

# 新型冠状病毒肺炎疫情确诊病例的统计分析 及自回归建模\*

曹文静<sup>1)</sup> 刘小菲<sup>1)</sup> 韩卓<sup>2)</sup> 冯鑫<sup>2)†</sup> 张琳<sup>1)‡</sup>  
刘肖凡<sup>3)</sup> 许小可<sup>4)</sup> 吴晔<sup>5)</sup>

1) (北京邮电大学理学院, 北京 100876)

2) (河北地质大学管理科学与工程学院, 石家庄 050031)

3) (香港城市大学媒体与传播系, 香港 999077)

4) (大连民族大学信息与通信工程学院, 大连 116600)

5) (北京师范大学计算传播学研究中心, 北京 100875)

(2020年4月6日收到; 2020年4月24日收到修改稿)

基于安徽省卫生健康委员会截至2020年2月19日公布的800余例新型冠状病毒肺炎病例信息, 根据病例中公布的接触史构建确诊患者间的有向传播关系, 发现源传染患者中男性居多, 被传染患者中女性居多. 从病例信息中可知, 安徽省新型冠状病毒肺炎疫情的发展从初期的具有武汉居住或接触史的输入病例转入后期本地传播为主的小范围社区传播, 且严格的防控隔离措施有效切断了社区内的进一步传播. 源传染患者与被传染患者的确诊时间间隔可用 $T$ 分布拟合, 确诊时间间隔的中位数为2 d, 平均值为2.67 d. 基于有向传播关系的统计特点, 构建安徽省疫情发展后期的自回归传播模型, 模型仿真结果与疫情发展数据符合. 对除湖北省的全国确诊病例数据同样采取自回归建模与仿真, 结果仍与疫情发展数据符合. 这一发现为控制疫情在湖北省以外区域的防控提供了参考: 通过严格的防控措施和隔离措施, 疫情在湖北省之外的传播具有很大的黏滞性, 多为家庭程度的密切接触传播, 且能有效控制新型冠状病毒肺炎在当地的传播深度, 有效控制了疫情的蔓延.

**关键词:** 新型冠状病毒肺炎疫情, 缓慢增长期, 有向传播关系, 自回归

**PACS:** 02.50.-r, 02.60.Ed, 05.10.-a

**DOI:** 10.7498/aps.69.20200503

## 1 引言

2020年1月23日, 武汉封城当日, 湖北省累计确诊549例, 2020年2月23日, 湖北省累计确诊64084例, 累计死亡2346例, 确诊人数是一个月前前的100倍以上. 2020年3月, 疫情在亚洲、欧

洲、北美洲大肆爆发, 截至2020年3月20日, 全球累计确诊人数已超过24万, 累计死亡近2万人……, 相比2003年的SARS病毒最终全球感染8422例, 死亡349例, 新型冠状病毒肺炎(简称为新冠肺炎)疫情的目前严重程度已超过其30倍<sup>[1,2]</sup>. “钻石公主”号邮轮感染率逾20%<sup>[3]</sup>. 环球网2月24日报道, 战“疫”一线感染新冠肺炎病毒的医护

\* 国家自然科学基金(批准号: 11875005, 61976025, 11905042)、教育部人文社会科学研究青年基金(批准号: 16YJC630022)和河北省普通高等学校青年拔尖人才计划(批准号: BJ2018111)资助的课题.

† 通信作者. E-mail: 149987543@qq.com

‡ 通信作者. E-mail: zhanglin2011@bupt.edu.cn

工作者已达 3387 例<sup>[4]</sup>……新冠肺炎病毒传染性强、潜伏期长、治疗期长,并且在无症状的潜伏期具有传染性<sup>[5-7]</sup>,使得疫情的防控工作困难重重。

面对疫情,一方面举全国医护工作者驰援武汉;另一方面,严格防控部署,各地居民小区实施严格管控措施,严防输入,十数亿民众主动居家隔离,企事业单位远程办公,全国大中小学停课在线学习;同时,及时追踪确诊患者的密切接触者并进行隔离观察……诸多措施,极大程度上斩断了新冠肺炎病毒的传播途径。目前,这场战“疫”,已经取得了阶段性的显著成效。全国各地每日新增确诊病例数降幅明显,目前新增病例均为海外输入病例,本地病例暂时已经实现零增长。随着疫情信息通报透明化,各地卫生健康委员会网站以及官方媒体及时公开确诊患者的患病过程信息及行动轨迹,不仅有助于缓解群众的恐慌心理,更有利于追踪密切接触人群进行及时隔离,阻断传播途径,进一步遏制疫情。

目前已有许多论文讨论新冠肺炎疫情,主要分为两方面。一方面是从疾病控制角度,研究平均潜伏期和基本再生数<sup>[8]</sup>,病例的临床统计信息<sup>[9-11]</sup>等。另一方面,以疫情传播的建模和预测为主<sup>[12-15]</sup>,主要以 SEIR 模型<sup>[16]</sup>和数据拟合为基础,加以适当的异质性假设,一定程度上能够复现疫情传播规律和规模,为疫情发展和防控提供参考。

然而 SEIR 模型与湖北省以外地区的疫情实际发展存在明显差异。自 2020 年 1 月 23 日武汉封城开始,至第一个 14 天的隔离期结束的 2020 年 2 月 5 日,各地区的输入病例已经得到了有效的隔离和控制,此后病毒绝大多数在家庭级别的密切接触程度内传播,传播广度与深度均被切断。在 SEIR 模型中,即使加入时滞和受控因素,其内在的根本逻辑仍是病毒经历很多代际的演化传播:SEIR 模型的演化方程每迭代一次,等价于病毒向前传播了一代<sup>[17-19]</sup>。而实际情况是,尽管有少数聚集病例经历多个代际的传播,大部分患者只经历了一至两代传播就得到了有效隔离和救治,也就是说,病毒从输入患者到本地患者,再在本地患者间传播的次数,仅能有少数几次的传播机会,不具备 SEIR 模型迭代所需要的传播深度。因此,基于 SIR 模型<sup>[20]</sup>的内在传染机制难以适用于新冠肺炎病毒在国内除湖北省以外地区演化方式的建模与

分析。

本文根据安徽省卫生健康委员会公开的 800 余病例信息,得到有效的病缘传播关系病例数据 233 份,分析了每日新增病例数中性别比例规律以及输入病例比例的规律。同时通过对有病缘关系的确诊时间间隔进行分析发现了其统计特征,并根据其统计特点对安徽省确诊数据和全国除湖北省的确诊数据进行了建模。根据疫情传播的真实情况和确诊时间间隔的统计特点,采用自回归模型建模,有效刻画了疫情发展中后期的演化规律,为理解疫情传播后期的内在机制提供帮助,为疫情防控决策提供参考。

## 2 数据信息

据截至 2020 年 2 月 19 日安徽省卫生健康委员会公布具体病例文本信息 887 例,统计每日新增病例的性别、是否具有武汉接触或居住史。并从 887 例病例中提取有效的源传染患者与被传染患者的关联信息 233 条,进行进一步分析与建模。

## 3 数据结论

### 3.1 不同性别每日新增人数

将每日新增确诊人数按性别分别统计,结果如图 1 所示。从图 1 可见女性感染人数的变化趋势滞后于男性,初步判断男性先感染,随后传染给家庭内共同生活的女性。每日新增确诊人数中男性占比如图 2 所示。开始阶段男性占比明显高于女性,在传染进入后期,比例下降至 0.5 左右,与图 1 结论符合。

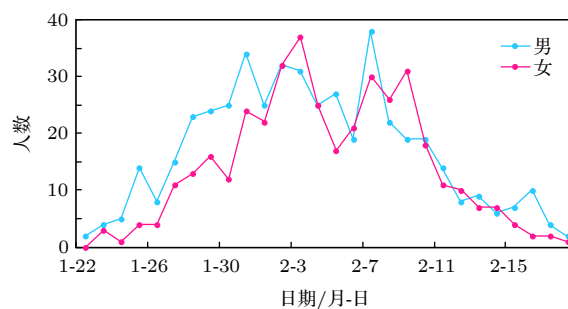


图 1 不同性别每日新增确诊人数

Fig. 1. Number of newly confirmed patients per day of different genders.

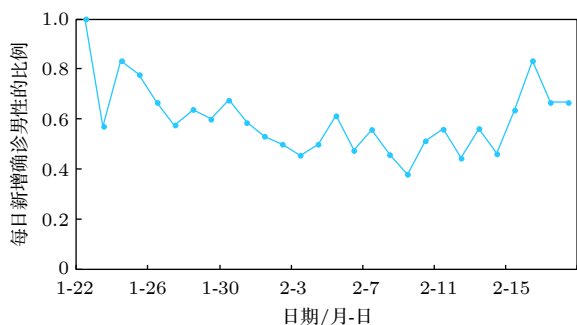


图 2 每日新增确诊男性的比例

Fig. 2. Proportion of newly confirmed males per day.

### 3.2 输入病例与本地传播关系

区分每日新增病例中, 输入病例与本地传播比例随时间变化趋势, 如图 3 所示. 从图 3 可见, 具有武汉居住和直接接触史的输入病例变化趋势领先于本地传播趋势, 输入病例防控严格, 武汉封城后, 快速降低. 后续病例转为本地传播, 并快速有效隔离. 图 4 为每日新增确诊输入病例的比例, 随着时间推移而下降至趋于 0.

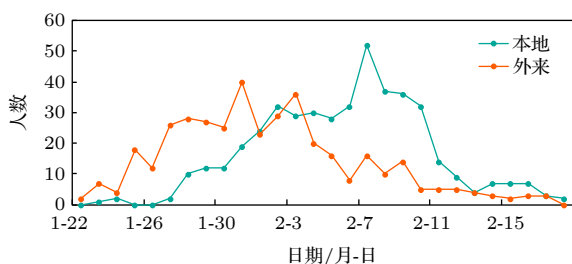


图 3 不同来源每日新增确诊人数

Fig. 3. Number of newly confirmed patients per day from different sources.

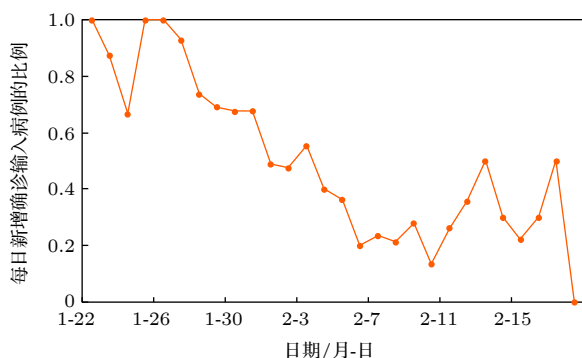


图 4 每日新增确诊输入病例的比例

Fig. 4. Proportion of newly confirmed imported cases per day.

### 3.3 有向传播关系

从 3.1 节结果可发现男性先感染, 女性后感染; 3.2 节内容展示了疫情由输入转为本地传播. 本节

进一步从病例数据中挖掘传染过程的病缘关系, 深入刻画源传染患者与被传染患者间的统计信息. 统计结果如表 1 和表 2 所列. 表 1 显示, 在病缘关系的性别分析中, 男性传染给女性的比例最大, 为 35.62%. 表 2 显示, 丈夫传染给妻子占比最多, 为 15.38%, 是第二类妻子传染给丈夫的 1.7 倍. 如此进一步确认了男性先感染, 并随后传染给女性占比最多. 从表 2 也能发现, 在计入表 2 的感染关系大于等于 5 例的排序中, 共有 9 类关系计入, 并且全都是核心家庭关系, 发生在夫妻、父母与子女、以及兄弟之间. 由此可见病毒在家庭内部的传染性之强, 在居家环境下, 几乎无法隔离.

表 1 传播过程中源传染患者与被传染患者性别统计 (有效样本 233 例)

Table 1. Gender statistics of source infected patients and infected patients during virus transmission. The number of valid sample is 233.

源传染者-被传染者	人数	比例
男-女	83	0.3562
男-男	54	0.2318
女-男	56	0.2403
女-女	40	0.1717

表 2 病毒传播过程中源传染患者与被传染患者关系统计 (有效样本 233, 仅截取统计样本量大于等于 5 的情况, 其中比例为每个类别内的人数除以有效样本数 233 得到)

Table 2. Relationship statistics between source infected patients and infected patients during virus transmission. The number of valid sample is 233. Use only statistical sample size greater than or equal to 5. The ratio is the number of people in each category divided by valid sample number 233.

源传染者-被传染者	人数	比例
丈夫-妻子	36	0.1538
妻子-丈夫	21	0.0897
儿子-母亲	11	0.0470
母亲-儿子	10	0.0427
儿子-父亲	10	0.0427
母亲-女儿	7	0.0299
父亲-女儿	5	0.0214
哥哥-弟弟	5	0.0214
父亲-儿子	5	0.0214

### 3.4 病缘关系确诊时间差的分布

在病缘关系的有效病例样本数据中, 进一步提取确诊时间差, 得确诊时间差的分布直方图如图 5 所示. 确诊日期相差 1 d 和 3 d 居多. 疫情转为本

地传播后,进一步的传播普遍发生在核心家庭内部. 实施严格的防控措施,能够有效地隔离具有传染性的确诊患者,并且通过进一步地追踪其密切接触者并及时隔离,能够有效地抑制疫情的进一步传播扩散.

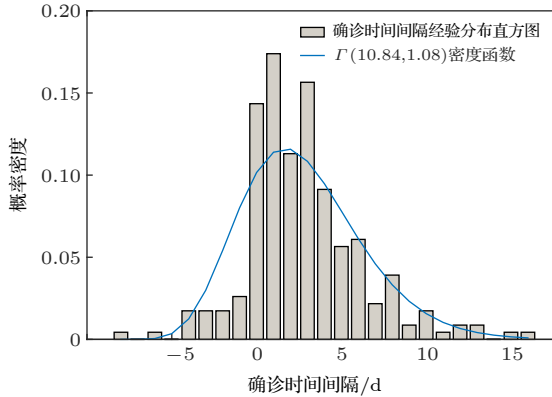


图5 源传染患者与被传染患者间确诊时间差的分布直方图,图中拟合密度参考线为  $\Gamma(10.84, 1.08)$  分布,并向左平移9

Fig. 5. Distribution histogram of the time interval of diagnosis between source infected patients and infected patients. In the figure, the fitting density reference line is the distribution of  $\Gamma(10.84, 1.08)$  and shifts 9 to the left.

在严格的防控机制下,基于SIR传播模型的迭代逻辑对于国内除湖北省以外地区的传播并不成立. 感染者并不是经历多代传播才患病的,而是从输入病例开始,加之较长的潜伏期,潜伏期内无症状传播,以及病毒在无法彻底隔离的共同生活的家人间传播,共同作用的结果. 得益于严防死守的隔离措施,患者一经发病就能够及时隔离治疗,并有效追踪密切接触者,彻底切断病毒传播途径.

对确诊时间间隔的频率直方图采用  $\Gamma$  分布进行拟合. 由于  $\Gamma$  分布的取值为正值,故将确诊时间数据用公式  $x - \min(x) + 1$  进行平移,其中  $x$  表示确诊时间间隔,得到拟合密度函数为  $\Gamma(10.84, 1.08)$ ,再平移还原回原始确诊时间间隔数据. 出现负值的确诊时间间隔,证实了病毒在无症状的潜伏期具有传染性,且由于患者的免疫力不同,病毒的潜伏期不同.  $\Gamma$  分布的拟合与负值确诊时间间隔都与已有的研究结果一致 [21].

## 4 模型与仿真

### 4.1 模型

结合第2部分的数据分析结果可知,在具有武

汉居住和接触史的输入病例发病确诊并隔离后,疫情的演化主要由继发感染者决定. 而继发感染者的传播普遍在家庭内部发生,且继发传染被有效的隔离措施阻断. 基于图5的统计结果,源传染者与被传染者确诊时间间隔主要集中在1—3 d. 本节对安徽省和全国除湖北省以外的每日新增确诊人数序列进行自回归迭代建模. 考虑到每日新增人数序列与其滞后1—3 d序列间的强相关关系,采用自回归模型进行建模.

自回归模型是处理时间序列常用方法之一,广泛应用于经济、信息、自然现象等领域的建模与预测. 模型基本形式为

$$X_t = c + \sum_{i=1}^p \varphi_i X_{t-i} + \varepsilon_t, \quad (1)$$

其中  $c$  为常数项,  $\varepsilon_t$  为随机误差,  $\varphi_i$  为  $i$  阶滞后的系数. 结合上文确诊时间间隔的分布特征,取  $p = 3$ . 将对安徽省和全国除湖北省以外每日新增确诊人数分别进行自回归建模,用来展示新冠肺炎在严格防控的情况下,在非湖北省的传播情况.

### 4.2 模型与仿真结论

安徽省每日新增数据的自回归拟合方程如下:

$$\begin{aligned} C_{t+1} - C_t &= 0.166(C_{t-2} - C_{t-3}) + 0.599(C_{t-1} - C_{t-2}) \\ &\quad - 0.059(C_t - C_{t-1}) - 2.763, \end{aligned} \quad (2)$$

其中  $C_t$  为安徽省累计确诊人数,数据从2020年2月6日开始. 初值取2月6日至2月9日累计确诊人数,代入(2)式迭代,得2月10日以后的每日新增确诊人数和累计确诊人数的仿真数据. 结果分别如图6和图7所示. 仿真数据与真实数据高度符合.

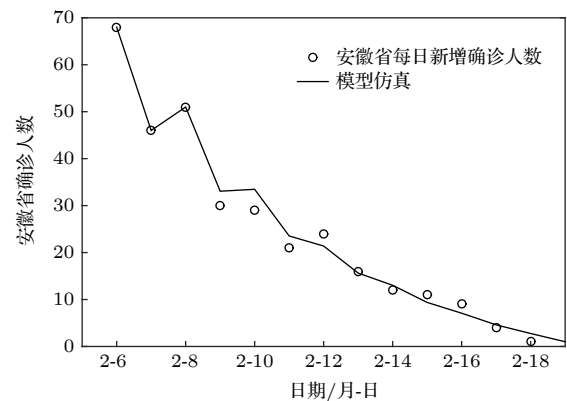


图6 安徽省每日新增确诊人数及模型仿真

Fig. 6. Number of newly confirmed patients per day and model simulation in Anhui province.



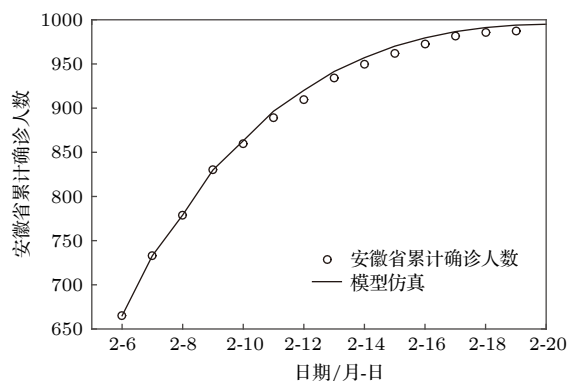


图7 安徽省累计确诊人数及模型仿真

Fig. 7. Cumulative number of confirmed cases and model simulation in Anhui province.

将安徽省的结论推广至全国, 对全国除湖北省的确诊病例数同样进行自回归建模, 拟合方程如下式:

$$\begin{aligned}
 & C_{t+1} - C_t \\
 &= 0.575(C_{t-2} - C_{t-3}) + 0.104(C_{t-1} - C_{t-2}) \\
 &+ 0.222(C_t - C_{t-1}) - 81.8,
 \end{aligned} \tag{3}$$

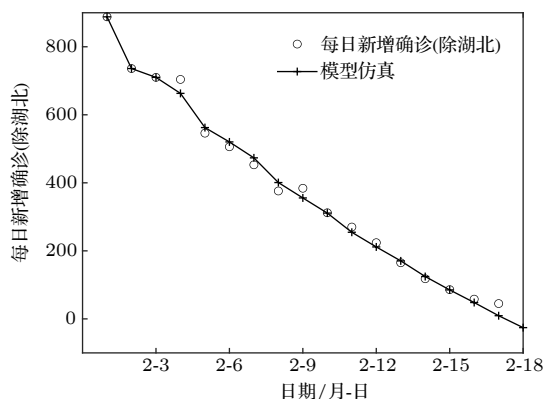


图8 每日新增确诊人数(除湖北省)及模型仿真

Fig. 8. Number of newly confirmed patients per day (except in Hubei province) and model simulation.

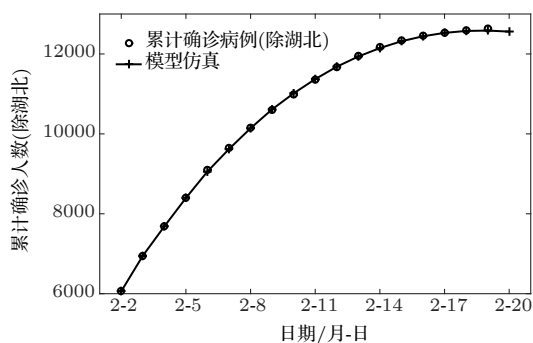


图9 累计确诊人数(除湖北省)及模型仿真

Fig. 9. Cumulative number of confirmed cases (excluding Hubei province) and model simulation.

其中  $C_t$  为除湖北省以外的全国累计确诊人数, 数据从 2020 年 2 月 2 日开始. 初值取 2 月 2 日至 2 月 5 日累计确诊人数, 代入 (3) 式迭代, 得 2 月 6 日以后的每日新增确诊人数和累计确诊人数的仿真数据. 结果分别如图 8 和图 9 所示. 仿真数据与真实数据高度符合.

## 5 讨论

本文对从安徽省 800 余确诊病例数据中提取的新型冠状病毒传播的统计特征和有向传播关系进行分析发现, 源传染患者中男性居多, 被传染患者中女性居多, 发生最多的为丈夫传给妻子. 疫情的发展从早期的具有武汉居住或接触史的输入病例转入本地传播为主的小范围传播. 传染过程中, 源传染患者与被传染患者间的确诊时间差的中位数为 2 d. 鉴于 SIR 模型的迭代机制的不适用性, 本文基于有向传播关系的统计特点和新冠肺炎防控的实际情况, 构建了安徽省和全国除湖北地区在疫情发展后期的自回归模型, 模型仿真结果与疫情发展数据能很好地符合. 通过病例数据中的发现, 建立符合实际情况的自回归模型, 能够为新冠肺炎疫情的传播机制提供参考. 同时为疫情的总体防控决策提供参考. 新冠肺炎疫情传染性强, 且潜伏期的无症状状态也具有传染性, 使得其防控形势困难重重. 中国的经验表明, 严格的防控措施是有效抑制新冠肺炎疫情扩散最重要的手段.

## 参考文献

- [1] National Health Commission of the People's Republic of China <http://www.nhc.gov.cn/> [2020-03-29]
- [2] Huaxia.com <http://www.huaxia.com/xw/dlxw/2003/08/234782.html> [2020-03-29]
- [3] Sina.com [https://news.sina.cn/zt\\_d/yiqing0121](https://news.sina.cn/zt_d/yiqing0121) [2020-03-29]
- [4] Sina.com [https://weibointl.api.weibo.cn/share/129902340.htm?weibo\\_id=4475585516343394](https://weibointl.api.weibo.cn/share/129902340.htm?weibo_id=4475585516343394) [2020-03-29]
- [5] Li C M, Zhong X Z, Yang J C 2004 *Chongqing Med. J.* **395** 947 (in Chinese) [李春梅, 钟晓祝, 杨金城 2004 *重庆医学* **395** 947]
- [6] Wu J T, Leung K, Leung G M 2020 *Lancet* **365** 689
- [7] Zhou T, Liu Q H, Yang Z M, Liao J Y, Yang K X, Bai W, Liu X, Zhang W 2020 *Chin. J. Evid.-Based Med.* **20** 117 (in Chinese) [周涛, 刘权辉, 杨紫陌, 廖敬仪, 杨可心, 白薇, 吕欣, 张伟 2020 *中国循证医学杂志* **20** 117]
- [8] Jasper F W C, Yuan S F, Kinhang K, Kelvin K W T, Hin C, Yang J, Xing F F, Liu J L, Cyril C Y, Rosana W S P, Hoiwah T, Simon K L, Kwokhung C, Vincent K P, Wanmui C, Jonathan D I, Jianpiao C, Vincent C C C, Honglin C, Christopher H, Kwokyung Y 2020 *Lancet* **395** 514

- [9] Wang C, Peter W H, Frederick G H, George F G 2020 *Lancet* **395** 470
- [10] Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, Liu L, Shan H, Lei C, Hui D S C, Du B, Li L, Zeng G, Yuen K Y, Chen R, Tang C, Wang T, Chen P, Xiang J, Li S, Wang J L, Liang Z, Peng Y, Wei L, Liu Y, Hu Y H, Peng P, Wang J M, Liu J, Chen Z, Li G, Zheng Z, Qiu S, Luo J, Ye C, Zhu S, Zhong N 2020 *The New England Journal of Medicine* DOI: 10.1056/NEJMoa2002032
- [11] Zhang L H, Li B H, Jia P, Pu j, Bai P, Zhu P J, Li L, Zeng G J, Zhao X, Dong S S, Liu M H, Zhang N 2020 *J. Biomedical Eng.* **37** 1 (in Chinese) [张龙浩, 李柏宏, 贾鹏, 蒲剑, 白蓓, 李音, 朱培嘉, 李雷, 曾国军, 赵欣, 董珊珊, 刘梦茵, 张楠 2020 *生物医学工程学杂志* **37** 1]
- [12] Maria N, Niamh O'Neill, Catrin S, Mehdi K, Maliha A, Riaz Ag 2020 *Int. J. Surg.* **76** 71
- [13] Yu Z, Zhang G Q, Liu Q Z, Lv Z Q 2020 *Journal of University of Electronic Science and Technology of China* DOI: 10.12178/1001-0548.2020027 (in Chinese) [喻孜, 张贵清, 刘庆珍, 吕忠全 2020 *电子科技大学学报* DOI: 10.12178/1001-0548.2020027]
- [14] Gerardo C, Cécile V, Simonsen L, Moghadas S M 2016 *The Royal Society. Collection* DOI: 10.6084/m9.figshare.c.3473634.v2
- [15] Zhang L 2020 *Journal of University of Electronic Science and Technology of China* DOI: 10.12178/1001-0548.2020037 (in Chinese) [张琳 2020 *电子科技大学学报* DOI: 10.12178/1001-0548.2020037]
- [16] Benjamin F M, Dirk B 2020 arXiv: 2002.07572 [q-bio.PE]
- [17] Liu Q, Jiang D Q, Tasawar H, Ahmed A, Bashir A 2020 *Physica A* DOI: 10.1016/j.physa.2020.124152
- [18] Cai J, Jia H Y, Wang K 2020 *Shandong Med. J.* **60** 1 (in Chinese) [蔡洁, 贾浩源, 王珂 2020 *山东医药* **60** 1]
- [19] Luo Y T, Tang S T, Teng Z D, Zhang L 2019 *Nonlinear Anal. Real World Appl.* **50** 365
- [20] Zhang Y P 2020 *China J. Epidemiol.* **41** 145 (in Chinese) [张彦平 2020 *中华流行病学杂志* **41** 145]
- [21] Du Z W, Xu X K, Wu Y, Wang L, Cowling B J, Meyers A L 2020 *Emerg. Infect. Dis.* DOI: 10.3201/eid2606.200357

# Statistical analysis and autoregressive modeling of confirmed coronavirus disease 2019 epidemic cases<sup>\*</sup>

Cao Wen-Jing<sup>1)</sup> Liu Xiao-Fei<sup>1)</sup> Han Zhuo<sup>2)</sup> Feng Xin<sup>2)†</sup> Zhang Lin<sup>1)‡</sup>  
Liu Xiao-Fan<sup>3)</sup> Xu Xiao-Ke<sup>4)</sup> Wu Ye<sup>5)</sup>

1) (*School of Science, Beijing University of Posts and Telecommunications, Beijing 100876, China*)

2) (*School of Management Science and Engineering, Hebei GEO University, Shijiazhuang 050031, China*)

3) (*School of Media and Communication, City University of Hong Kong, Hongkong 999077, China*)

4) (*School of Information and Telecommunication Engineering, Dalian Minzu University, Dalian 116600, China*)

5) (*Computational Communication Research Center, Beijing Normal University, Beijing 100875, China*)

( Received 6 April 2020; revised manuscript received 24 April 2020 )

## Abstract

Based on the information about more than 800 cases reported by Anhui provincial health commission on February 19 2020, the directional transmission relationship between the confirmed patients is constructed according to the contact history published in the cases, and it is found that the majority of the patients who can infect others are male and most of the patients who are infected are female. According to the analysis of case information, the early confirmed coronavirus disease 2019 (COVID-19) cases in Anhui province had the history of residence or exposure in Wuhan. In the later stage, the cases spread in small communities mainly by local transmission, and the strict prevention and control measures can effectively cut off further transmission in the communities. The time interval between diagnosing the source-infected patients and infected patients is fitted by  $\Gamma$  distribution, with a median of 2 days and an average of 2.67 days. Based on the statistical characteristics of directional transmission relationship, an autoregressive transmission model is constructed in the late stage of epidemic development in Anhui province, and the simulation results are consistent with the epidemic development data. Autoregressive model and simulation are also used for predicting the data of confirmed cases in the whole country except for Hubei province. This discovery can be referenced by regional epidemic prevention and control except for where it originated. Through strict protection measures and isolation measures, the spread of the epidemic outside the original place is highly viscous. It usually spreads by close contact between family members, and the local spread of COVID-19 can be effectively controlled.

**Keywords:** coronavirus disease 2019 epidemic, slow growth period, directed transmission, autoregression

**PACS:** 02.50.-r, 02.60.Ed, 05.10.-a

**DOI:** 10.7498/aps.69.20200503

<sup>\*</sup> Project supported by the National Natural Science Foundation of China (Grant Nos. 11875005, 61976025, 11905042), the Humanity and Social Science Youth foundation of Ministry of Education of China (Grant No. 16YJC630022), and the Top-notch Talent Project of Hebei Education Department, China (Grant No. BJ2018111) .

<sup>†</sup> Corresponding author. E-mail: [149987543@qq.com](mailto:149987543@qq.com)

<sup>‡</sup> Corresponding author. E-mail: [zhanglin2011@bupt.edu.cn](mailto:zhanglin2011@bupt.edu.cn)