

耦合不同年龄层接触模式的新冠肺炎传播模型

王国强 张烁 杨俊元 许小可

Study of coupling the age-structured contact patterns to the COVID-19 pandemic transmission

Wang Guo-Qiang Zhang Shuo Yang Jun-Yuan Xu Xiao-Ke

引用信息 Citation: *Acta Physica Sinica*, 70, 010201 (2021) DOI: 10.7498/aps.70.20201371

在线阅读 View online: <https://doi.org/10.7498/aps.70.20201371>

当期内容 View table of contents: <http://wulixb.iphy.ac.cn>

您可能感兴趣的其他文章

Articles you may be interested in

新型冠状病毒肺炎的流行病学参数与模型

Epidemiological parameters and models of coronavirus disease 2019

物理学报. 2020, 69(9): 090202 <https://doi.org/10.7498/aps.69.20200389>

基于连续感染模型的新冠肺炎校园传播与防控策略分析

Analysis of COVID-19 spreading and prevention strategy in schools based on continuous infection model

物理学报. 2020, 69(24): 240201 <https://doi.org/10.7498/aps.69.20201106>

新型冠状病毒肺炎疫情确诊病例的统计分析及其自回归建模

Statistical analysis and autoregressive modeling of confirmed coronavirus disease 2019 epidemic cases

物理学报. 2020, 69(9): 090203 <https://doi.org/10.7498/aps.69.20200503>

基于虚拟源原理的源边界参数蒙特卡罗反演技术

Source boundary parameter of Monte Carlo inversion technology based on virtual source principle

物理学报. 2019, 68(23): 232901 <https://doi.org/10.7498/aps.68.20191095>

新型冠状病毒肺炎早期时空传播特征分析

Analysis on early spatiotemporal transmission characteristics of COVID-19

物理学报. 2020, 69(8): 080701 <https://doi.org/10.7498/aps.69.20200285>

基于3.5 MeV射频四极质子加速器硼中子俘获治疗装置的束流整形体设计

Design of beam shaping assembly based on 3.5 MeV radio-frequency quadrupole proton accelerator for boron neutron capture therapy

物理学报. 2018, 67(14): 142801 <https://doi.org/10.7498/aps.67.20180380>

耦合不同年龄层接触模式的新冠肺炎传播模型*

王国强¹⁾²⁾ 张烁¹⁾²⁾ 杨俊元^{1)2)†} 许小可^{3)‡}

1) (山西大学复杂系统研究所, 太原 030006)

2) (山西大学, 疾病防控的数学技术与大数据分析山西省重点实验室, 太原 030006)

3) (大连民族大学信息与通信工程学院, 大连 116600)

(2020年8月20日收到; 2020年10月14日收到修改稿)

搜集广东省自1月23日到2月16日期间944例新冠肺炎样本信息. 对确诊人群进行年龄特征分析, 将人群分为儿童组(0—5岁)、青少年组(6—19岁)、中青年组(20—64岁)、老年组(65岁及以上), 耦合不同年龄层的接触模式, 建立离散年龄结构新冠肺炎模型, 得出模型的基本再生数及最终规模. 通过蒙特卡罗数值算法(MCMC)辨识模型的参数、拟合累计病例数、计算消亡时间、感染峰值及到达时间等有关生物量. 研究发现中青年人群感染人数最多; 相比于居家模式, 社区模式下中青年人群感染峰值上升41%, 峰值推迟一周到达. 通过分析不同年龄层的最终规模在对应年龄层的占比, 发现老年人的易感性较高, 青少年的易感性相对较低. 在居家模式下, 若各年龄层患者能及时就诊, 住院峰值将进一步减少, 但住院高峰将提前一周到达. 此模型可揭示个体接触行为对新冠肺炎的传播的影响, 定量评价居家隔离措施的有效性.

关键词: 新冠肺炎, 年龄结构接触模式, 蒙特卡罗模拟, 基本再生数**PACS:** 02.30.Hq, 02.60.Cb, 02.70.Uu**DOI:** 10.7498/aps.70.20201371

1 引言

新冠肺炎“coronavirus disease 2019”(COVID-19)肆虐全球10个多月以来, 截至目前已有7000多万人确诊, 170多万人死亡, 给人类的健康和生活带来极大的危害. 广东作为中国人口第一大省、经济强省, 人口密集且流动频繁, 感染风险颇大. 截至2020年7月16日, 广东省新冠肺炎累计确诊1650例, 新增病例均为输入性病例, 疫情基本得到控制. 这主要得益于广东各级政府及时启动新冠肺炎疫情一级响应^[1], 如封城、限制或者停止集会、居家隔离、佩戴口罩等一系列措施. 在此基础上, 了解和掌握新冠肺炎的流行特征和内在传播机制是有效遏制新冠肺炎再次爆发传播的关键.

分析广东省2020年1月9日到2月16日944例新冠肺炎患者的详细病例数据信息(年龄、性别、接触史、症状、症状出现时间、住院确诊时间等), 发现大多数患者以发热为主要临床症状, 所占比例为73.31%, 而少数患者伴有咳嗽、头晕、肌肉酸痛、乏力等症状; 男性患者占比49.79%, 女性患者占比50.21%. 通过对广东省每日新增病例分段统计(如图1所示), 发现0—5岁、6—19岁、20—64岁、65岁及以上各年龄组每日新增病例分别于1月31日、2月1日、1月27日及2月2日到达高峰; 由图1发现, 20—64岁每日新增病例高峰到达时刻早于其他组别, 且变化幅度明显高于其他组别.

为了更直观地观测每日新增确诊病例年龄构成, 图2给出了各年龄组每日新增病例占当日新增总病例百分比. 发现每日新增病例占当日新增总病例的百分比20—64岁超过80%, 65岁及以上约为

* 国家自然科学基金(批准号: 61573016, 61873154, 61773091)、山西省留学回国项目(批准号: 2015-094)、山西省1331工程项目和山西省新型冠状病毒(COVID-19)预防和控制项目(批准号: 202003D31011/GZ)资助的课题.

† 通信作者. E-mail: yjyang66@sxu.edu.cn

‡ 通信作者. E-mail: xuxiaoke@foxmail.com

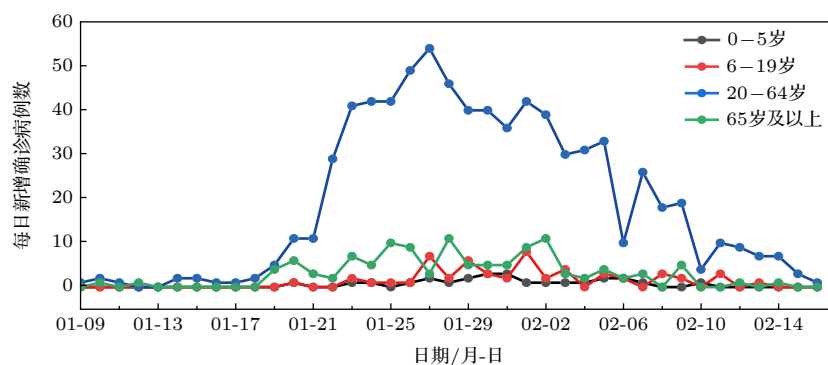


图1 广东省各年龄组每日新增病例折线图

Fig. 1. Line chart of new cases for each age group in Guangdong province.

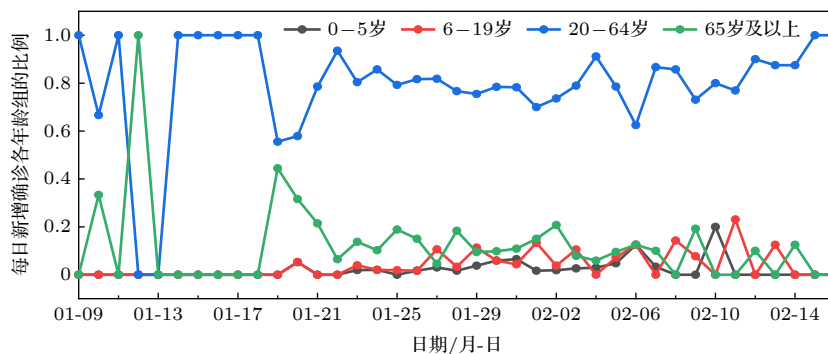


图2 广东省各年龄组每日新增病例占当日新增总病例百分比图

Fig. 2. Chart of the percentage of new cases for each age group in the total new cases of each day in Guangdong province.

15%, 0—19岁仅为5%. 广东省新冠肺炎患者分布呈现明显的年龄异质性. 如何利用数学模型揭示新冠肺炎传播的年龄异质性, 对新冠肺炎的优化控制至关重要.

现有新冠肺炎模型主要包括以数据驱动的统计模型^[2-5]及通过机理分析建立的动力学模型^[6-8], 统计建模以数据驱动为特点, 能直观描述疾病的发生和发展过程, 如曹文静等^[5]构建了安徽省疫情发展后期的自回归传播模型, 为新冠肺炎疫情的传播早期预测预警提供参考. 动力学模型主要发掘疾病传播的内在传播规律, 评估和预测疾病流行特征. 李盈科等^[6]详细阐述了几类新冠肺炎动力学模型及其在早期疾病防控方面的作用, 模型假设人群的接触是均匀混合的, 忽略了接触行为的异质性. 了解和掌握不同年龄段个体接触模式可以更精准地刻画传染病传播的传播途径, 是传染病建模的重要组成部分^[9-11]. 广东省采取了严格的管控遏制新冠肺炎的传播, 该措施改变人的行为模式及活动范围. 本文考虑以个体家庭为中心, 将人群的活动距离分为家庭内部(居家模式)和家庭外0.5 km内活动范围(社区模式), 并在两种模式下对不同年龄组之间的接触进行量化, 得到不同模式下的接触矩

阵, 结合传播动力学模型, 刻画新冠肺炎在不同人群间的传播. 考虑各年龄层接触模式的异质性对提高新冠肺炎传播的认识及控制具有重要意义, 并提供理论依据.

近期, 已有一些学者耦合社会接触模式和动力学建模方法来研究年龄异质性对新冠肺炎的影响^[12,13]. 如 Kiesha 等^[12]建立了一个离散年龄结构的 SEIR 模型评估不同时间复工对新冠肺炎的影响. 文献^[13]评估不同年龄段的错峰复工感染新冠肺炎的风险. 已有的结果主要讨论不同年龄间接触模式对新冠肺炎的风险评估, 对数据的拟合及实际情况的预测略有欠缺. 本文拟合广东省新冠肺炎各年龄层高维数据、辨识模型参数, 揭示社会接触行为对新冠肺炎传播的影响, 评估各种策略对新冠肺炎传播的影响.

2 模型假设和构建

2.1 模型假设

1) 2020年1月23日至2月19日时间跨度较短; 1月23日距除夕仅一天, 春运基本结束, 而且

各级政府已于1月23日实行封城,人口流动因素可以忽略.因此可忽略人口出生和自然死亡,不考虑人口迁移,假设广东省人口为固定常数.

2)自1月23日起,全国各地开始实行严格的居家式隔离,人与人之间主要以家庭单位接触,故假设广东省人群接触方式为家庭式接触.

3)假设广东省的医疗资源充足,被确诊的染病者能得到及时救治.

4)假设确诊者入院后被完全隔离,不再具有传染性.

2.2 模型构建

依据广东省新冠肺炎传播特征,将广东省人口分成0—5岁、6—19岁、20—64岁以及65岁及以上4个不同层次,刻画广东省新冠肺炎在儿童、青少年、中青年、老年人群中的流行特征.用 N_1 表示0—5岁的总人口, N_2 表示6—19岁的总人口, N_3 表示20—64岁总人口, N_4 表示65岁及以上老年人的数量.依据SEIHR仓室建模理论,将第 $k(k=1,2,3,4)$ 年龄段总人口 $N_k(t)$ 分成五种不同的人群,即易感者、潜伏者、染病者、确诊住院者和康复者,分别用 $S_k(t)$, $E_k(t)$, $I_k(t)$, $H_k(t)$ 和 $R_k(t)$ 表示.第 k 年龄阶段易感者的个体和 j 年龄阶段染病个体平均接触数为 c_{kj} ,第 k 年龄段和 j 年龄段染病者接触数为 $c_{kj} \frac{I_j}{N_j}$,总接触数为 $\sum_{j=1}^4 c_{kj} \frac{I_j}{N_j}$.第 k 年龄段的易感个体以传播概率 β_k 被感染而转变成潜伏个体 E_k .潜伏个体经过 $1/\alpha$ 天潜伏变成染病个体.染病个体经过 $1/h$ 天进入医院确诊.确诊住院个体经过 $1/\gamma$ 天转化成康复个体.上述新冠肺炎状态间的转化可用如下微分方程组表述:

$$\begin{cases} \frac{dS_k(t)}{dt} = -\beta_k S_k(t) \sum_{j=1}^4 c_{kj} \frac{I_j(t)}{N_j(t)}, \\ \frac{dE_k(t)}{dt} = \beta_k S_k(t) \sum_{j=1}^4 c_{kj} \frac{I_j(t)}{N_j(t)} - \alpha E_k(t), \\ \frac{dI_k(t)}{dt} = \alpha E_k(t) - h I_k(t), \\ \frac{dH_k(t)}{dt} = h I_k(t) - \gamma H_k(t), \\ \frac{dR_k(t)}{dt} = \gamma H_k(t). \end{cases} \quad (1)$$

假设(1)式表示的模型1具有如下初值:

$$\begin{aligned} S_k(0) = N_k, E_k(0) \geq 0, I_k(0) \geq 0, H_k(0) \geq 0, \\ R_k(0) \geq 0, k = 1, 2, 3, 4. \end{aligned} \quad (2)$$

模型各参数的生物学意义及取值参见表1.第 k 个年龄组初始时刻的易感者数量、潜伏者数量、染病者数量、确诊者数量的具体取值见表2.

表1 参数定义和参数值

Table 1. Parameter definitions and parameter values.

参数	定义	数值	置信区间	数据来源
β_1	0—5岁年龄组感染率	0.0164	[0.0163, 0.0165]	MCMC
β_2	6—19岁年龄组感染率	0.0266	[0.0265, 0.0267]	MCMC
β_3	20—64岁年龄组感染率	0.0869	[0.0868, 0.0870]	MCMC
β_4	65岁及以上年龄组感染率	0.0358	[0.0357, 0.0359]	MCMC
$1/\alpha$	潜伏期	5.2	—	文献[14]
$1/h$	确诊周期	2.0383	[2.0379, 2.0387]	MCMC
$1/\gamma$	康复期	10	—	文献[15]

表2 变量生物学意义和变量初始值

Table 2. The meaning of variables and initial values.

参数	定义	数值	置信区间	数据来源
$S_1(0)$	初始时刻0—5岁人群易感者数量	7153800	—	表4
$S_2(0)$	初始时刻6—19岁人群易感者数量	16207482	—	表4
$S_3(0)$	初始时刻20—64岁人群易感者数量	78760477	—	表4
$S_4(0)$	初始时刻65岁及以上人群易感者数量	10664042	—	表4
$E_1(0)$	初始时刻0—5岁人群潜伏者数量	16	[15, 17]	MCMC
$E_2(0)$	初始时刻6—19岁人群潜伏者数量	31	[30, 32]	MCMC
$E_3(0)$	初始时刻20—64岁人群潜伏者数量	317	[316, 318]	MCMC
$E_4(0)$	初始时刻65岁及以上人群潜伏者数量	64	[63, 65]	MCMC
$I_1(0)$	初始时刻0—5岁人群染病者数量	2	[1, 3]	MCMC
$I_2(0)$	初始时刻6—19岁人群染病者数量	2	[1, 3]	MCMC
$I_3(0)$	初始时刻20—64岁人群染病者数量	68	[67, 69]	MCMC
$I_4(0)$	初始时刻65岁及以上人群染病者数量	17	[16, 18]	MCMC
$H_1(0)$	初始时刻0—5岁人群确诊住院数量	2	—	文献[16]
$H_2(0)$	初始时刻6—19岁人群确诊住院数量	3	—	文献[16]
$H_3(0)$	初始时刻20—64岁人群确诊住院数量	109	—	文献[16]
$H_4(0)$	初始时刻65岁及以上人群确诊住院数量	24	—	文献[16]

2.3 基本再生数

基本再生数是传染病防控中一个非常重要的量. 基本再生数 R_0 表示在一个全部都是易感者的环境中一个染病者在整个染病期内二次传染人数的平均值. 在疾病传播初期, 如果 $R_0 < 1$, 疾病消亡; 否则, 疾病流行. 模型 1 有惟一的无病平衡点

$$P_0 = (S_1^0, 0, 0, 0, 0, \dots, S_4^0, 0, 0, 0, 0) \\ = (N_1, 0, 0, 0, 0, \dots, N_4, 0, 0, 0, 0),$$

这里 $N_k (k=1, 2, 3, 4)$ 分别表示广东省 0—5 岁、6—19 岁、20—64 岁及 65 岁及以上人口数.

利用文献 [17,18] 计算模型 1 基本再生数. 在无病平衡点 P_0 处线性化模型 1, 定义

$$F = \begin{pmatrix} 0 & F_1 \\ 0 & 0 \end{pmatrix} \\ = \begin{pmatrix} \beta_1 c_{11} & \beta_1 \frac{c_{12}N_1}{N_2} & \beta_1 \frac{c_{13}N_1}{N_3} & \beta_1 \frac{c_{14}N_1}{N_4} \\ \beta_2 \frac{c_{21}N_2}{N_1} & \beta_2 c_{22} & \beta_2 \frac{c_{23}N_2}{N_3} & \beta_2 \frac{c_{24}N_2}{N_4} \\ \beta_3 \frac{c_{31}N_3}{N_1} & \beta_3 \frac{c_{32}N_3}{N_2} & \beta_3 c_{33} & \beta_3 \frac{c_{34}N_3}{N_4} \\ \beta_4 \frac{c_{41}N_4}{N_1} & \beta_4 \frac{c_{42}N_4}{N_2} & \beta_4 \frac{c_{43}N_4}{N_3} & \beta_4 c_{44} \end{pmatrix}, \quad (3)$$

和

$$V = \begin{pmatrix} \alpha I_d & 0 \\ -\alpha I_d & h I_d \end{pmatrix}, \quad (4)$$

其中 I_d 为 4×4 单位矩阵, 0 为 4×4 零矩阵,

$$F_1 = \begin{pmatrix} \beta_1 c_{11} & \beta_1 \frac{c_{12}N_1}{N_2} & \beta_1 \frac{c_{13}N_1}{N_3} & \beta_1 \frac{c_{14}N_1}{N_4} \\ \beta_2 \frac{c_{21}N_2}{N_1} & \beta_2 c_{22} & \beta_2 \frac{c_{23}N_2}{N_3} & \beta_2 \frac{c_{24}N_2}{N_4} \\ \beta_3 \frac{c_{31}N_3}{N_1} & \beta_3 \frac{c_{32}N_3}{N_2} & \beta_3 c_{33} & \beta_3 \frac{c_{34}N_3}{N_4} \\ \beta_4 \frac{c_{41}N_4}{N_1} & \beta_4 \frac{c_{42}N_4}{N_2} & \beta_4 \frac{c_{43}N_4}{N_3} & \beta_4 c_{44} \end{pmatrix}, \quad (5)$$

则

$$R_0 = \rho(FV^{-1}) = \frac{1}{h} \rho(F_1), \quad (6)$$

其中 ρ 表示矩阵的谱半径.

容易验证, 当初值条件满足 (2) 式时模型 1 的

解是非负的. 注意到 $N_k(t) = S_k(t) + E_k(t) + I_k(t) + H_k(t) + R_k(t)$, 那么 $\frac{dN_k(t)}{dt} = 0$, 故各年龄层人口数 N_k 保持不变. 从而区域

$$\Gamma = \{(S_k, E_k, I_k, H_k, R_k) \in R_+^5 (k=1, 2, 3, 4) \mid \\ 0 < S_k \leq N_k, 0 \leq E_k, I_k, H_k, R_k \leq N_k\}$$

是模型 (1) 的一个正向不变集.

定理 1 假设矩阵 F_1 不可约且 $R_0 < 1$, 则无病平衡点 P_0 全局渐近稳定.

证明 由矩阵 F_1 的不可约性, 结合 Perron-Frobenius^[19] 定理, 矩阵 F_1 存在一个单的特征根 $\rho(F_1) = hR_0$ 和一个正的左特征向量 $\omega = (\omega_1, \omega_2, \omega_3, \omega_4)$, 满足 $\omega F_1 = h\rho(F_1)\omega = hR_0\omega$.

令 $E = (E_1, E_2, E_3, E_4)^T$ 和 $I = (I_1, I_2, I_3, I_4)^T$, 构建如下 Lyapunov 函数:

$$L[E, I](t) = \sum_{k=1}^4 \omega_k (E_k(t) + I_k(t)), \quad (7)$$

其中 $\omega_k (k=1, 2, 3, 4)$ 是 $h\rho(F_1)$ 对应特征向量的分量.

对 L 沿模型 1 解轨线求全导数得

$$\frac{dL[E, I](t)}{dt} \Big|_1 = \sum_{k=1}^4 \omega_k \left(\frac{E_k(t)}{dt} + \frac{I_k(t)}{dt} \right) \\ = \sum_{k=1}^4 \omega_k \left[\sum_{j=1}^4 \beta_k c_{kj} S_k \frac{I_j}{N_j} - h I_k \right] \\ \leq \sum_{k=1}^4 h \omega_k \left[\sum_{j=1}^4 \frac{\beta_k c_{kj} S_k^0}{h N_j} I_j - I_k \right] \\ = h \omega \left(\frac{1}{h} F_1 - I_d \right) I = h(R_0 - 1) \omega I. \quad (8)$$

令

$$M = \{(S_1, E_1, I_1, H_1, R_1, \dots, S_4, E_4, I_4, H_4, R_4) \\ \in \Gamma \mid L' = 0\}.$$

如果 $R_0 < 1$, 由 ω 正性得到 $L' = 0$ 的充要条件为 $I = 0$. 从而最大不变集 M 只包含惟一的无病平衡点 P_0 . 由 LaSalle 不变集原理^[20] 知, 当 $R_0 < 1$ 时, 无病平衡点 P_0 是全局渐近稳定的.

定理 1 表明只要基本再生数小于 1, 就不会有新增新冠肺炎患者, 疾病得以控制.

最终规模是指在整个疾病流行期间内感染的总人数. 在传染病传播中, 用来刻画疾病传播的危害程度. 实际上由模型 1 可得, 第 k 年龄层的累计

病例数

$$h \int_0^{\infty} I_k(t) dt$$

即为各年龄组的最终规模.

首先, 由模型 1 的第一个方程得出 $S_k(t)$ 单调递减, 故 $S_k(\infty)$ 存在. 将模型 1 第一个方程分离变量积分得

$$\ln \left(\frac{S_k(0)}{S_k(\infty)} \right) = \beta_k \sum_{j=1}^4 \frac{c_{kj}}{N_j} \int_0^{\infty} I_j(t) dt. \quad (9)$$

因此, 反常积分 $\int_0^{\infty} I_k(t) dt$ 存在. 由 $I_k(t)$ 的非负性得 $I_k(\infty) = 0$. 同理, 由第三个方程得 $E_k(\infty) = 0$.

其次, 将模型 1 的前三个方程相加并积分得

$$\begin{aligned} h \int_0^{\infty} I_k(t) dt \\ = S_k(0) + E_k(0) + I_k(0) - S_k(\infty) - E_k(\infty) - I_k(\infty) \\ = S_k(0) + E_k(0) + I_k(0) - S_k(\infty). \end{aligned} \quad (10)$$

最后, 将 (10) 式代入到 (9) 式, 得到模型 1 各年龄层最终规模为

$$\begin{aligned} \ln \left(\frac{S_k(0)}{S_k(\infty)} \right) = \frac{\beta_k}{h} \sum_{j=1}^4 \frac{c_{kj}}{N_j} (S_j(0) + E_j(0) + I_j(0) \\ - S_j(\infty)), \quad k = 1, 2, 3, 4, \end{aligned} \quad (11)$$

其中 $S_k(\infty)$ 表示第 k 年龄层易感者的最终规模. 事实上, 新冠肺炎的易感初始人群 $S_k(0)$ 已知; 若通过数据辨识模型参数 β_k 和 h 、接触矩阵 c_{kj} 及初值 $E_j(0)$ 和 $I_j(0)$, 根据 (11) 式可求出 $S_k(\infty)$, 进而得到第 k 年龄层的染病最终规模为 $S_k(0) + E_k(0) + I_k(0) - S_k(\infty)$.

2.4 人口数据及部分参数值的确定

通过查阅中国 2010 年人口普查资料^[21], 得到 2010 年广东省各年龄层统计数据 (见表 3).

随后依据广东省 2010 年到 2019 年各年度人口增长和出生率^[22], 估算出现有年龄段人口分布 (见表 4).

由文献^[23]可知新冠肺炎的平均潜伏期为 5.2 天, 故可以假设 $\alpha = 1/5.2$; 根据国家卫健委新闻发布会^[15], 湖北省外患者从确诊住院到恢复出院时长约为 10 天, 故可假设 $\gamma = 1/10$.

从模型 1 不难发现, 计算接触矩阵 $c_{ij}(i, j = 1, 2, 3, 4)$ 非常关键^[7,11]. 文献^[14]对广东省城乡人

口的出行和接触方式进行了调查, 考虑个体活动范围 (家庭内部和外部接触), 得到了不同活动范围下的接触矩阵——家庭内部接触矩阵、以家庭为中心半径 0.5 km 内的接触矩阵及半径超过 0.5 km 的接触矩阵. 本文将家庭内部接触矩阵定义为居家模式接触矩阵, 将活动半径 0.5 km 以内和居家模式接触矩阵的和定义为社区模式接触矩阵. 本次新冠肺炎的传播主要考虑儿童、青少年、中青年及老年人的流行病学特征和文献^[14]各年龄层的划分相契合.

广东省于 1 月 23 日出台一系列严格的控制措施, 如封城、居家隔离、停工、停课等, 各年龄段的人群接触近似于居家模式. 将文献^[14]中家庭模式下的接触矩阵 (图 3(a)) 作为疫情期间采取居家隔离的接触矩阵. 引入社区模式接触矩阵 (图 3(b)) 评估居家隔离措施对新冠肺炎的影响. 接触矩阵中的元素 c_{kj} 表示 k 年龄组的个体与 j 年龄组人群的接触数.

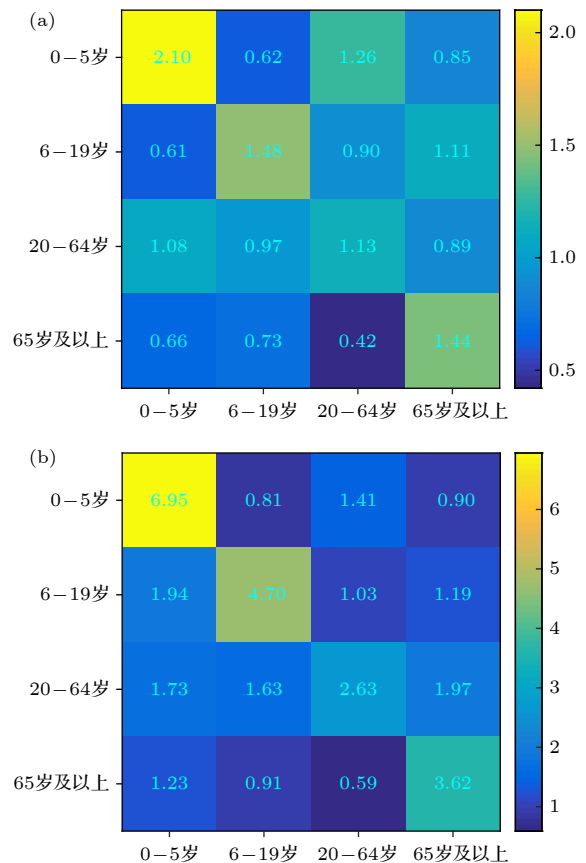


图 3 (a) 广东省居家模式接触矩阵; (b) 广东省社区模式接触矩阵

Fig. 3. (a) Household-mode contact matrix of Guangdong; (b) community-mode contact matrix of Guangdong.

表 3 2010 年广东省各年龄段人口 (单位: 1000 人)

Table 3. The number of population for each age group of Guangdong province in 2010 (Unit: 1000).

年龄	0	1—4岁	5—9岁	10—14岁	15—19岁	20—24岁	25—29岁	30—34岁	35—39岁	40—44岁
人口	1081	4543	5167	6812	9990	12276	10621	8935	9533	8864
年龄	45—49岁	50—54岁	55—59岁	60—64岁	65—69岁	70—74岁	75—79岁	80—84岁	85岁以上	
人口	7047	4899	4399	3066	2168	1911	1494	903	611	

表 4 2019 年广东省各年龄段人口分布 (单位: 人)

Table 4. The number of population for aggregated age groups in Guangdong province in 2019 (unit: person).

年龄	N_1 (0—5岁)	N_2 (5—19岁)	N_3 (20—64岁)	N_4 (65岁以上)
人口数	7153800	16207482	78760477	10664042

3 模型拟合结果

模型 1 中第 k 年龄组每日新增确诊病例为 $hI_k(t)$, 从而第 T 天的累计确诊病例为

$$z_k(T) = h \int_0^T I_k(t) dt = \sum_{t=0}^T hI_k(t),$$

$$T = 1, 2, \dots, 25. \quad (12)$$

结合马尔科夫链蒙特卡罗方法 (MCMC)^[24] 拟合累计病例数据, 同时采取 Metropolis-Hastings(M-H)

算法校准参数值和初值 (见表 1 与表 2). 此算法随机模拟 20000 次, 从第 8000 次模拟截断, 计算后 12000 次模拟平均值.

基于各年龄组每日累计确诊病例, 分别拟合 0—5 岁 (儿童组), 6—19 岁 (青少年组), 20—64 岁 (中青年组), 65 岁及以上 (老年组) 四个年龄组的累计病例随时间的变化图 (如图 4 和图 5 所示). 不难发现拟合的整体效果比较理想. 但 0—5 岁年龄组的拟合偏差稍大, 主要是由于 0—5 岁病例样本量较少, 拟合噪声影响较强, 导致实际数据和模型拟合结果有一定偏差.

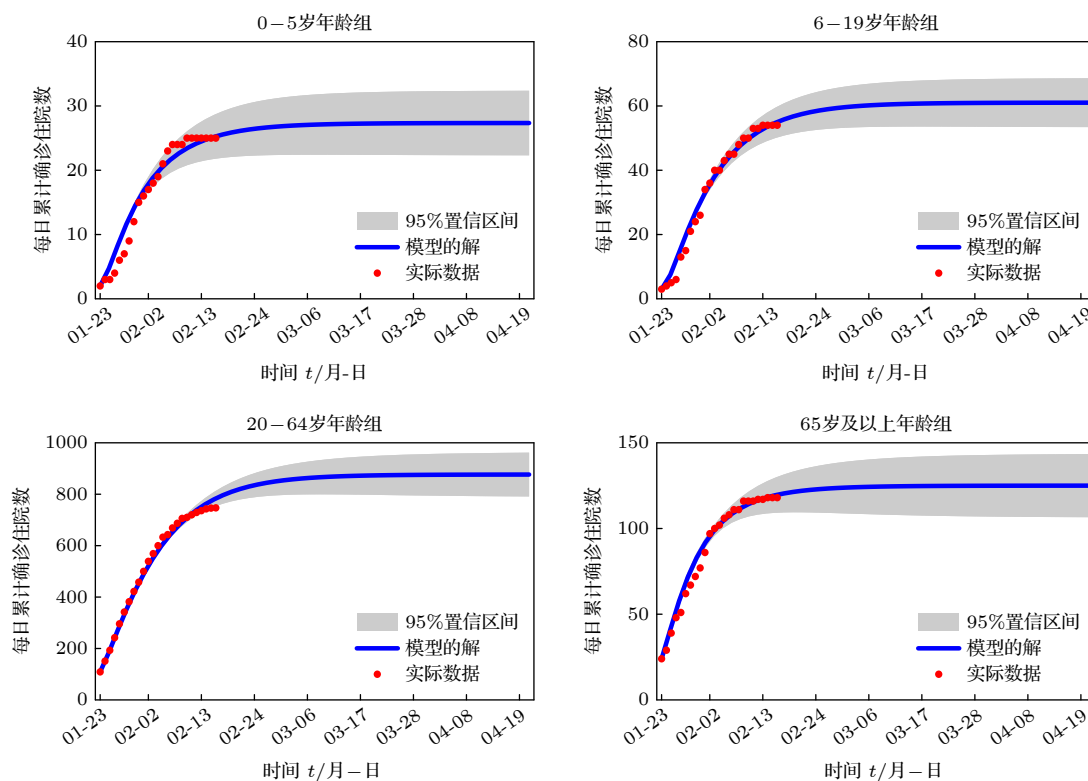


图 4 各年龄组每日累计确诊住院数. 红点表示实际数据, 蓝线表示模型预测均值, 灰色阴影代表 95% 的置信区间

Fig. 4. The cumulative number of confirmed cases per day in each age group. The red dots represent the actual data, the blue line represents the solution of model 1, and the shaded gray area represents the 95% confidence interval.

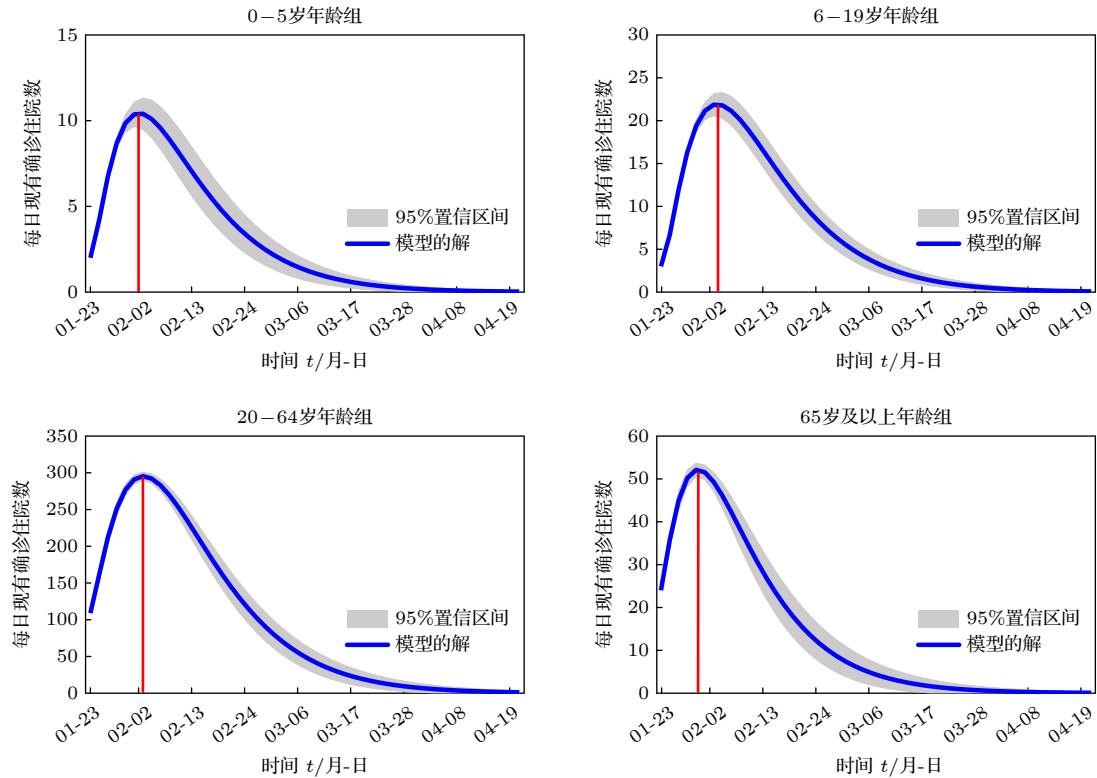


图 5 各年龄组每日现有确诊住院数 $H_k(t)$. 其中蓝线表示模型的解, 灰色阴影部分表示 95% 置信区间, 红线表示峰值

Fig. 5. The number of confirmed hospitalized cases per day in each age group. The blue line represents the solution of the model, the shaded gray area represents the 95% confidence interval, and the red line represents the peak.

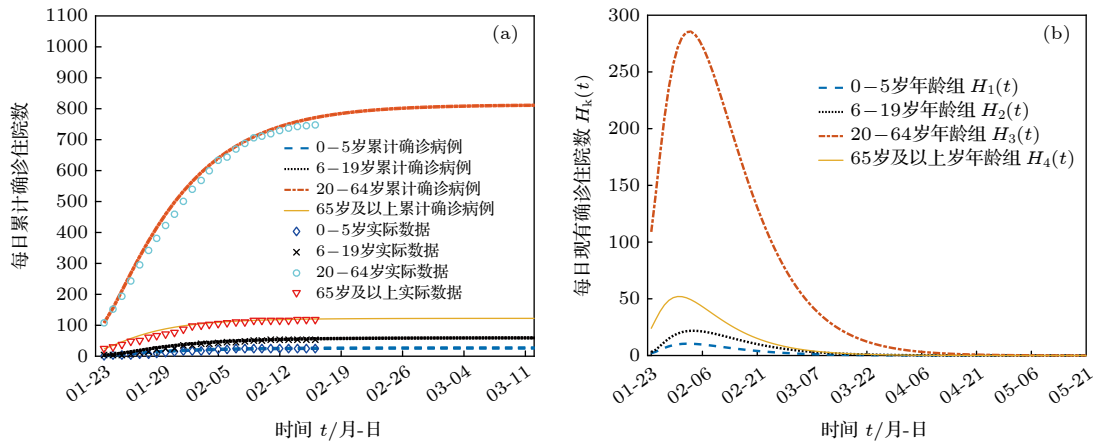


图 6 (a) 各年龄组每日累计确诊住院数 (b) 各年龄组每日现有确诊住院数 $H_k(t)$.

Fig. 6. (a) Total number of confirmed hospitalizations per day in each age group. (b) Each age group has the number of confirmed hospitalizations $H_k(t)$ per day.

根据拟合结果, 通过比较各年龄组的累计确诊住院病例曲线 (图 6(a)) 和每日现有确诊住院曲线 (图 6(b)), 发现儿童及青少年感染人数明显少于老年组和中青年组, 呈现明显的年龄异质性; 图 6(b) 表明老年组每日现有确诊峰值最先到达、儿童组紧随其后、中青年组最后到达. 具体各个年龄组的住院病例峰值到达时间、住院峰值、最终规模、消亡时间见表 5.

表 5 表明每日现有确诊住院数高峰老年组 (65 岁及以上) 于 1 月 31 日到达, 儿童 (0—5 岁) 于 2 月 2 日到达, 青少年组 (6—19 岁) 和中青年组 (20—64 岁) 于 2 月 3 日到达. 由图 7(a) 可以看出 20—64 岁年龄组每日现有确诊住院峰值占比最多, 65 岁及以上年龄组次之, 0—5 岁年龄组最少; 各年龄层最终规模 (图 7(a)) 和高峰占比构成相似; 住院病例多为中青年和老年人. 该结论为针对不同

表 5 拟合结果分析
Table 5. Analysis of fitting results.

名称	峰值到达时间	住院峰值/人	最终规模/人	疾病消亡时间
0—5岁年龄组	2月2日 [2月1日, 2月3日]	11 [10, 12]	28 [23, 33]	3月10日 [3月3日, 3月17日]
6—19岁年龄组	2月3日 [2月1日, 2月5日]	21 [20, 23]	61 [54, 69]	3月22日 [3月15日, 3月29日]
20—64岁年龄组	2月3日 [2月2日, 2月4日]	296 [290, 301]	877 [791, 962]	4月22日 [3月31日, 4月30日]
65岁及以上年龄组	1月31日 [1月29日, 2月1日]	52 [50, 54]	125 [107, 144]	3月23日 [3月9日, 4月5日]

的年龄人群预备和筹集相应的医疗资源提供了关键信息. 通过比较表 1 中传播概率 $\beta_1, \beta_2, \beta_3, \beta_4$ 以及各年龄组最终规模, 得出青少年感染新冠肺炎风险较低, 20—64 岁中青年染病风险较高. 结合表 4 和表 5 的数据计算各年龄层最终规模占对应年龄层人数占比, 结果如图 7(b) 所示. 不难发现, 65 岁及以上年龄患者占比最高, 而 6—19 岁年龄组患者占比最低. 因此, 老年人群易感染新冠肺炎, 该年龄组人群应该加强日常新冠肺炎预防.

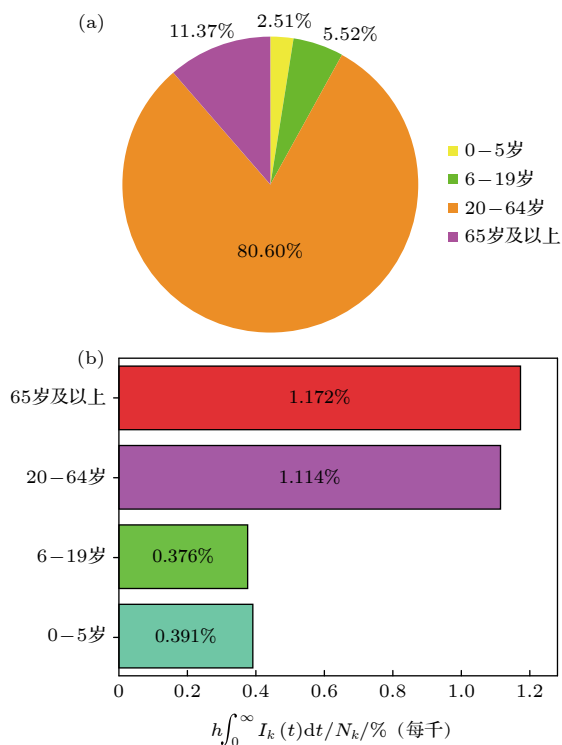


图 7 (a) 各年龄组最终规模占比图; (b) 各年龄层最终规模占对应年龄层人数占比图

Fig. 7. (a) The proportions of final size in each age group; (b) The proportion of the final size of each age group in the associated age group.

4 控制策略评估

为了评估居家隔离效果, 改变接触模式, 允许个体以家庭为中心, 在半径为 0.5 km 的范围内活

动, 该接触模式定义为图 3(b) 所示的社区模式. 比较两种接触模式下各年龄组感染新冠肺炎的变化.

图 8 表明在社区模式下, 各年龄组的每日现有确诊病例数峰值及到达时间有显著差异, 其中, 0—5 岁年龄组每日现有确诊病例数住院高峰时间将推迟 2 天, 峰值增加 9%; 6—19 岁年龄组高峰到达时间推迟 3 天, 峰值增加 29%; 20—64 岁年龄组高峰时间推迟近 1 周, 峰值增加 41%; 65 岁及以上年龄组峰值到达时间推迟 1 天, 峰值增加 8% (表 6).

表 6 社区模式住院峰值到达时间及峰值
Table 6. Daily peak arrival time and peak values of hospitalized cases in the community mode.

名称	0—5岁年龄组	6—19岁年龄组	20—64岁年龄组	65岁及以上年龄组
峰值到达时间	2月4日	2月6日	2月8日	2月1日
住院峰值	12	27	417	56

定理 1 表明基本再生数小于 1 能有效控制新冠肺炎的传播. 为此, 采用拉丁超立方抽样方法分析模型参数对基本再生数的影响, 结果如图 9(a) 所示. 可以看出, 减小传染概率 $\beta_k (k = 1, 2, 3, 4)$ 和提高确诊率能有效降低新冠肺炎的传播. 在居家模式下, 提高确诊率比其他控制策略更有效.

在居家隔离政策下, 对比 $1/h$ 天后确诊对疫情高峰的影响, 分别取参数值 $h = 0.5, h = 1.0, h = 2.0$, 即确诊期分别为 2 天, 1 天和 0.5 天, 其他参数值保持不变, 结果如图 9(b) 所示. 可以看出, 如果感染者出现症状后立刻去医院检测, 特别是在出现症状后半天内确诊, 会明显降低住院高峰人数, 同时也能缩短疫情消亡时间. 说明及时就诊能有效遏制新冠肺炎的蔓延.

通过以上控制措施评估效果可以发现, 实施居家隔离措施、及时隔离确诊病例并对其进行治疗是疫情防控的关键. 以上所有分析都基于医疗资源充足, 若考虑感染者在出现症状后不能及时得到确诊, 该模型还需要进一步修改和完善.

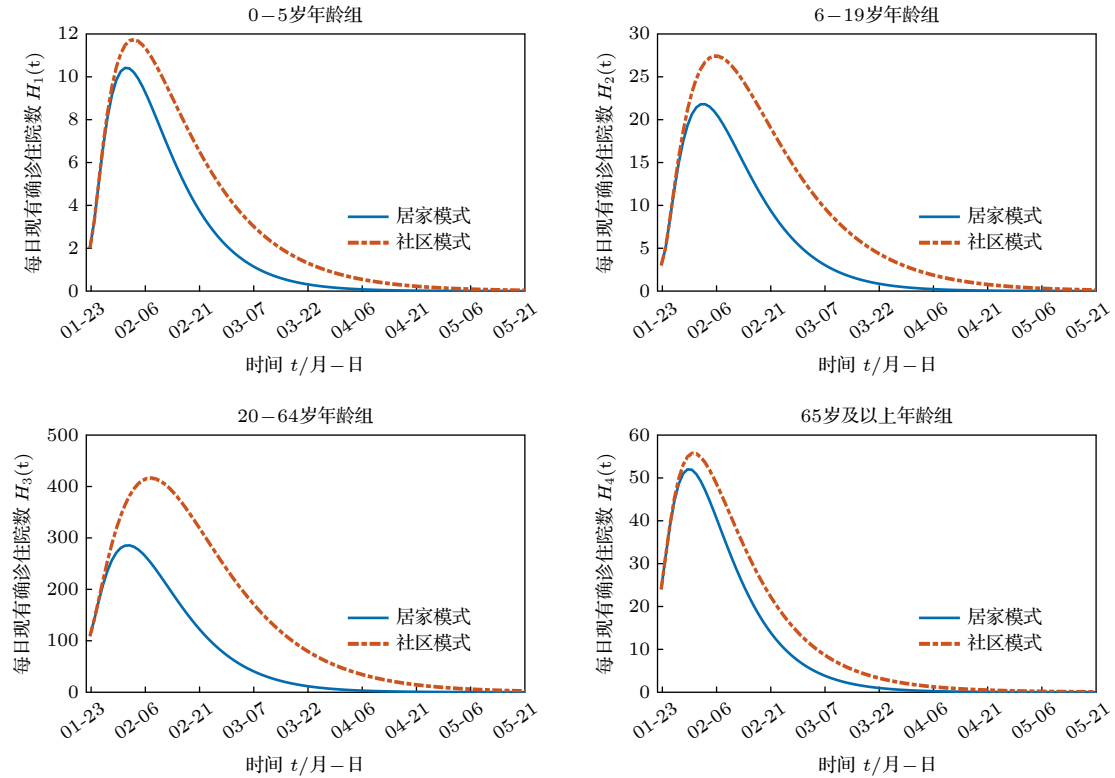


图 8 各年龄组每日确诊病例时间序列图. 蓝实线表示居家模式下的每日确诊病例时间序列图, 红虚线表示社区模式下每日确诊病例时间序列图

Fig. 8. The daily hospitalized cases for each age group over time t . The blue line represents the daily hospitalized cases in the household mode, the red dotted line represents the daily hospitalized cases in the community mode.

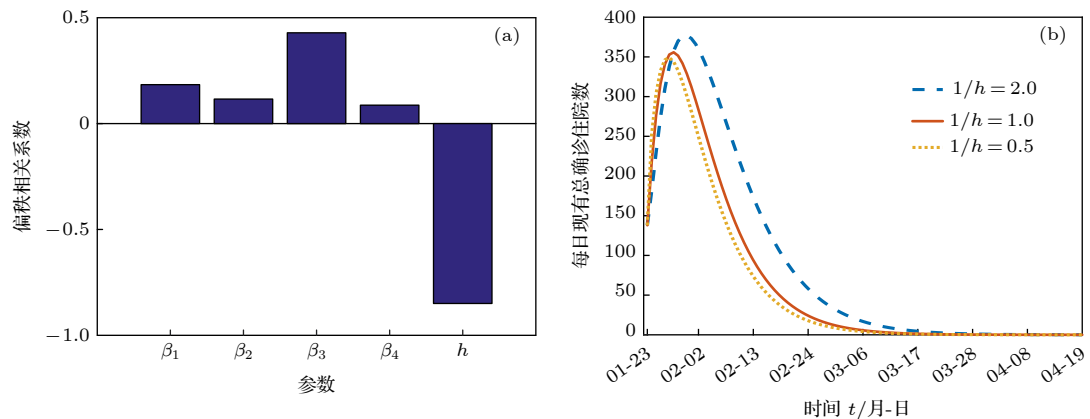


图 9 (a) 参数对基本再生数 R_0 的敏感性分析; (b) 每日现有总确诊住院数 $\sum_{k=1}^4 H_k(t)$ 取不同确诊率时间序列图

Fig. 9. (a) The sensitivity analysis of model parameters to R_0 ; (b) the time series diagram of the total hospitalized cases with parameter h .

5 讨 论

本文分析了广东省 944 例新冠肺炎患者信息, 发现广东省新冠肺炎患者各年龄患者的构成比例有显著差异. 通过建立耦合不同年龄层接触模式的 SEIHR 模型, 基于现实病例数据, 利用 MCMC

方法, 辨识模型参数, 分析不同年龄段人群的新冠肺炎的传播趋势, 计算出基本再生数 $R_0 = 1.2323$ (95% CI [1.2316, 1.2329]) 和最终规模分别为 0—5 岁年龄组 28 例 (95% CI [23, 33]), 6—19 岁年龄组 61 例 (95% CI [54, 69]), 20—64 岁年龄组 877 例 (95% CI [791, 962]), 65 岁及以上年龄组 125 例 (95% CI [107, 144]). 研究发现居家模式对疫情的

控制非常有效, 如果政策适当放松, 即社区模式下染病人数上升明显, 其中 20—64 岁 (中青年组) 染病峰值比居家模式增加 41%. 通过分析不同年龄层感染最终规模在对应年龄层人口数的占比, 发现老年人感染新冠肺炎风险较高, 青少年的感染风险较低.

耦合不同年龄层人群接触模式的传染病动力学模型, 本质上刻画了不同社会接触模式 (物理距离) 对新冠肺炎传播的影响, 揭示不同年龄段人群的新冠肺炎的流行特征和规律. 同时能准确捕捉不同控制策略下, 人的社会行为对新冠肺炎的影响. 结果表明在严格管控措施下, 提高确诊和隔离率是遏制新冠肺炎传播的关键. 耦合年龄结构的传染病模型为合理调派医疗资源和节约成本提供了理论依据.

图 6 较准确地拟合了各年龄组的确诊病累计例数, 表明模型 1 能客观反映广东省新冠肺炎在各年龄组间的实际传播情况. 与传统的均匀混合的 SEIHR 仓室模型相比, 模型 1 能较准确了解新冠肺炎与年龄及物理接触距离的关系, 能捕捉各个年龄组人群之间的传播动态, 能较清晰地评估居家隔离对各年龄组新冠肺炎传播的影响. 通过数值模拟发现, 经典的均匀混合模型会高估新冠肺炎在广东省的传播, 可能对及时复工复产发出误导信息 (见图 10).

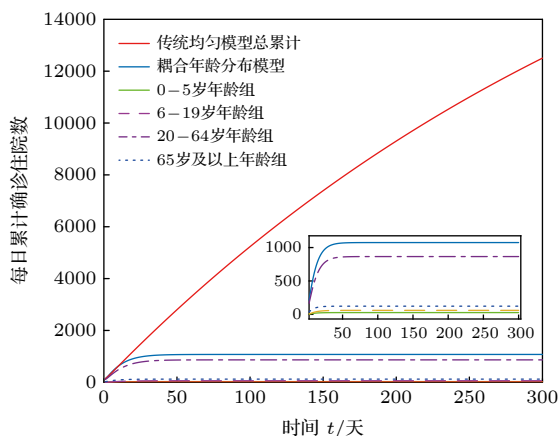


图 10 均匀混合模型与模型 (1) 比较图

Fig. 10. The comparison diagram of uniformly mixed model and model (1).

本文虽然详细地讨论了年龄异质性及物理接触距离对新冠肺炎的影响, 但模型的建立仍有许多不足. 首先接触矩阵的数据由文献 [14] 直接获得,

虽然我们采用和文献 [25,26] 类似的聚合方法更新接触数据, 但获得的信息与实际情况有一定的偏差. 受限于获得信息的不完整, 目前只能依赖文献 [14] 提供的数据将年龄进行分层. 将来如果采用更加精确的遥感设备, 精确记录每天各年龄组的接触数据, 进而能设计更加合理的分组方式来研究新冠肺炎的传播. 其次, 由于缺乏实际详细信息, 模型 1 假设参数 (除感染率外) 都和年龄无关, 实际上由于各年龄层个体体质不同, 不同年龄层的康复期略有不同, 那么康复率应该是一个与年龄有关的函数. 我们将对这些遗留问题进行进一步改进和完善.

参考文献

- [1] Guangdong Provincial Health Commission. http://www.gd.gov.cn/gdywdt/gdyw/content/post_2878982.html/ [2020-01-24]
- [2] Ankarali H, Ankarali S, Caskurlu H 2020 *Asia-Pac. J. Public Health* **32** 157
- [3] Jin Q X 2020 *Statistics & Decision* **36** 11 (in Chinese) [金启轩 2020 统计与决策 **36** 11]
- [4] Zhang L 2020 *J. Univ. Electron. Sci. Technol. China* **49** 345 (in Chinese) [张琳 2020 电子科技大学学报 **49** 345]
- [5] Cao W J, Liu X F, Han Z, Feng X, Zhang L, Liu X F, Xu X K, Wu Y 2020 *Acta Phys. Sin.* **69** 090203 (in Chinese) [曹文静, 刘小菲, 韩卓, 冯鑫, 张琳, 刘肖凡, 许小可, 吴晔 2020 物理学报 **69** 090203]
- [6] Li Y K, Zhao S, Lou Y J, Gao D Z, Yang L, He D H 2020 *Acta Phys. Sin.* **69** 090202 (in Chinese) [李盈科, 赵时, 楼一均, 高道舟, 杨琳, 何岱海 2020 物理学报 **69** 090202]
- [7] Yang J Y, Wang G Q, Zhang S 2020 *Math. Biol. Eng.* **17** 4500
- [8] Tang B, Wang X, Qian L, Nicola L B, Tang S Y, Xiao Y N, Wu J H 2020 *J. Clin. Med.* **9** 462
- [9] Bai Y 2017 *Ph. D. Dissertation* (Changchun: Jilin University) (in Chinese) [白媛 2017 博士学位论文 (长春: 吉林大学)]
- [10] Prem K, Alex R, Mark J 2017 *PLOS Comput. Biol.* **13** e1005697
- [11] Li X Z, Yang J Y, Maia M 2020 *Age Structured Epidemic Modelling* (Switzerland: Springer) p153
- [12] Kiesha P, Liu Y, Timothy W R, Adam J K, Rosalind M E, Nicholas D 2020 *Lancet Public Health* **5** e261
- [13] Zhao H, Feng Z L 2020 *Math. Biosci.* **326** 108405
- [14] Read J M, Lessler J, Riley S, Wang S, et al. 2014 *Proc. R. Soc. B* **281** 20140268
- [15] The Economic Observer <https://baijiahao.baidu.com/s?id=1657595901837318521&wfrspider&forpc/> [2020-02-04]
- [16] Health Commission of Guangdong Province. http://wsjkw.gd.gov.cn/zwyw_yqxx/content/post_2911721.html/ [2020-03-01]
- [17] Martcheva M 2015 *An Introduction to Mathematical Epidemiology* (New York: Springer) p104
- [18] Van Den Driessche P, Watmough J 2002 *Math. Biosci.* **180** 29
- [19] Horn R A, Johnson C R 1994 *Matrix Analysis* (Cambridge: Cambridge University Press) p534
- [20] Lasalle J P 1976 *The Stability of Dynamical Systems* (Philadelphia: SIAM) p30
- [21] Census Office of the State Council, Division of Population

- and Employment Statistics 2012 *Data from China's 2010 population census-on* (Beijing: China Statistical Press) p265 (in Chinese) [国务院人口普查办公室, 国家统计局人口和就业统计司2012中国2010年人口普查资料-上 (北京: 中国统计出版社) 第265页]
- [22] Guangdong Statistical Yearbook 2019 http://stats.gd.gov.cn/gdtjnj/content/post_2639622.html/ [2019-09-29]
- [23] Li Q, Guan X H, Wu P, Wang X Y, Zhou L, Tong Y Q, Ren R Q, Kathy S M, Leung, Eric H Y L, Jessica Y W, Xing X S, Xiang N J 2020 *N. Engl. J. Med.* **382** 1199
- [24] Xiao Y N, Zhou Y C, Tang S Y 2012 *Principle of Biomathematics* (Xi'an: Xi'an Jiaotong University Press) p213 (in Chinese) [肖燕妮, 周义仓, 唐三一 2012 生物数学原理 (西安: 西安交通大学出版社) 第213页]
- [25] Moghadas S M, Shoukat A, Fitzpatrick M C, et al. 2020 *Proc. Natl. Acad. Sci. USA* **117** 9122
- [26] Zhang J, Litvinova M, Liang Y, et al. 2020 *Science* **368** 1481

Study of coupling the age-structured contact patterns to the COVID-19 pandemic transmission*

Wang Guo-Qiang¹⁾²⁾ Zhang Shuo¹⁾²⁾ Yang Jun-Yuan^{1)2)†} Xu Xiao-Ke^{3)‡}

1) (Complex Systems Research Center, Shanxi University, Taiyuan 030006, China)

2) (Shanxi Key Laboratory of Mathematical Techniques and Big Data Analysis on Disease Control and Prevention, Shanxi University, Taiyuan 030006, China)

3) (College of Information & Communication Engineering, Dalian Minzu University, Dalian 116600, China)

(Received 20 August 2020; revised manuscript received 14 October 2020)

Abstract

Background: The coronavirus disease 2019 (COVID-19) has raged more than 10 months and it has become a major public health concern. It is necessary to account for the intrinsic mechanisms and reveal the transmission pattern. **Method:** We collect detailed information of 944 COVID-19 cases in Guangdong province from January 23rd to February 16th. According to the age-structured characteristics, the population is divided into four groups such as child group (0–5 years old), adolescent group (6–19 years old), young and middle-aged group (20–64 years old), elderly group (65 and over years old). Coupling with different age-structured contact patterns, we establish a discrete age-structured COVID-19 model, obtain the basic reproduction number and final size. By Markov Chain Monte Carlo numerical method (MCMC), we identify the model parameters, fit the cumulative cases, calculate eradication time of disease, infection peak and the peak arrival time, etc. **Results:** We found that the most infected people are the young and middle-aged individuals; Compared with household quarantine measure, the peak value of hospitalizations among young and middle-aged group in community mode will increase of 41%, and the peak will delay two weeks. By analyzing the proportions of the final sizes associated age groups, it is found that the elderly have a higher susceptibility, while the adolescents have a lower susceptibility. Under the household quarantine measure, if infected individuals have been confirmed in time of half a day, the peak size of hospitalizations will be further reduced, and the peak hospitalization will advance one week. The model reveals social contact patterns for impacting on COVID-19 transmission, and evaluates the effectiveness of household quarantine.

Keywords: coronavirus disease 2019, age structured contact pattern, Markov Chain Monte Carlo simulation, the basic reproduction number

PACS: 02.30.Hq, 02.60.Cb, 02.70.Uu

DOI: 10.7498/aps.70.20201371

* Project supported by the National Natural Science Foundation of China (Grant Nos. 61573016, 61873154, 61773091), the Fund for Returned Chinese Scholars of Shanxi Province, China (Grant No. 2015-094), the 1331 Engineering Project of Shanxi Province, China, and the New Coronavirus (COVID-19) Prevention and Control Project of Shanxi Province, China (Grant No. 202003D31011/GZ).

† Corresponding author. E-mail: yjyang66@sxu.edu.cn

‡ Corresponding author. E-mail: xuxiaoke@foxmail.com