利用人口流动数据以及两阶段模型预测 2019 冠状病毒病流行趋势

叶元庆^{1,2,3}, 雷 浩^{1,2}, 陈 辰^{1,2}, 胡可嘉^{1,2}, 徐小林^{1,2}, 袁长征^{1,2}, 曹淑殷^{1,2},
王思思^{1,2}, 王思聪^{1,2}, 李 舒^{1,2}, 应智峻^{1,2}, 贾君麟^{1,2}, 王秦川^{2,4},
Sten H. VERMUND⁵, 许正平^{6,7}, 吴息凤^{1,2,3}

1. 浙江大学医学院附属第二医院生物统计、生物信息学和大数据中心, 浙江 杭州 310009
2. 浙江大学医学院公共卫生学院大数据健康科学系, 浙江 杭州 310058
3. 浙江大学健康医疗大数据国家研究院, 浙江 杭州 310058
4. 浙江大学医学院附属邵逸夫医院肿瘤外科, 浙江 杭州 310016
5. 美国耶鲁大学耶鲁公共卫生学院, 康涅狄格州 纽黑文市 06520
6. 浙江大学医学院环境医学研究所, 浙江 杭州 310058
7. 浙江大学感染性疾病诊治协同创新中心, 浙江 杭州 310003

[摘 要] 目的: 基于人口流动和疫情防控措施信息构建数学模型, 预测 2019 冠状病毒病(COVID-19)疫情的发展趋势。方法: 获取截至 2020 年 2 月 8 日浙江省乐清市 151 例确诊病例的患病过程以及乐清和武汉之间的人口流动等信息, 采用两阶段的方法构建数学模型, 将人口流动数据与确诊病例症状开始时间以及输入性病例和本地病例的传播动力学特征整合起来, 模拟并预测乐清的 COVID-19 疫情发展趋势。结果: 在疫情初期, 每日从武汉来的输入性病例数(症状出现日)与当日、6 d 前和 9 d 前从武汉到乐清的旅客数量成正相关。利用根据武汉到乐清的旅客数量估计的输入性病例数以及易感-暴露-感染-恢复模型预测乐清最终的病例数为 170 例。如果根据每日报告的输入性病例数来预测, 预计最终病例数为 165 例。截至 2020 年 4 月 27 日, 乐清实际监测到的病例数为 170 例, 这两个预测值与实际值均接近。结论: 本研究建立的两阶段预测模型能够对 COVID-19 的流行趋势做出准确预测。

[关键词] 2019 冠状病毒病; 中国; 输入病例; 本地病例; 易感-暴露-感染-恢复模型; 预测

[中图分类号] R181.2 **[文献标志码]** A

收稿日期: 2020-10-19 接受日期: 2020-12-25

基金项目: 浙江大学新型冠状病毒(2019-nCoV)肺炎应急科研专项(2020XGZX003); 浙江省创新团队(2019R01007); 浙江省重点实验室(2020E10004); 浙江省自然科学基金(LEZ20H260002)

第一作者: 叶元庆, 研究员, 博士生导师, 主要从事遗传流行病学研究; E-mail: yuanqing99@zju.edu.cn; <https://orcid.org/0000-0001-5708-8961>. 雷 浩, 讲师, 主要从事传染病流行病学研究; E-mail: leolei@zju.edu.cn; <https://orcid.org/0000-0002-1850-7040>

通信作者: 许正平, 教授, 博士生导师, 主要从事生物电磁学和分子医学研究; E-mail: zpxu@zju.edu.cn; <https://orcid.org/0000-0003-0922-736X>. 吴息凤, 教授, 博士生导师, 主要从事肿瘤分子流行病学和健康医疗大数据研究; E-mail: xifengw@zju.edu.cn; <https://orcid.org/0000-0003-2980-8129>

Predicting COVID-19 epidemiological trend by applying population mobility data in two-stage modeling

YE Yuanqing^{1,2,3}, LEI Hao^{1,2}, CHEN Chen^{1,2}, HU Kejia^{1,2}, XU Xiaolin^{1,2}, YUAN Changzheng^{1,2}, CAO Shuyin^{1,2}, WANG Sisi^{1,2}, WANG Sicong^{1,2}, LI Shu^{1,2}, YING Zhijun^{1,2}, JIA Junlin^{1,2}, WANG Qinchuan^{2,4}, Sten H. VERMUND⁵, XU Zhengping^{6,7}, WU Xifeng^{1,2,3} (1. Center for Biostatistics, Bioinformatics and Big Data, the Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou 310009, China; 2. Department of Big Data in Health Science, School of Public Health, Zhejiang University School of Medicine, Hangzhou 310058, China; 3. National Institute for Data Science in Health and Medicine, Zhejiang University, Hangzhou 310058, China; 4. Department of Surgical Oncology, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou 310016, China; 5. Yale School of Public Health, Yale University, New Haven 06520, Connecticut, USA; 6. Institute of Environmental Medicine, Zhejiang University School of Medicine, Hangzhou 310058, China; 7. Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Zhejiang University, Hangzhou 310003, China)

YE Yuanqing and LEI Hao contributed equally to this article.

Corresponding authors: XU Zhengping, E-mail: zpxu@zju.edu.cn, <https://orcid.org/0000-0003-0922-736X>; WU Xifeng, E-mail: xifengw@zju.edu.cn, <https://orcid.org/0000-0003-2980-8129>

[Abstract] Objective: To predict the epidemiological trend of coronavirus disease 2019 (COVID-19) by mathematical modeling based on the population mobility and the epidemic prevention and control measures. **Methods:** As of February 8, 2020, the information of 151 confirmed cases in Yueqing, Zhejiang province were obtained, including patients' infection process, population mobility between Yueqing and Wuhan, etc. To simulate and predict the development trend of COVID-19 in Yueqing, the study established two-stage mathematical models, integrating the population mobility data with the date of symptom appearance of confirmed cases and the transmission dynamics of imported and local cases. **Results:** It was found that in the early stage of the pandemic, the number of daily imported cases from Wuhan (using the date of symptom appearance) was positively associated with the number of population travelling from Wuhan to Yueqing on the same day and 6 and 9 days before that. The study predicted that the final outbreak size in Yueqing would be 170 according to the number of imported cases estimated by consulting the population number travelling from Wuhan to Yueqing and the susceptible-exposed-infectious-recovered (SEIR) model; while the number would be 165 if using the reported daily number of imported cases. These estimates were close to the 170, the actual monitoring number of cases in Yueqing as of April 27, 2020. **Conclusion:** The two-stage modeling approach used in this study can accurately predict COVID-19 epidemiological trend.

[Key words] Coronavirus disease 2019; China; Imported cases; Indigenous cases; Susceptible-exposed-infectious-recovered model; Prediction

[J Zhejiang Univ (Med Sci), 2021, 50(1): 68–73.]

[缩略语] 2019 冠状病毒病(coronavirus disease 2019, COVID-19); 易感-暴露-感染-恢复(susceptible-exposed-infectious-recovered, SEIR)

2020 年 3 月 11 日,世界卫生组织正式宣布 COVID-19 为“一种大范围的流行病”。人口流动在 COVID-19 空间传播中起着十分重要的作用^[1-3]。之前的研究表明,人口流动对某地输入病例的数量影响较大^[4],但输入病例如何影响本地病例的传播动力学特征,进而影响该地区整体的疫情发展趋势尚未深入研究。乐清市位于浙江省,是温州都市圈的一个城市。截至 2020 年 2 月 8 日,乐清大部分早期病例和超过 45% 的确诊病例都是从武汉返回的人员,初步说明从武汉到乐清市的人口流动对乐清疫情的发生和持续进展起到了至关重要的作用。本文提出了两阶段的模拟方法,并以乐清市 COVID-19 疫情传播为例,将武汉到乐清的人口流动数据与输入性病例(以症状发生日期计)、输入性病例与本地病例的动态信息整合在一起,构建数学模型模拟乐清本地病例的流行趋势,通过与每日实际确诊病例数进行对比来验证模型。本文旨在构建一个疫情预测模型的框架,为传染病疫情的精准预测和预防提供一个方法。

1 资料与方法

1.1 数据来源

COVID-19 病例的发病日期、暴露史、旅行史、诊断日期、乐清总人口数、从湖北省及其他高风险地区到乐清的旅客总人数、武汉到乐清的旅客总人数等数据均来自乐清市卫生健康局以及乐清市疾病预防控制中心。2020 年 1 月 1 日至 2 月 12 日,武汉至中国各城市(除台湾、香港和澳门)的每日人口流动数据由百度慧眼提供(<https://huiyan.baidu.com>)。鉴于短时间内百度慧眼的人群覆盖率相对固定,本研究认为可以将其作为一个常数,假设其覆盖率为 ε_1 。对于非输入性病例,本研究将首次报告症状(如咳嗽、发热等)的日期作为发病日期。对于输入性病例,发病日期定义为首次报告症状(咳嗽、发热等)的日期和到达乐清的日期中较晚者,其中有 1 例病例在诊断前没有出现任何症状,则将该病例的诊断日期作为发病日期。截至 2020 年 2 月 8 日,本研究详细收录了乐

清 151 例确诊病例的资料。

1.2 输入病例与输入人口流动定量关系建模

首先,将每日从武汉输入的病例数进行建模:

$$f(t) = \gamma_0 + \gamma t + \sum_{i=0}^{10} \alpha_i w_{t-i} + \sum_{i=0}^{10} \beta_i w_{t-i} \times t \quad (1)$$

其中 w_{t-i} 表示 t 时刻发病前第 i 天武汉发送至乐清的旅客量/100。由于人口流动数据只包含了武汉到温州市区(包含乐清)的旅客量,定义 t 时刻发病前第 i 天通过百度慧眼查询到的武汉发送温州的旅客量为 w'_{t-i} ,假设武汉到乐清日旅客量占武汉到温州的日旅客量的比例为 ε_2 ,可得:

$$w_{t-i} = \varepsilon_2 w'_{t-i} / \varepsilon_1 \quad (2)$$

将公式(2)带入公式(1)中,可根据百度慧眼查询到的武汉发送温州的旅客量 w'_{t-i} 来估计公式(1)中参数值 $\alpha_i \varepsilon_2 / \varepsilon_1$ 和 $\beta_i \varepsilon_2 / \varepsilon_1$ 。该模型分别考虑 w_{t-i} 以及其交互作用项 $w_{t-i} \times t$ 。交互项 $w_{t-i} \times t$ 可用于解释武汉市旅客量在不同时间段的不同贡献。以 0.05 为显著性水平,采用前向逐步选择法来拟合选择最优模型。通过观察实际值与拟合值的散点图和残差图,并进行 Breush-Pagan 线性异方差检验和 Ramsey RESET 检验来评估模型的效能。然后,利用基于选择的最优方程^[1]估计输入病例数的动态变化。

1.3 输入病例与本地病例定量关系建模

根据构建的输入病例与输入人口流动定量关系模型,基于输入人口数据预测输入病例数,本研究构建了第二阶段模型来定量分析输入病例数与本地病例数的关系,即 SEIR 模型。由于该模型所考虑的时间跨度相对较短(不到 3 个月),因此假设已经恢复的患者不会再成为模型中的易感人群,同时排除死亡患者。第二阶段的 SEIR 模型方程如下:

$$S(t+1) = S(t) - \frac{\beta(t)[I(t) + E(t) + I_m(t)]S(t)}{N} \quad (3)$$

$$E(t+1) = E(t) + \frac{\beta(t)[I(t) + E(t) + I_m(t)]S(t)}{N} - \frac{E(t)}{k} \quad (4)$$

$$I(t+1) = I(t) + \frac{E(t)}{k} - \frac{I(t)}{D+k} \quad (5)$$

$$R(t+1) = R(t) + \frac{I(t)}{D+k} \quad (6)$$

$S(t)$ 、 $E(t)$ 、 $I(t)$ 和 $R(t)$ 分别表示第 t 天的易感人数、暴露人数、感染人数和恢复人数; $\beta(t)$ 表示第 t 天的感染率并随着时间的变化而变化; N 表示乐清的总人口数, 研究期间虽然有人口的迁入和迁出, 但是相对于庞大的固定人口数而言, 迁入和迁出人口相对很小, 因此假定 N 在研究期间是固定值, 即 $N=1\,418\,500$; k 和 D 分别表示潜伏期和入院前的感染期。基于 Li 等^[5]的研究, 设 $k=5.2\text{ d}$, $D=2.3\text{ d}$ 。COVID-19 的有效再生数 $R(t)=\beta(t)(D+k)$ 。 $I_{in}(t)$ 表示第 t 天报道的输入病例数。

假定基本再生数 $R_0=2.2$ ^[4]。浙江省在 2020 年 1 月 23 日启动了重大突发公共卫生事件一级响应来防控 COVID-19 疫情, 所以在 2020 年 1 月 23 日以后, 假定基本再生数为 $0.01\,R_0$ 。

2 结果

2.1 人口流动对输入病例的影响

截至 2020 年 2 月 8 日, 乐清市共有确诊病例 151 例, 其中 71 例为输入病例(占确诊病例的 47%以上), 在这 71 例输入病例中有 69 例(97%)是从武汉输入的。图 1 显示了每日来自武汉及武汉以外城市的输入病例数和乐清市本地病例数。

在第一阶段, 通过前向逐步选择法, 每日由武汉输入病例数的最优推导是基于当日、6 d 前和 9 d 前从武汉到乐清的旅客数量, 其函数关系式为:

$$\begin{aligned} f(t) = & -0.13t - 0.41w_t + 0.10w_{t-6} \\ & + 0.13w_{t-9} + 0.01w_{t-6} \times t - 1.96w_{t-9} \\ & + 0.08w_{t-9} \times t + 4.33 \end{aligned} \quad (7)$$

最终模型的 $R^2=0.875$ (图 2), Breush-Pagan 检验和 Ramsey RESET 检验结果差异均不显著, 说明方差齐, 因此模型是合理的。

2.2 输入病例对本地病例的影响

在第二阶段, 本研究模拟了 COVID-19 本地感染病例的动态学特征, 利用公式(1)预测乐清的每日输入病例数。由于乐清实施了严格的防控措施, 本研究假设 2020 年 2 月 8 日以后每日输入病例数为 0, 如图 3A、B 所示, 该模型预测出乐清最终累计确诊病例数将达到 170 例。另外, 本研究再次利用 2020 年 1 月 8 日至 2 月 8 日每日报告的输

图 1 2020 年 1 月 8 日至 2 月 8 日乐清市每日 2019 冠状病毒病输入病例数及本地病例数

Figure 1 Number of daily imported and local COVID-19 cases in Yueqing from January 8 to February 8, 2020

图 2 模型预测 2020 年 1 月 8 日至 2 月 8 日乐清市每日 2019 冠状病毒病输入病例数

Figure 2 Model estimation of the number of daily imported COVID-19 cases in Yueqing from January 8 to February 8, 2020

入病例数来模拟乐清本地病例的传播动态, 结果如图 3C、D 所示, 模型预测出乐清最终累计确诊病例数将达到 165 例。截至 2020 年 4 月 27 日, 乐清 COVID-19 实际确诊病例数为 170 例, 并且自 2 月 16 日以来连续 71 d 无新增确诊病例, 表明该模型能够利用乐清 2 月 8 日以前的数据对其 2 月 8 日以后的疫情发展进行准确预测。

3 讨论

本文构建两阶段模型, 通过分析 COVID-19 确诊病例的发病日期以及疫情起源地的人口流动数据来预测 COVID-19 的流行趋势。在第一阶段, 利用武汉市每日到乐清的旅客数量模拟从武汉输入的病例数; 在第二阶段, 利用 SEIR 模型分别基于观察到的每日输入病例数和预测的每日输入病例数来预测未来的累计确诊病例数。此外, 在分析

A、B: 基于公式(7)预测的每日输入病例数预测; C、D: 基于每日实际输入病例数预测. 平滑后的本地病例指根据3 d平滑而得到的病例数.

图3 模拟和实际报道乐清每日新增和累计确诊2019冠状病毒病病例数

Figure 3 Predicted and reported number of daily new infections of COVID-19 and cumulative cases in Yueqing

过程中还考虑了乐清实施的严格防控措施。根据构建的两阶段方法和截至2020年2月8日的数据,预测乐清最终累计确诊病例数为170例,与截至2020年4月27日实际监测的病例数相同。因此,本研究提出的两阶段预测模型可用来预测某一地区COVID-19疫情发展趋势。

自COVID-19疫情暴发以来,研究报道了许多模型,部分模型也纳入了人口流动数据^[1-3,6-13],且几乎所有模型都使用了与确诊病例发病日期相关的信息。一份关于中国70 000多例确诊病例的报告显示,基于确诊日期的病例数和基于发病日期的确诊病例数之间随时间变化差异很大^[14],主要是因为病例的诊断可能会受到诊断测试工具的可及性或实验室确诊滞后性的影响,尤其是在COVID-19暴发的早期阶段^[3]。虽然有的模型在分析时确实考虑了症状出现的日期,但超过25%的受试者都缺少这一信息^[9]。此外,可能

由于缺少旅行史相关信息,之前所有模型均未将输入病例和本地病例分开进行分析。

本文资料中,每日输入的COVID-19病例数(以发病日期计)与症状出现前第6天和第9天从武汉来的旅客数量正相关,这表明部分患者到达乐清时已经出现症状。据报道,COVID-19的潜伏期可能为1~14 d^[15],每日输入的COVID-19病例数(以发病日期计)与症状出现前第6天和第9天旅客数量之间正相关或许能够解释病例潜伏期的变化。

综上所述,本研究构建了两阶段模型来预测COVID-19的流行趋势,并且以浙江省乐清市的COVID-19疫情为例,验证了两阶段模型的有效性。这表明根据人口流动数据、确诊病例发病时间以及病例旅行史等信息构建两阶段模型来预测COVID-19的流行趋势是可行的。虽然现阶段乐清的COVID-19疫情已经结束,但本文的目的在

于构建一个疫情预测模型的理论框架,希望为可能常态化的COVID-19疫情乃至未来可能发生的其他传染病的精准预测和预防提供方法指导。

参考文献

- [1] 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]. **Lancet**, 2020, 395(10225):689–697. DOI:10.1016/S0140-6736(20)30260-9.
- [2] READ J M, BRIDGEN J R, CUMMINGS D A, et al. Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions [J/OL]. **medRxiv**, 2020. DOI: 10.1101/2020.01.23.20018549.
- [3] AI S, ZHU G, TIAN F, et al. Population movement, city closure and spatial transmission of the 2019-nCoV infection in China[J/OL]. **medRxiv**, 2020. DOI: 10.1101/2020.02.04.20020339.
- [4] ZHANG C, CHEN C, SHEN W, et al. Impact of population movement on the spread of 2019-nCoV in China[J]. **Emerg Microbes Infect**, 2020,9(1):988–990. DOI:10.1080/22221751.2020.1760143.
- [5] LI Q, GUAN X, WU P, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia[J]. **N Engl J Med**, 2020,382(13):1199–1207. DOI:10.1056/NEJMoa2001316.
- [6] SONG X, LIU X, WANG C. The role of telemedicine during the COVID-19 epidemic in China—experience from Shandong province[J]. **Crit Care**, 2020,24(1):178. DOI:10.1186/s13054-020-02884-9.
- [7] WADHERA R K, WADHERA P, GABA P, et al. Variation in COVID-19 hospitalizations and deaths across New York city boroughs[J]. **JAMA**, 2020,323(21):2192. DOI:10.1001/jama.2020.7197.
- [8] CAO Z, ZHANG Q, LU X, et al. Incorporating human movement data to improve epidemiological estimates for 2019-nCoV[J/OL]. **medRxiv**, 2020. DOI: 10.1101/2020.02.07.20021071.
- [9] DU Z, WANG L, CAUCHEMEZ S, et al. Risk for transportation of 2019 novel coronavirus (COVID-19) from Wuhan to cities in China[J]. **medRxiv**, 2020. DOI:10.1101/2020.01.28.20019299.
- [10] ZHAO H, MAN S, WANG B, et al. Epidemic size of novel coronavirus-infected pneumonia in the epicenter Wuhan; using data of five-countries' evacuation action [J/OL]. **medRxiv**, 2020. DOI: 10.1101/2020.02.12.20022285.
- [11] SANCHE S, LIN Y T, XU C, et al. High contagiousness and rapid spread of severe acute respiratory syndrome coronavirus 2[J]. **Emerg Infect Dis**, 2020,26(7):1470–1477. DOI:10.3201/eid2607.200282.
- [12] CHINAZZI M, DAVIS J T, GIOANNINI C, et al. **Preliminary assessment of the international spreading risk associated with the 2019 novel coronavirus (2019-nCoV) outbreak in Wuhan city** [R]. Boston: Northeastern University, 2020.
- [13] IMAI N, DORIGATTI I, CORI A, et al. **Estimating the potential total number of novel Coronavirus cases in Wuhan City, China** [R]. London: Imperial College London, 2020.
- [14] 中国疾病预防控制中心新型冠状病毒肺炎应急响应机制流行病学组. 新型冠状病毒肺炎流行病学特征分析[J]. **中华流行病学杂志**, 2020,41(2):145–151. DOI:10.3760/cma.j.issn.0254-6450.2020.02.003.
Epidemiology Working Group for NCIP Epidemic Response. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China[J]. **Chinese Journal of Epidemiology**, 2020,41(2):145–151. DOI:10.3760/cma.j.issn.0254-6450.2020.02.003. (in Chinese)
- [15] GUAN W J, NI Z Y, HU Y, et al. Clinical characteristics of coronavirus disease 2019 in China [J]. **N Engl J Med**, 2020,382(18):1708–1720. DOI:10.1056/NEJMoa2002032.

[本文编辑 沈 敏 余 方]