

## · 新型冠状病毒肺炎疫情防控 ·

# 新型冠状病毒肺炎疫情影响预测建模、 数据融合与防控策略分析

唐三一<sup>1</sup> 肖燕妮<sup>2</sup> 彭志行<sup>3</sup> 沈洪兵<sup>4</sup>

<sup>1</sup>陕西师范大学数学与信息科学学院, 西安 710119; <sup>2</sup>西安交通大学数学与统计学院  
数学与生命科学交叉中心 710049; <sup>3</sup>南京医科大学全球健康中心 210029; <sup>4</sup>南京医科  
大学公共卫生学院 211166

通信作者:沈洪兵, Email:hbshen@njmu.edu.cn; 彭志行, Email:zhihangpeng@njmu.edu.cn

**【摘要】** 自 2019 年 12 月以来,武汉暴发的 COVID-19 疫情由于春节人口流动快速蔓延,自 2020 年 1 月 23 日起全国大范围实施围堵缓疫策略,并不断提高检测和检出率,有效地抑制了疫情快速蔓延的趋势。在 COVID-19 爆发的早期,如何利用数学模型并结合少量和实时更新的多源数据,对疫情进行风险分析,评估防控策略的有效性和时效性等具有非常重要的现实意义。本研究将结合前期研究基础,系统介绍如何依据疫情发展的不同阶段和数据的完善,逐步建立符合我国防控策略的 COVID-19 传播动力学模型,给出模型由自治到非自治,风险评估指标由基本再生数到有效再生数,疫情发展与评估由早期的 SEIHR 传播动力学决定到最终取决于隔离人群和疑似人群规模的演变等的重要研究思路。

**【关键词】** 新型冠状病毒肺炎; 动力学模型; 基本再生数

**基金项目:**国家自然科学基金(81673275); 国家科技重大专项(2018ZX10715002-004-002, 2018ZX10713001-001)

DOI: 10.3760/cma.j.cn112338-20200216-00107

## Prediction modeling with data fusion and prevention strategy analysis for the COVID-19 outbreak

Tang Sanyi<sup>1</sup>, Xiao Yanni<sup>2</sup>, Peng Zhihang<sup>3</sup>, Shen Hongbing<sup>4</sup>

<sup>1</sup>School of Mathematics and Information Science, Shaanxi Normal University, Xi'an 710119, China; <sup>2</sup>Center for the Intersection of Mathematics and Life Sciences, School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, China; <sup>3</sup>Center for Global Health, Nanjing Medical University, Nanjing 210029, China; <sup>4</sup>School of Public Health, Nanjing Medical University, Nanjing 211166, China

Corresponding authors: Shen Hongbing, Email: hbshen@njmu.edu.cn; Peng Zhihang, Email: zhihangpeng@njmu.edu.cn

**【Abstract】** Since December 2019, the outbreak of COVID-19 in Wuhan has spread rapidly due to population movement during the Spring Festival holidays. Since January 23rd, 2020, the strategies of containment and contact tracing followed by quarantine and isolation has been implemented extensively in mainland China, and the rates of detection and confirmation have been continuously increased, which have effectively suppressed the rapid spread of the epidemic. In the early stage of the outbreak of COVID-19, it is of great practical significance to analyze the transmission risk of the epidemic and evaluate the effectiveness and timeliness of prevention and control strategies by using mathematical models and combining with a small amount of real-time updated multi-source data. On the basis of our previous research, we systematically introduce how to establish the transmission dynamic models in line with current Chinese prevention and control strategies step by step, according to the different epidemic stages and the improvement of the data. By summarized our modelling and assessing ideas, the model formulations vary from autonomous to non-autonomous dynamic systems, the risk assessment index changes from the basic regeneration number to the effective regeneration number, and the epidemic development and assessment evolve from the early SEIHR transmission model-based dynamics to the recent dynamics which are mainly associated with the variation of the isolated and suspected population sizes.

**【Key words】** COVID-19; Dynamical model; Reproductive number

**Fund programs:** National Natural Science Foundation of China (81673275); National Science and Technology Major Project of China (2018ZX10715002-004-002, 2018ZX10713001-001). DOI:10.3760/cma.j.cn112338-20200216-00107

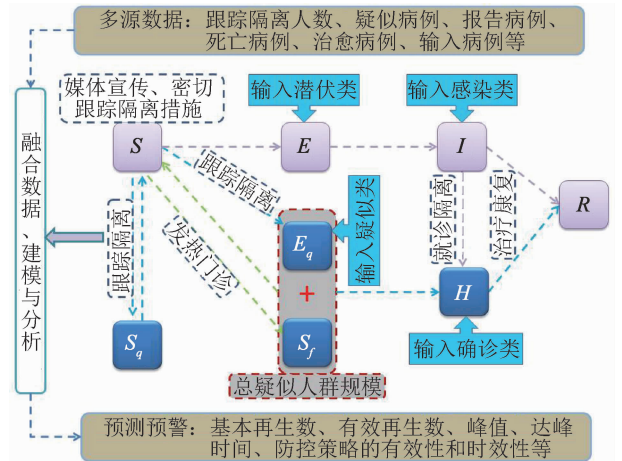
自进入 21 世纪,中国相继暴发了包括 2003 年的非典型肺炎(SARS)、2009 年的甲型流感(H1N1)、2013 年的禽流感(H7N9)和 2019 年 COVID-19 等在内的恶性传染性疾病<sup>[1-2]</sup>。而在 2019 年底暴发的 COVID-19 疫情,由武汉市首发并快速向全国蔓延,截至 2020 年 2 月 10 日 24:00,我国累计报告 42 078 例感染,死亡 1 017 例<sup>[3-4]</sup>。

突发性传染病的共同特点是突发性和无有效的治疗方法,因此,在其暴发的初期,基于数学模型和少量实时更新的数据,分析其传播能力和传播风险<sup>[5-7]</sup>,给出包括基本再生数( $R_0$ )等在内的关键技术指标,有利于疾病预防控制部门快速了解其传播力,从而采取相应的防控措施。而随着疫情的发展和控制措施的加强,针对疫情的复杂演化和围堵缓疫策略的有效性、时效性以及影响疫情严重程度的关键防控因子的甄别进行系统研究,则可为决策部门提供重要的决策依据。同时基于模型,采用敏感性和不确定性分析,评估复工、复学、局部集聚性爆发等事件对疫情防控成功的影响<sup>[8-9]</sup>,都具有非常重要的意义。

因此,根据 COVID-19 疫情暴发早期、封城策略的实施、治疗检测与检出率的不断提高,以及密切跟踪隔离和疑似病例这两类人群规模的不断攀升的事实,通过发展模型和依据实时更新的数据,结合前期的研究工作基础,本研究重点介绍如何建立数学模型分阶段分析 COVID-19 疫情传播的风险、评估不断加强防控策略措施的有效性。

一、具多防控策略的 COVID-19 疫情传播模型

COVID-19 疫情暴发以来,湖北省卫生健康委员会和国家卫生健康委员会及时通报了疫情数据,主要包括累积报告病例、累计治愈病例、累积死亡病例、跟踪隔离人数、疑似病例人数等,同时多个省份还报告了输入病例、本地病例等详细数据。上述详尽数据为建立数据驱动的 COVID-19 传播模型奠定了基础,因此,研究期间依据 COVID-19 传播的机制以及我国采取的防控策略,给出了如下具多防控策略的一般性 COVID-19 疫情传播模型示意图(图 1)。依据具有密切跟踪隔离策略建立传染动力学模型的基本思路,根据图 1 可得如下数学模型:



注:融合多源数据、构建模型、实现预测、预警和风险分析及决策评估

图 1 基于 COVID-19 传播机制、防控策略的模型构建示意图

$$\begin{cases}
 S' = -\frac{(\beta c(t) + c(t)q(t)(1-\beta))SI}{N} - mS + \lambda S_q + b(1-f)B, \\
 E' = \frac{\beta c(t)(1-q(t))SI}{N} - \sigma E + P_E(t), \\
 I' = \sigma E - (\delta_i(t) + \alpha + \gamma_I)I + P_I(t), \\
 B' = \frac{\beta c(t)q(t)SI}{N} + mS - bB + P_B(t), \\
 S'_q = \frac{(1-\beta)c(t)q(t)SI}{N} - \lambda S_q, \\
 H' = \delta_i(t)I + b_f B - (\alpha + \gamma_H)H + P_H(t), \\
 R' = \gamma_I I + \gamma_H H.
 \end{cases} \quad (1)$$

模型(1)是适合 COVID-19 传播和我国追踪隔离等防控特点的数学模型。在该模型中,自然传播过程的人群分为易感者类(S)、潜伏者类(E)、感染者类(I)、疑似病例类(B)、住院者类(H)和恢复者类(R)。由于密切跟踪隔离措施的实施,与感染者密切接触的人群分为隔离的易感者类( $S_q$ )和潜伏者类( $E_q$ )。通过感染者密切接触追踪,假设  $q$  比率的接触者被隔离,其中被隔离的个体若被感染,则该个体隔离在  $E_q$  仓室,否则隔离在  $S_q$  仓室。若比率为  $1-q$  的接触者在追踪中被遗漏,一旦被有效感染则移动到  $E$  仓室,否则仍然留在仓室  $S$  中。设每次接触时传播概率为  $\beta$ ,接触数为  $c$ 。被隔离的个体中,如果被感染(或未感染),则以  $\beta c q$  [或  $(1-\beta)c q$ ] 的速率移动到仓室  $E_q$  (或  $S_q$ )。若未被隔离且被感染,则以  $\beta c(1-q)$  的速率转移到仓室  $E$ 。感染者个体被确诊的速率为  $\delta_i$ ,并以  $\gamma_H$  的速率恢复到仓室  $R$ 。如果考虑由于

发热门诊有人进入到疑似病例仓室的实际情况,模型中统一记成仓室B。函数 $P_E(t), P_I(t), P_B(t)$ 和 $P_H(t)$ 分别表示输入到潜伏类、感染类、疑似类和确诊类的病例数,可由泊松分布刻画。模型其他参数和意义见表1。

表1 模型(1)的参数及含义

参数	定义	参数	定义
$c_0$	疾病爆发早期接触数	$f$	疑似病例的检出率
$c_b$	防控策略实施后的最小接触数	$\sigma$	从潜伏类到感染类的转化率
$r_1$	接触数的指数递减率	$\lambda$	解除隔离率
$\beta$	每次接触的传播概率	$\delta_0$	从感染者到确诊的初始确诊率
$q_0$	早期暴露着的跟踪隔离率	$\delta_H$	从感染者到确诊的最大确诊率
$q_m$	控制措施实施后的最大跟踪隔离率	$r_3$	确诊率的指数增长率
$r_2$	跟踪隔离率指数增加率	$\gamma_I$	感染者的康复率
$m$	通过发热门诊进入疑似人群的隔离率	$\gamma_H$	在院病例的康复率
$b$	疑似人群的检测率	$\alpha$	因病死亡率

在 COVID-19 疫情传播的早期(即 2020 年 1 月 23 日前),相应的防控措施比较弱,这样模型(1)中的 3 个与防控策略密切相关的 3 个函数接触数  $c(t)$ , 隔离率  $q(t)$  和确诊率  $\delta_I(t)$  都为常数,此时模型(1)为一个自治系统,由系统参数确定的传播阈值为  $R_0$ 。为了刻画不断加强的防控、确诊和检出率等,上述 3 个参数应该是基于时间的单调递减或递增函数。为了方便大家理解,简单介绍如下。

随着追踪隔离措施的加强,假设接触数是一个关于时间的递减函数  $c(t)$ ,可以定义如下:

$$c(t) = (c_0 - c_b)e^{-r_1(t-t_c)} + c_b$$

其中的阈值时间  $t_c$  为 1 月 23 日,其他参数意义参考表 1。类似地,定义隔离率为时间递增的函数  $q(t)$ ,即有

$$q(t) = (q_0 - q_m)e^{-r_2(t-t_c)} + q_m \quad (\text{满足 } q_m > q_0)$$

同样假设确诊率  $\delta_I(t)$  是一个关于时间增加的函数,那么确诊周期  $1/\delta_I(t)$  即为时间递减的函数,记为

$$\frac{1}{\delta_I(t)} = \left( \frac{1}{\delta_0} - \frac{1}{\delta_H} \right) e^{-r_3(t-t_c)} + \frac{1}{\delta_H}$$

基于上述假设,模型(1)的有效再生数可以定义为

$$R(t) = \frac{\beta c(t)(1 - q(t))}{\delta_I(t) + \alpha + \gamma_I} \quad (2)$$

当相应的率函数为常数时,上式定义了  $R_0$ 。

## 二、COVID-19 疫情的风险和防控决策分析

自 2019 年 12 月以来, COVID-19 在武汉暴发,直到 2020 年 1 月 24 日凌晨以前,相应的防控措施较少,累计报告确诊病例数基本按照疾病传播的自然状态演化,即指数增长模式。在这种条件下利用常系数的 SEIR 或 SEIHR 模型,就能较好地刻画和预测

COVID-19 疫情的动态变化,即模型(1)中的隔离率  $q$  很小或为 0 且无输入病例。文献[9]利用 2020 年 1 月 10 日至 1 月 22 日 24:00 之间的累积报告病例数,采用最小二乘法和 MCMC 方法,估计模型未知参数。基于公式(2)给出的计算公式得到 COVID-19 传播早期的  $R_0$  为 6.47(95%CI: 5.71 ~ 7.23),同时利用似然函数法(用新报告病例数)并采用极大似然估计,取续代时间为 6 时加以验证,得到了基本一致的再生数。依据确定的参数,对未来一周的疫情做了较准确的预测:截至 1 月 29 日 24:00 模型预测报告病例数为 7 723 例(实际全国报告病例数为 7 711 例)<sup>[9]</sup>。

$R_0=6.47$  是一个很大的一个值,相比 2003 年 SARS 的  $R_0=3.6$ ,说明了 COVID-19 的传播能力远大于 SARS,也说明了疫情发展的紧急性和严重性。本研究的估计值比包括 WHO 在内的组织或团队公布的  $R_0=2.2$  要高出很多,但本研究基于动力学模型和统计方法相互验证得到了几乎相同的数值,具有一定的可信度。2 月 7 日 WHO 通过分析中国约 1.7 万例患者数据后指出:COVID-19 的传染性远高于 SARS。近期基于大样本分析,有关工作相继报道了  $R_0=3.7$  和 5.5,与本研究前期结果近似,说明本研究的结果具有较强的可信性和时效性。

自 2020 年 1 月 23 日起武汉市开始实施封城策略,各地也相继采取了出行限制策略。由此可见,当第二节给出的 3 个率函数为常数时的自治模型,不能很好地刻画自 2020 年 1 月 23 日凌晨后实施的不断加强的封城、密切跟踪隔离、疑似病例的检测率、筛查率等措施。因此我们及时发展文献[9]中的模型,在图 1 中除了没有考虑由于发热门诊发现的疑似病例外,得到了能够刻画防控措施不断加强的非自治模型,通过参数估计得到模型的有效再生数,发现在我国采取围堵缓疫策略后,有效再生数随着防控措施的加强逐渐减小,即新发感染数开始逐渐下降。严控措施实施一周左右的时间后,有效再生数可能已经降到 1 以下,这充分揭示了在我国实施的封城、密切跟踪隔离措施的有效性,为政府决策提供了重要的决策依据。在不同条件下得到报告病例数将在 2 月 4—8 日左右达到峰值。同时该模型可以用来评估从 2 月 8 日开始复工复学的人员流动规模对疫情的影响,并能分析疫情再度暴发的风险。研究表明:在维持 1 月 23 日以来的防控措施下,每日 100 万 ~ 200 万的人口流动是风险较低的,但是更多的人口流动比如  $\geq 300$  万就极有可能导致二次暴发。当然非疫区的人员流动规模适当扩大,二次暴

发的风险也是比较低的。上述风险分析部分结果为相关决策机构提供了重要的决策依据。

截至 2 月 11 日 24:00 全国密切跟踪隔离累积人数超过 45 万,其中累积疑似病例超过 7 万,且与普通发热门诊病例交叉混合,对疫情未来的发展动态预测带来很大挑战。同时依据有关专家报告,得知每日报告确诊病例数很大比例来源于跟踪隔离和疑似人群,这说明我国当前 COVID-19 疫情的最终走向与这两个群体的大小有关,而且依赖针对这两个群体的检测率、检出比例等关键因子。由此可见,我国 COVID-19 的传播已经超出经典 SEIR/SEIHR 动力学模型的范畴,需要发展符合国情的防控策略的全新模型。

因此,研究过程中及时调整研究思路,创新性地提出了适合我国当前防控特点的 COVID-19 传播疫情模型,如同时考虑密切跟踪隔离的疑似病例和发热门诊发现的疑似病例(B)。该仓室就像一个黑匣子,包含了很多重要的信息却又很难明显区别开来,但却能很好地刻画检测强度和检出率对整个疫情的影响。该模型的提出,使得本研究的模型能够利用累积报告病例、累积死亡病例、累积跟踪隔离人数(不含疑似病例)、累积疑似病例等 4 列数据进行模型拟合和参数估计,并能实现多源数据之间的交叉验证,保证模型参数估计和预测的准确性。

研究结果显示,累积跟踪隔离人数和累积疑似病例这两个群体规模的稳定是我国 COVID-19 疫情峰值到来的前提。敏感性和不确定分析都显示累积跟踪隔离人数、累积疑似病例数群体规模基本趋于稳定,意味着峰值即将到来,累积报告病例数增幅减缓。为了分析疫情数据的随机性、不确定性等因素对未来疫情的影响,本研究采用泊松随机数生成方法,重新得到 1 000 组上述列举的 4 列数据,通过 1 000 次的模型拟合和参数估计,得到了相应的均值和 95%CI。结果显示,未来疫情发展中累积疑似病例和累积报告死亡病例规模不确定性相对较大,受随机性影响较强。注意到只有累积死亡病例数的预测曲线完全落在不确定性分析的 95%CI 上限的偏中间位置,因此关注重症病例提高治愈率,谨防复工复产等返程集聚性暴发的随机发生对疫情带来的冲击应该得到高度关注。为了进一步确定防控策略的有效性,特别是对完整数据区间内有效再生数的影响,我们采用相应的统计方法,得到的有效再生数估计值及其相应的 95%CI(图 2)。该结果再一次验证了严控措施实施一段时间后,有效再生数已经  $<1$ ,及新

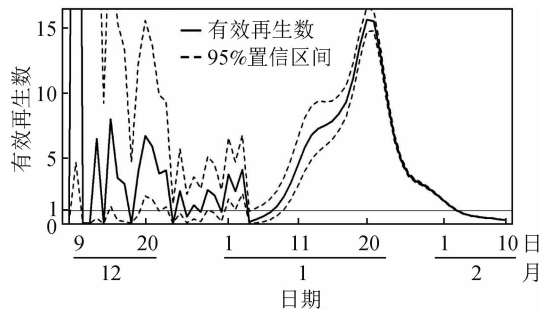


图 2 从 2019 年 12 月 9 日到 2020 年 2 月 10 日之间的有效再生数估计值及其相应的 95%CI

发感染已经得到了有效控制,目前去库存和防治二次暴发成了疫情防控的重中之重。

### 三、讨论与建议

依据图 1 给出的带有追踪隔离等防控措施的 COVID-19 传播示意图,本研究通过在前期研究中分阶段构建并发展相应的数学模型,对不同阶段的疫情以及防控策略的有效性进行了系统深入的研究,得到了很多有价值的研究结果。比如敏感性和不确定性分析等得到防疫后期应该关注重症病例并提高治愈率,谨防复工复产等返程集聚性暴发的随机事件对疫情防控带来新的冲击等,因此在当下国家不断加强隔离、检测和检出率的同时,大众持续的、严格地做好自我隔离和保护是最佳的策略。模型(1)具有一定的普适性,根据 COVID-19 的疫情和我国采用不同防控措施的时间节点,在 COVID-19 疫情发展的不同阶段如何使用该模型框架将成为讨论的重点。

首先,疫情的第一个阶段,既 2020 年 1 月 23 日前,由于封城、密切跟踪隔离措施没有实施,则模型(1)中的隔离率为常数  $q_0$ (特别地,可以假设为零),此时模型(1)简化为经典的 SEIHR 模型。这样采用 1 月 23 日前的疫情数据,可以估计模型的  $R_0$ ,得到 COVID-19 疫情早期的传播风险。

其次,在疫情的第二阶段,既 2020 年 1 月 23 日起,封城、密切跟踪隔离等措施逐步加强,确诊率和检出率逐步提高,此时模型(1)中的隔离率、接触数、确诊率等参数不再是常数,而是依赖于时间的函数。此时需要依据多源数据,估计有效再生数,通过这些率函数的敏感性分析可以评估逐步加强的控制措施的有效性以及疫情的发展趋势。

之后,在疫情的第三阶段,随着围堵缓疫策略的加强,COVID-19 疫情的传染率大幅下降(接触数明显降低),大多感染者、疑似病例都以隔离。此时 COVID-19 疫情的发展已经不是经典的传播动力学

所确定的了,应该重点关注密切跟踪隔离人群、疑似病例人群的动态变化。此时的治愈率、检出率、解除隔离率等成为敏感性分析的关键参数。

最后,在疫情的第四阶段,模型(1)中给出了输入病例对疫情的影响。但是如何分析输入病例对其他疫情不严重的省份的影响,由于报告病例相对较少,需要将模型(1)发展成为相应的离散随机模型,这为分析复工等对二次暴发风险分析提供了重要的模型支持。

通过上面由点到面的分阶段、分层次分析,才是有效的研究春节前后人员大流动对各地区甚至是全国疫情影响的最有效方法之一,也为围堵缓疫策略的依从性研究提供了模型和研究思路。上面依据疫情动态发展、防控措施的实施等适时建立和发展相应的传播动力学模型,能够有效预测疫情的动态发展,为疫情的传播风险提供预警,具有重要的理论和现实意义。

如何基于突发性传染病的传播机制、防控措施的实施,构建适合中国防控特色的数学模型并服务于地方和国家防控策略是具有挑战性的问题。比如根据传统的 SEIR 模型并依据单一的数据源,对我国疫情的风险进行预测和评估这本身就具有一定的风险性,这完全有可能低估或高估疫情。特别是传统的 SEIR 模型不能有效刻画各种防控策略的联合作用,因而不能综合评价疫情防控措施的有效性和时效性,这样可能得出比如封城效果不明显不准确的结论。

**利益冲突** 所有作者均声明不存在利益冲突

### 参 考 文 献

- [1] WHO. Novel Coronavirus-China, Disease outbreak news: Update [EB/OL]. (2020-01-12) [2020-01-23]. <https://www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/>.
- [2] WHO. Situation report [EB/OL]. (2020-01-23) [2020-02-01]. <https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200123-sitrep-3-2019-ncov.pdf>.
- 湖北省卫生健康委员会. 湖北省新冠肺炎疫情情况 [EB/OL].

(2020-01-23) [2020-02-01]. <http://wjw.hubei.gov.cn/bmdt/ztzl/fkxgzbdgrfyyq/>.

Health Commission of Hubei Province. COVID-19 situation report for Hubei province [EB/OL]. (2020-01-23) [2020-02-01]. <http://wjw.hubei.gov.cn/bmdt/ztzl/fkxgzbdgrfyyq/>.

- [3] 中华预防医学会新型冠状病毒肺炎防控专家组. 新型冠状病毒肺炎流行病学特征的最新认识 [J]. 中华流行病学杂志, 2020, 41 (2): 139-144. DOI: 10.3760/cma.j.issn.0254-6450.2020.02.002.
- Special Expert Group for Control of the Epidemic of Novel Coronavirus Pneumonia of the Chinese Preventive Medicine. An update on the epidemiological characteristics of novel coronavirus pneumonia (COVID-19) [J]. Chin J Epidemiol, 2020, 41 (2): 139-144. DOI: 10.3760/cma.j.issn.0254-6450.2020.02.002.
- [4] 国家卫生健康委员会. 截至 2 月 10 日 24 时新型冠状病毒肺炎疫情最新情况. <http://www.nhc.gov.cn/xcs/yqtb/202002/4a611bc7fa20411f8ba1c8084426c0d4.shtml>. 2020 年 2 月 11 日.
- National Health Commission of the People's Republic of China. An update of Novel Coronavirus Pneumonia outbreak as of 24:00 on 10 February. <http://www.nhc.gov.cn/xcs/yqtb/202002/4a611bc7fa20411f8ba1c8084426c0d4.shtml>. 11 February 2020.
- [5] Huang CL, Wang YM, Li XW, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China [J]. Lancet, 2020, 395(10223): 497-506. DOI: 10.1016/S0140-6736(20)30183-5.
- [6] Chen NS, Zhou M, Dong X, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study [J]. Lancet, 2020, 395(10223): 507-513. DOI: 10.1016/S0140-6736(20)30211-7.
- [7] Li Q, Guan X, Wu P, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected Pneumonia [J]. N Engl J Med, 2020. DOI: 10.1056/NEJMoa2001316.
- [8] 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]. J Clin Med, 2020, 9 (2): 462. DOI: 10.3390/jcm9020462.
- [9] Tang B, Bragazzib NL, Li Q, et al. An updated estimation of the risk of transmission of the novel coronavirus (2019-nCoV) [J]. Infect Dis Model, 2020, 5: 248-255. DOI: 10.1016/j.idm.2020.02.001.

(收稿日期: 2020-02-16)

(本文编辑: 李银鸽)