

·新型冠状病毒肺炎疫情防控·

基于马尔科夫链蒙特卡罗模拟方法的COVID-19年龄别病死率估计

杜志成¹ 郝元涛¹ 魏永越² 张志杰³ 沈思鹏² 赵杨² 唐金陵⁴ 陈峰² 姜庆五³
李立明⁵

¹中山大学公共卫生学院,全球卫生研究中心,热带病防治研究教育部重点实验室,广州510080; ²南京医科大学公共卫生学院,全球健康中心 211166; ³复旦大学公共卫生学院,上海 200032; ⁴广州市妇女儿童医疗中心 510623; ⁵北京大学公共卫生学院 100191

通信作者:郝元涛, Email:haoyt@mail.sysu.edu.cn

【摘要】目的 新型冠状病毒肺炎疫情已席卷全球,疫情结束前,其病死率的估计受现有确诊病例和发病到死亡时间分布的影响,且结论尚不明确,本研究旨在对新型冠状病毒肺炎的年龄别病死率进行估计。**方法** 收集国家卫生健康委员会和CDC发布的新型冠状病毒肺炎疫情数据信息,采用Gamma分布拟合发病到死亡时间分布规律,采用马尔科夫链蒙特卡罗模拟估计年龄别病死率。**结果** 新型冠状病毒肺炎的发病到死亡时间 $M=13.77$ ($P_{25} \sim P_{75}$: 9.03 ~ 21.02) d, 总病死率为 4.1% (95%CI: 3.7% ~ 4.4%), 0 ~、10 ~、20 ~、30 ~、40 ~、50 ~、60 ~、70 ~ 和 ≥ 80 岁组病死率分别为 0.1%、0.4%、0.4%、0.4%、0.8%、2.3%、6.4%、14.0% 和 25.8%。**结论** 校正删失的马尔科夫链蒙特卡罗模拟方法适用于新发突发传染病疫情期间的病死率估计,尽早明确新型冠状病毒肺炎的病死率有助于疫情防控。

【关键词】 病死率; 马尔科夫链蒙特卡罗模拟; 新型冠状病毒肺炎

基金项目:国家自然科学基金(81773543, 81973150); 中山大学2020年度“三大”建设大科研项目培育专项

DOI:10.3760/cma.j.cn112338-20200609-00823

Using Markov Chain Monte Carlo methods to estimate the age-specific case fatality rate of COVID-19
Du Zhicheng¹, Hao Yuantao¹, Wei Yongyue², Zhang Zhijie³, Shen Sipeng², Zhao Yang², Tang Jinling⁴, Chen Feng², Jiang Qingwu³, Li Liming⁵

¹School of Public Health, Global Health Institute, Key Laboratory of Tropical Disease Control for the Ministry of Education, Sun Yat-sen University, Guangzhou 510080, China; ²School of Public Health, Center for Global Health, Nanjing Medical University, Nanjing 211166, China; ³School of Public Health, Fudan University, Shanghai 200032, China; ⁴Guangzhou Women and Children Medical Center, Guangzhou 510623, China; ⁵School of Public Health, Peking University, Beijing 100191, China
Corresponding author: Hao Yuantao, Email: haoyt@mail.sysu.edu.cn

【Abstract】Objectives The COVID-19 epidemic has swept all over the world. Estimates of its case fatality rate were influenced by the existing confirmed cases and the time distribution of onset to death, and the conclusions were still unclear. This study was aimed to estimate the age-specific case fatality rate of COVID-19. **Methods** Data on COVID-19 epidemic were collected from the National Health Commission and China CDC. The Gamma distribution was used to fit the time from onset to death. The Markov Chain Monte Carlo simulation was used to estimate age-specific case fatality rate. **Results** The median time from onset to death of COVID-19 was $M=13.77$ ($P_{25} \sim P_{75}$: 9.03~21.02) d. The overall case fatality rate of COVID-19 was 4.1% (95%CI: 3.7%~4.4%) and the age-specific case fatality rate were 0.1%, 0.4%, 0.4%, 0.4%, 0.8%, 2.3%, 6.4%, 14.0 and 25.8% for 0-, 10-, 20-, 30-, 40-, 50-, 60-, 70- and ≥ 80 years group, respectively. **Conclusions** The Markov Chain Monte Carlo simulation method adjusting censored is suitable for case fatality rate estimation during the epidemic of a new infectious disease. Early identification of the COVID-19 case fatality rate is helpful to the prevention and control of the epidemic.

【Key words】 Case fatality rate; Markov Chain Monte Carlo simulation; COVID-19

Fund programs: National Natural Science Foundation of China (81773543, 81973150); Sun Yat-sen University 2020 “Three Major” Construction Major Scientific Research Projects Cultivation Project

DOI:10.3760/cma.j.cn112338-20200609-00823

新型冠状病毒肺炎(COVID-19)疫情已席卷全球,截至2020年4月10日,全球已有超过185个国家暴发疫情,报道160余万病例,其中死亡病例9.6万^[1]。病死率(case fatality rate)是流行病学特征最重要的指标之一,特别在有效治疗方法和药物问世之前,病死率是疾病严重性的重要指标。针对一种新发突发传染病,人们对其病死率的认识往往受病程的长短、现有病例数和现有重症病例数等影响。WHO发布的COVID-19全球研究路线中关于严重性的知识空缺包括:“轻症发展到重症的临床表现是什么?不同年龄和性别的严重性有何不同?危重症高风险人群有哪些?”等^[2]。因此,在疫情结束前,尽早获得较为准确的COVID-19年龄别病死率信息,将为卫生行政部门作出疫情防控部署决策起到重要的参考意义。

截至3月24日,我国COVID-19本土疫情已基本阻断,31个省份累计报告确诊病例81 218例,现有确诊病例尚有4 287例,其中重症病例1 399例^[3]。现有确诊病例,特别是重症病例,可能随着病情的发展成为死亡病例,这种由于发病到死亡的时间分布规律导致死亡结局延迟出现的现象,在流行病学观察性研究中被称作删失(censored),而删失对病死率的估计存在影响,且疫情处在越早期则影响越大,易出现明显低估情况。因此,本研究旨在校正删失的情况下准确估计COVID-19各年龄组的病死率。

资料与方法

1. 数据来源:收集我国31个省份:①2020年1月20日至3月24日每日报告病例数和报告死亡病例数,相关数据从国家卫生健康委员会(卫健委)官方网站获取;②48例死亡病例信息,从各级卫健委官方网站获取,纳入标准为报告具有明确发病日期和死亡日期的病例,收集的信息包括省份、性别、年龄、发病日期、死亡日期和来源链接等,如国家卫健委1月23日公布的“曾XX,男,61岁,2019年12月20日左右开始发热,……,(2020年1月9日)23时13分,……,宣布临床死亡”^[4];③2019年12月2日至2020年2月1日的每日发病病例数,从WHO-中国联合调查组COVID-19报告中提取^[5-6];④我国2月11日年龄别累计死亡病例数和病例数比例,从中国CDC新型冠状病毒肺炎应急响应机制流行病学组的报告中获取^[7]。

2. 统计学分析:

(1)发病到死亡时间分布:发病到死亡的时间分

布将作为年龄别病死率马尔科夫链蒙特卡罗(MCMC)模拟估计中的先验信息。本研究采用Gamma分布对48例具有明确发病日期和死亡日期的死亡病例进行分布拟合,并计算COVID-19发病到死亡天数的百分位数($P_{2.5}$ 、 P_{25} 、 P_{50} 、 P_{75} 、 $P_{97.5}$)。Gamma分布估计方法见文献[8]。本部分内容主要利用R 3.6.3软件coarseDataTools软件包完成。

(2)每日发病病例数估计:每日发病病例数将作为年龄别病死率MCMC模拟估计中病死率的分母。本研究采用改进的SEIR模型对12月2日至2月1日的每日发病病例数和3月24日的累计病例数进行拟合(决定系数 $R^2=0.997$),并对2月2日至3月24日的每日发病病例数进行预测。微分方程如下:

$$\begin{aligned}\frac{dS}{dt} &= -\frac{\beta SE}{N} - \frac{\theta \beta SI_2}{N} - \delta SI_1 + \varepsilon S_q \\ \frac{dS_q}{dt} &= \delta SI_1 - \varepsilon S_q \\ \frac{dE}{dt} &= \frac{\beta SE}{N} + \frac{\theta \beta SI_2}{N} - \alpha E \\ \frac{dI_1}{dt} &= \rho \frac{\beta SE}{N} - \gamma I_1 \\ \frac{dI_2}{dt} &= (1 - \rho) \frac{\beta SE}{N} - \gamma I_2 \\ \frac{dR}{dt} &= \gamma I_1 + \gamma I_2\end{aligned}$$

式中, S 为易感者, S_q 为隔离的易感者, E 为潜伏感染者, I_1 为收治患者, I_2 为无症状和轻症等未被收治患者, R 为治愈和病死等病例, N 为总人群。各人群初始值设置: $N=1.4 \times 10^9$, $S=N$, $S_q=0$, $E=0$, $I_1=1$, $I_2=0$, $R=0$ 。各参数拟合结果: $\beta=0.414$, $\delta=4.4 \times 10^{-6}$, $\varepsilon=5.0 \times 10^{-6}$, $\theta=0.75$, $\gamma=1/14$, $\rho=0.95$ 。本部分内容利用Berkely Madonna 9.1.19软件完成。

(3)年龄别病死率估计:本研究基于贝叶斯框架采用MCMC方法对年龄别病死率进行估计。设各年龄组为 $A=\{a=1, 2, \dots, 9; 1=0 \sim 岁, 2=10 \sim 岁, \dots, 9=\geq 80 岁\}$,则对应的病死率为 θ_a ;发病到死亡时间的均数±标准差为 $m_{od} \pm s_{od}$,则对应的分布为 $f_{od}(\cdot | m_{od}, s_{od})$;研究期间的日期为 $T=\{0, 1, \dots, t\}$,则12月2日为研究初始日期 t_0 ,2月11日为 t_{11} ;各年龄组每日发病病例数的比例为 P_a ,以全国2月11日年龄别累计病例数比例统一赋值;每日发病病例数为 C ,则各年龄组每日发病病例数为 $C_a=C \times P_a$,而日期 t 年龄组 a 每日发病病例数 $C_{t,a}$ 。那么,日期 t 年龄组 a 的病例在日期 T 的死亡概率 λ 和期望死亡病例数 $E(D_a)$ 为:

$$\lambda(a, t | \theta_a, m_{od}, s_{od}) = \theta_a \int_t^{T-t} f_{od}(\tau | m_{od}, s_{od}) d_\tau$$

$$E(D_a) = \sum_{t=t_0}^T C_{t,a} \lambda(a, t)$$

采用极大似然函数法对 θ_a 进行求解,似然函数 L 包括3个部分 $L=L_1+L_2+L_3$,分别拟合1月20日至3月24日的每日累计死亡病例数与对应期望死亡病例数(L_1),2月11日年龄别累计死亡病例数(L_2),3月24日的累计病死率(L_3):

$$L_1 = \text{Poisson}(D, E(D))$$

$$L_2 = \text{Poisson}(D_{t_{71,a}}, E(D_{t_{71,a}}))$$

$$L_3 = \text{Binomial}(D_T/C_T, E(D_T/C_T))$$

本部分内容主要利用R 3.6.3中的drjacoby软件包完成。

结 果

1. 发病到死亡时间分布:基于卫健委公布数据的动态粗病死率见图1。本研究共收集48例死亡病例信息,年龄(70.1 ± 11.6)岁,其中男性31例。发病到死亡时间的中位数 $M=13.77$ ($P_{25} \sim P_{75}$: 9.03 ~ 21.02) d,其Gamma分布的形状参数3.215,尺度参数4.879(图2)。

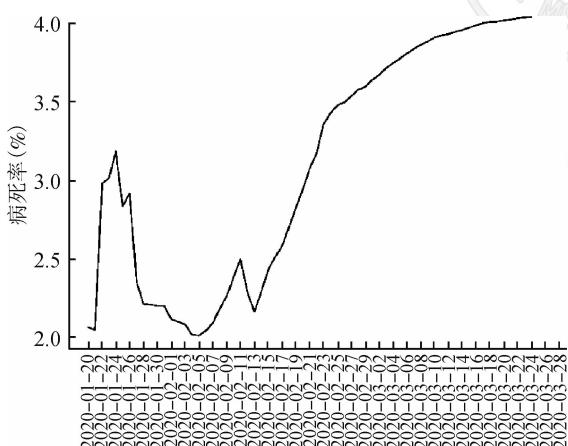


图1 基于卫健委公布数据的动态粗病死率

2. 年龄别病死率模型拟合情况:模型估计期望死亡病例数与1月20日至3月24日的每日累计死亡病例数的拟合情况见图3,对称平均绝对百分比误差(SMAPE)=3.26%,决定系数(R^2)=0.999。与2月11日年龄别累计死亡病例数的拟合情况见图4,SMAPE=30.85%, R^2 =1.000。模型估计的累计病死率为4.05%(3 294/8 1236),与3月24日的累计病死率4.04%(3 281/81 218)的比较结果为 $\chi^2=0.018$, $P=0.892$ 。

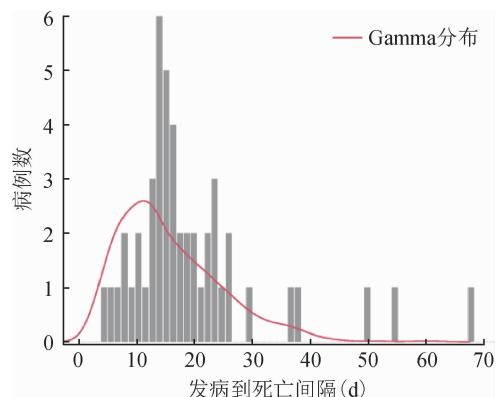


图2 发病到死亡时间及其概率密度分布

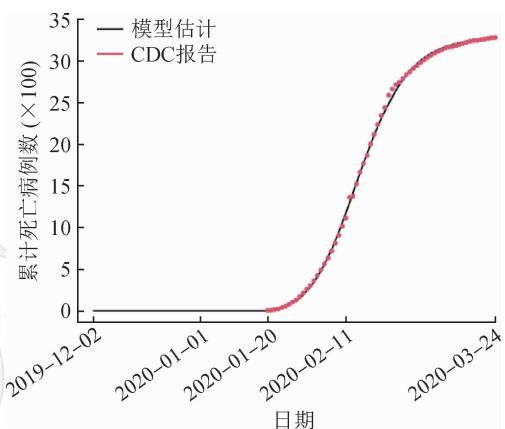


图3 模型预测及卫健委公布每日累计死亡病例数

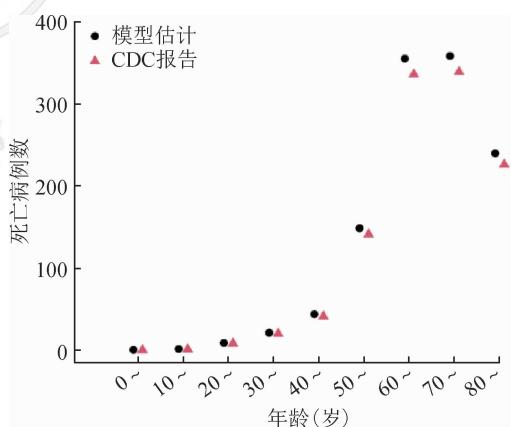


图4 模型预测及CDC公布的年龄别死亡病例数
(截至2月11日)

3. 年龄别病死率估计结果:模型估计的年龄别病死率结果见表1,总病死率为4.1% (95% CI: 3.7% ~ 4.4%),0~、10~、20~、30~、40~、50~、60~、70~和≥80岁组病死率分别为0.1%、0.4%、0.4%、0.4%、0.8%、2.3%、6.4%、14.0%和25.8%。

讨 论

年龄别病死率是重要的流行病学特征,可为疫

表1 CDC报告与模型估计的年龄别病死率

年龄组 (岁)	CDC报告 ^a		模型估计 ^b	
	死亡数/病例数	病死率 (%)	死亡数/病例数	病死率 (%, 95%CI)
全部	1 023/44 672	2.3	3 294/81 236	4.1(3.7~4.4)
0~	0/416	-	1/731	0.1(0.0~0.7)
10~	1/549	0.2	4/975	0.4(0.1~1.1)
20~	7/3 619	0.2	25/6 580	0.4(0.2~0.6)
30~	18/7 600	0.2	59/13 810	0.4(0.3~0.6)
40~	38/8 571	0.4	123/15 597	0.8(0.6~1.0)
50~	130/10 008	1.3	415/18 197	2.3(2.1~2.5)
60~	309/8 583	3.6	994/15 597	6.4(6.0~6.8)
70~	312/3 918	8.0	1 002/7 149	14.0(13.2~15.0)
≥80	208/1 408	14.8	671/2 600	25.8(24.0~27.3)

注: ^aCDC报告数据截至2月11日; ^b模型估计结果截至3月24日

情的防控决策提供重要科学依据。在疫情结束前,或者卫生行政部门发布年龄别病死率信息前,利用统计学方法对其进行估计,具有重要的意义。本研究基于中国CDC和卫健委等多来源的数据,采用MCMC模拟的方法,结合发病到死亡的时间分布规律,对COVID-19的年龄别病死率进行估计。得到COVID-19发病到死亡时间 $M=13.77$ ($P_{25} \sim P_{75}$: 9.03~21.02) d, 总病死率为4.1% (95%CI: 3.7%~4.4%), ≤40岁组病死率均<1.0%。50~、60~、70~和≥80岁组病死率分别为2.3%、6.4%、14.0%和25.8%。模型预测与实际信息的拟合效果较好,如SMAPE较小、 R^2 接近1、 χ^2 拟合优度检验拒绝零假设等,说明本研究的发现稳健可靠。

全球各地区的病死率差异较大。对全球截至4月6日报告COVID-19病例超过1 000例的63个国家的粗病死率进行Meta分析^[9],得到病死率的95%CI为0.72%~10.37%,其中病死率最高的前3位国家分别为阿尔及利亚(13.04%)、意大利(12.67%)和英国(11.69%),而中国(4.07%)则排在第21位。而截至4月10日中国31个省份现有确诊病例1 089例(其中重症病例141例),此时的粗病死率为3 339/81 953=4.07%,与本研究估计的总病死率4.1%较为接近。

年龄别病死率需要权威部门更新数据的支持。Verity等^[6]估计的年龄别病死率在≥30岁的各年龄组较本研究估计的结果低,可能与其总病死率(3.67%)存在低估有关。Hauser等^[10]估计的湖北省年龄别病死率,各组别均高于本研究估计的全国结果高,可能与其设定较高的总病死率为6.0%有关,高于截至4月10日的湖北省粗病死率(3 219/67 803=4.7%)。因此,本研究估计的年龄别病死率可能更具有参考意义。

疫情结束前病死率的估计需要考虑删失的影

响。截至3月24日,中国31个省份报告累计报告确诊病例81 218例,累计死亡病例3 281例,粗病死率为4.0%^[3]。若采用2月11日的数据计算,则粗病死率为2.3%^[7]。在疫情不同时期计算的粗病死率不同,是累计确诊病例和死亡病例都可能在增长的固有影响。疫情早期粗病死率存在较大波动,随着疫情稳定,粗病死率逐渐稳定在4.0%水平。因此,本研究基于发病到死亡的时间分布规律,考虑现有确诊病例特别是重症病例的结局删失影响,符合理论实际的要求。

考虑删失影响估计病死率的主要统计学方法有MCMC模拟、传染病动力学模型和Kaplan-Meier估计等^[11]。传染病动力学模型与MCMC模拟方法一样,仅要求整合的数据信息(如发病到死亡分布和每日累计死亡病例数等),但该方法需要对模型中各个仓室间的关系进行估算,准确的估计要求更多的仓室,而保守或激进的假设都将影响目标参数(即年龄别病死率)的求解;而Kaplan-Meier估计可以较为准确的估计不同人群(如年龄别)的病死率,但该方法需基于个案数据信息(如病例的发病时间、死亡时间、治愈时间和删失时间等)进行分析,在实际应用中难以推广。因无法获得国内的病例个案信息,Verity等^[6]采用Kaplan-Meier估计方法对1 334例中国31个省份以外的COVID-19病例进行分析,得到病死率为4.1% (95%CI: 2.1%~7.8%)。Hauser等^[10]采用传染病动力学模型对湖北省的疫情进行分析,得到病死率为6.0% (95%CI: 5.3%~6.9%),存在高估的可能。因此,本研究采用的MCMC模拟方法无需考虑传染病动力学模型中仓室的个数以及仓室间的关系,也无需Kaplan-Meier方法要求的个案数据信息,具有更广泛的应用前景。但从公共卫生领域长远发展来看,建立疫情数据的有效共享利用制度非常重要,不仅能最大化发挥监测体系的作用,并且有利于疫情的防控遏制。

本研究存在局限性。第一,本研究基于48例死亡病例对发病到死亡的分布进行估计,按照中心极限定理,当样本量≥30时,即可准确获得分布的集中趋势,但离散程度的精确估计需要更多的样本量^[12]。第二,本研究假设各年龄组发病到死亡的时间分布一致,对可能具有更长延迟时间的年龄组存在高估的风险。第三,本研究未考虑无症状和轻症等未被收治病例,但截至4月7日尚在医学观察的无症状感染者占1 095/81 802=1.3%^[13],结果可能存在高估风险,但风险较小。第四,本研究未考虑武汉市4月16日

核增的死亡病例情况^[14],结果可能存在低估的风险,但因无法获取同期的发病例数,导致风险的大小难以量化。

综上所述,校正删失的MCMC模拟的方法适用于新发传染病疫情结束前的病死率估计,尽早明确COVID-19的病死率有助于制定更有效的疫情防控策略。

利益冲突 所有作者均声明不存在利益冲突

参 考 文 献

- [1] Johns Hopkins University. Coronavirus COVID-19 global cases [EB/OL]. (2020-04-10) [2020-04-10]. <https://coronavirus.jhu.edu/map.html>.
- [2] WHO. Coronavirus disease (COVID-19) R&D [EB/OL]. (2020-02-12) [2020-03-12]. <https://www.who.int/blueprint/priority-diseases/key-action/novel-coronavirus/en/>.
- [3] 国家卫生健康委员会. 截至3月24日24时新型冠状病毒肺炎疫情最新情况[EB/OL]. (2020-03-25) [2020-04-11]. <http://www.nhc.gov.cn/xcs/yqtb/202003/b882c06edf184fbf800d4c7957e02dad.shtml>. National Health Commission. The latest situation of the novel coronavirus epidemic as of 24: 00 on March 24 [EB/OL]. (2020-03-25) [2020-04-11]. <http://www.nhc.gov.cn/xcs/yqtb/202003/b882c06edf184fbf800d4c7957e02dad.shtml>.
- [4] 国家卫生健康委员会. 17例死亡病例病情介绍[EB/OL]. (2020-01-23) [2020-04-08]. <http://www.nhc.gov.cn/xcs/yqtb/202001/5d19a4f6d3154b9fae328918ed2e3c8a.shtml>. National Health Commission. Introduction on the 17 death cases [EB/OL]. (2020-01-23) [2020-04-08]. <http://www.nhc.gov.cn/xcs/yqtb/202001/5d19a4f6d3154b9fae328918ed2e3c8a.shtml>.
- [5] WHO. Report of the WHO-China joint mission on coronavirus disease 2019 (COVID-19) [EB/OL]. (2020-02-28) [2020-04-08]. [https://www.who.int/publications-detail/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-\(covid-19\)](https://www.who.int/publications-detail/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-(covid-19)).
- [6] Verity R, Okell LC, Dorigatti I, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis [J]. Lancet Infec Dis, 2020, 20 (6) : 669–677. DOI: 10.1016/S1473-3099(20)30243-7.
- [7] 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]. Chin J Epidemiol, 2020, 41 (2) : 145–151. DOI: 10.3760/cma.j.issn.0254-6450.2020.02.003.
- [8] 杜志成, 顾菁, 李菁华, 等. 基于区间删失数据估计方法的COVID-19潜伏期分布估计[J]. 中华流行病学杂志, 2020, 41 (7) : 1000–1003. DOI: 10.3760/cma.j.cn112338-20200313-00331. Du ZC, Gu J, Li JH, et al. Estimating the distribution of COVID-19 incubation period by interval-censored data estimation method [J]. Chin J Epidemiol, 2020, 41 (7) : 1000–1003. DOI: 10.3760/cma.j.cn112338-20200313-00331.
- [9] Oke J, Heneghan C. Global Covid-19 case fatality rates [EB/OL]. (2020-03-17) [2020-04-11]. <https://www.cebm.net/covid-19/global-covid-19-case-fatality-rates/>.
- [10] Hauser A, Counotte MJ, Margossian CC, et al. Estimation of SARS-CoV-2 mortality during the early stages of an epidemic: a modelling study in Hubei, China and northern Italy [J]. medRxiv, 2020:2020–2023. DOI: 10.1101/2020.03.04.20031104.
- [11] Ghani AC, Donnelly CA, Cox DR, et al. Methods for estimating the case fatality ratio for a novel, emerging infectious disease[J]. Am J Epidemiol, 2005, 162 (5) : 479–486. DOI: 10.1093/aje/kwi230.
- [12] Chang H, Wu C, Ho J, et al. On sample size in using central limit theorem for gamma distribution [J]. Inform and Manag Sci, 2008, 19(1):153–174. DOI: 10.1.1.574.9623.
- [13] 吴尊友. 新型冠状病毒肺炎无症状感染者在疫情传播中的作用与防控策略[J]. 中华流行病学杂志, 2020, 41(6) : 801–805. DOI: 10.3760/cma.j.cn112338-20200406-00517. Wu ZY. Contribution of asymptomatic and pre-symptomatic cases of COVID-19 in spreading virus and targeted control strategies [J]. Chin J Epidemiol, 2020, 41 (6) : 801–805. DOI: 10.3760/cma.j.cn112338-20200406-00517.
- [14] 湖北省卫生健康委员会. 关于订正2020年4月16日湖北省新冠肺炎疫情的情况说明[EB/OL]. (2020-04-17) [2020-04-18]. http://wjw.hubei.gov.cn/fbjd/dtyw/202004/t20200417_2233509.shtml. Hubei Health Commission. Explanation on revising the pneumonia epidemic situation in COVID-19, Hubei Province on April 16, 2020 [EB/OL]. (2020-04-17) [2020-04-18]. http://wjw.hubei.gov.cn/fbjd/dtyw/202004/t20200417_2233509.shtml.

(收稿日期:2020-06-09)

(本文编辑:李银鸽)