science, 2020, 1126(10): abb6105.