



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



# Distinctive trajectories of the COVID-19 epidemic by age and gender: A retrospective modeling of the epidemic in South Korea

Xinhua Yu\*, Jiasong Duan, Yu Jiang, Hongmei Zhang

Division of Epidemiology, Biostatistics, and Environmental Health, School of Public Health, University of Memphis, Memphis, TN 38152, United States



## ARTICLE INFO

### Article history:

Received 27 May 2020

Received in revised form 26 June 2020

Accepted 29 June 2020

### Keywords:

Epidemic dynamics

Elderly

Age interaction

Gender interaction

COVID-19

South Korea

Generalized additive model

Negative binomial distribution

## ABSTRACT

**Objectives:** Elderly people had suffered a disproportionate burden of COVID-19. We hypothesized that males and females in different age groups might have different epidemic trajectories.

**Methods:** Using publicly available data from South Korea, daily new COVID-19 cases were assessed using generalized additive models, assuming Poisson and negative binomial distributions. Epidemic dynamics by age and gender groups were explored using interactions between smoothed time terms and age and gender.

**Results:** A negative binomial distribution fitted the daily case counts best. The relationship between the dynamic patterns of daily new cases and age groups was statistically significant ( $p < 0.001$ ), but this was not the case with gender groups. People aged 20–39 years led the epidemic processes in South Korean society with two peaks — one major peak around March 1 and a smaller peak around April 7, 2020. The epidemic process among people aged 60 or above trailed behind that of the younger age group, and with smaller magnitude. After March 15, there was a consistent decline of daily new cases among elderly people, despite large fluctuations in case counts among young adults.

**Conclusions:** Although young people drove the COVID-19 epidemic throughout society, with multiple rebounds, elderly people could still be protected from infection after the peak of the epidemic.

© 2020 The Author(s). Published by Elsevier Ltd on behalf of International Society for Infectious Diseases. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Introduction

The novel severe acute respiratory syndrome-associated beta-coronavirus (SARS-CoV-2), of unknown origin, appeared in Wuhan, China in late December 2019 and swept across the world over the months that followed (Anderson et al., 2020; Li et al., 2020a; Zhu et al., 2020), causing over 491,500 deaths worldwide (<https://coronavirus.jhu.edu/map.html>, accessed on June 26, 2020) and significantly disrupting both societal activities and personal lives (Pew Research Center, 2020). Although several early studies described the dynamics of the epidemic process in details (Li et al., 2020a; Wu and McGoogan, 2020), many uncertainties remained. For example, diagnosis criteria varied significantly across countries. During the early epidemic in Wuhan, China, patients were required to have serious pneumonia symptoms plus lab-confirmed

virus detection (Huang et al., 2020; Zhu et al., 2020), thus missing most mildly symptomatic and all asymptomatic patients. According to a modeling study, probably 86% of COVID-19 cases might have been undocumented in Wuhan (Li et al., 2020b). Many epidemic measures, such as the basic reproduction number based on the early epidemic in Wuhan, were questioned by later studies due to the possible underestimating of parameters (Nishiura et al., 2020; Zhao et al., 2020a,b). On the other hand, some countries, such as South Korea and Singapore, classified patients based only on lab tests, yielding a better picture of the epidemic.

To fully understand the epidemic process of COVID-19, accurate and complete epidemic data are indispensable. Data from South Korea have been generally considered to be of the highest quality, mainly due to two notable strategies adopted by the South Korean government from the beginning of the epidemic: extensive contact tracing and large-scale testing to identify possible cases, in addition to case isolation (Shim et al., 2020). South Korea identified the first COVID-19 case on January 20, 2020, and the outbreak started its exponential growth after February 19, 2020. In an outbreak that occurred in a call center, 1143 people were tested and

\* Corresponding author.

E-mail addresses: [xyu2@memphis.edu](mailto:xyu2@memphis.edu) (X. Yu), [jduan@memphis.edu](mailto:jduan@memphis.edu) (J. Duan), [yjiang4@memphis.edu](mailto:yjiang4@memphis.edu) (Y. Jiang), [h Zhang@memphis.edu](mailto:h Zhang@memphis.edu) (H. Zhang).

97 were positive and confirmed (positive rate 8.5%) (Park et al., 2020). After tracing all contacts of those 97 cases, about 16% were tested positive (secondary attack rate). In addition, South Korea also installed roadside testing stations to test any person who had concerns about his/her infectious status, in addition to those who had contacted known patients. Such extensive controlling measures not only halted the epidemic successfully but also produced a more complete picture of the COVID-19 epidemic.

A striking COVID-19 phenomenon was that people aged 65 or older suffered the heaviest burden of the disease (Richardson et al., 2020; Wu and McGoogan, 2020), while the proportion of cases was higher in men than in women. According to a recent CDC report, about 80% of deaths occurred among elderly people, with those aged 80 or above having almost a 15% chance of dying if infected (CDC, 2020; Garg et al., 2020). In our previous analysis based on Florida COVID-19 data, we found that people aged 65 or older accounted for 54% of hospitalizations and 82% of deaths. The mortality rate was 14% among elderly people who were infected with coronavirus (Yu, 2020a).

Since May 1, 2020, the COVID-19 pandemic has been waning across the world (<https://coronavirus.jhu.edu/map.html>), pressing many countries to consider re-opening businesses. Many public health experts warned of a possible rebound of new cases if current interventions were relaxed (Chowell and Mizumoto, 2020; Ferguson et al., 2020; Kissler et al., 2020). A recent model predicted that the COVID-19 epidemic might last more than a year and that multiple waves of outbreaks were possible (Kissler et al., 2020). It is likely that elderly people will still suffer the heaviest disease burden during a return of the outbreak (Hay et al., 2020).

However, it was unknown whether or how the epidemic processes were different between young and old people. Our study aimed to statistically examine the dynamics of the COVID-19 pandemic, based on data from South Korea. In addition to identifying the best fit of the epidemic process, we explored gender- and age group-specific trajectories of COVID-19 to facilitate our understanding of the disease and its impact on different populations, and to assess the potential for, and severity of, a COVID-19 rebound.

## Materials and methods

Daily counts of confirmed new COVID-19 cases and deaths were obtained from an open source (<https://github.com/jihoo-kim/Data-Science-for-COVID-19>, accessed on May 2, 2020), which used data systematically gathered from the Korea Center for Disease Control (KCDC) daily reports. All cases were verified against KCDC reports. The line list file included patient's age, gender, and date of virus infection confirmation. However, this file excluded almost all cases occurring in the city of Daegu (more than 6000 cases), and thus cases from Daegu were excluded from our study. We further excluded cases with missing confirmation date ( $n=3$ ). Age was grouped (in years) as 0–19, 20–39, 40–59, and 60 or above. Those with missing gender information ( $n=78$ ) or missing age information ( $n=86$ ) were retained in the analysis for overall trajectories (total sample size  $n=3349$ ), but were excluded in the gender- or age-specific analysis.

Since our purpose in this study was not to predict new cases in the future but to model the epidemic process, we adopted a semi-parametric generalized additive model (GAM) to obtain fitted daily case counts, and also to account for non-linear patterns in the epidemic process (Wood, 2017). Time was modeled as a continuous variable with smoothing terms (thin-plate regression splines with eight knots). Interactions between smooth terms and gender (or age group) were modeled as separate smoothing functions for each

group. Specifically, for interaction models:

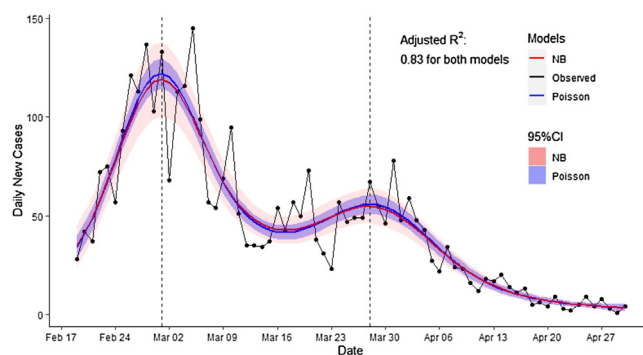
$$\log_e(E(Y_{ij})) = \beta_0 + \sum_{j=1}^J I_j \left( \sum_{k=1}^K \beta_{j,k} \cdot b_k(\text{time}_i) \right)$$

where  $Y_{ij}$  represented the observed case count for day  $i$  and group  $j$  that followed a certain distribution. For this study, we focused on negative binomial (NB) or Poisson distributions due to their robustness. We used the variable  $\text{time}_i$  to represent day starting from 1,  $I_j()$  as an indicator variable (0/1), denoting whether daily counts of new cases were for group  $j$  (1) or not (0),  $b_k()$  to represent a basis function for the  $k$ th term to smooth the temporal trend, and  $\beta_{j,k}$  as regression coefficients for smooth term  $k$  and group  $j$  (representing group-specific effects). Parameters were estimated via the restricted maximum likelihood (REML) approach. The generalized cross validation criterion with Mallows' Cp (GCV.Cp) and maximum likelihood (ML) methods were also explored. Therefore, the above GAM framework allowed us to compare different trajectories through examining the interactions between smoothed time terms and age/gender groups, with a focus on comparing the overall trajectories rather than point-wise comparisons. The statistics  $R^2$  and percent of deviance explained by the models were used to identify the best-fit model. The R package *mgcv* was used to fit the GAM model (Wood, 2017). The aforementioned data and programs are available online at <https://github.com/Jiasong-Duan/COVID-19-epidemic-trajectories>.

## Results

From February 19 to April 30, 2020, there were 3349 COVID-19 cases (1439 males; 43%) identified outside Daegu city. Those aged 0–19 accounted for 6% ( $n=202$ ) of total cases, ages 20–39 for 37% ( $n=1227$ ), ages 40–59 for 31% ( $n=1034$ ), and those aged 60 or above for 24% ( $n=800$ ). As shown in Figure 1, the epidemic outside Daegu city peaked around March 1, 2020 and declined afterwards, except for a second, small peak around March 28, 2020.

The fitted curves for the observed daily new cases were overlaid on the observed counts in Figure 1. Predictions from both NB and Poisson models were indistinguishable. However, the confidence intervals from the NB model were much wider than those for the Poisson model. As shown in the model comparison table (Table 1), there were no differences between different estimating methods in the adjusted  $R^2$  and percent deviance explained by the same model. The adjusted  $R^2$  was 0.839 with 89.2% deviance explained by the Poisson model, while the adjusted  $R^2$  from the NB model was 0.838 with 90.3% deviance (Table 1). Although both models resulted in similar model-fitting parameters, the NB model also



**Figure 1.** Epidemic curve of COVID-19 and predictions from generalized additive models, South Korea, February 19 to April 30, 2020. Note: NB: negative binomial.

**Table 1**  
Model comparisons for fitting the COVID-19 epidemic curves, South Korea.

Model	Method	$R^2_{adjusted}$	Deviance explained
Poisson	REML (default)	0.8327	89.10%
	GCV.Cp	0.8332	89.10%
	ML	0.8328	89.10%
NB	REML (default)	0.8319	90.30%
	GCV.Cp	0.8319	90.30%
	ML	0.8322	90.30%

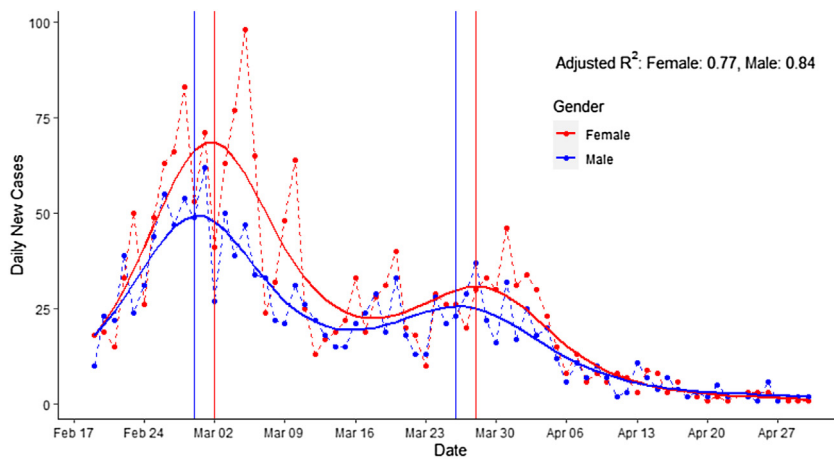
REML: restricted maximum likelihood; GCV.Cp: generalized cross validation criterion with Mallows' Cp; ML: maximum likelihood; NB: negative binomial.

estimated a dispersion factor of 18.2, implying that the Poisson distribution might not be a suitable choice to fit the data. In addition, the wider confidence intervals from the NB model covered a greater range of observed values. Thus, to be conservative, the model based on NB distribution was selected and implemented in the subsequent analyses. The confidence intervals from the fitted models were omitted in the subsequent plots to emphasize the different overall patterns in the epidemic process.

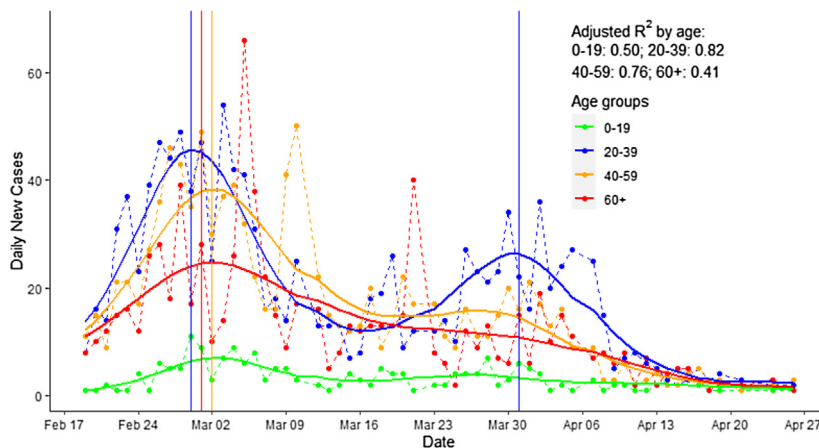
Figure 2a and b shows the fitted epidemic processes by gender and age groups. The epidemic curve for males fell significantly below that for females ( $p = 0.0006$ ). Although the epidemic curve for males peaked about 1 day earlier than that for females, as

shown in Figure 2a, the shapes of the curves were not significantly different between males and females ( $p$  for interaction = 0.35). On the other hand, the age-specific epidemic curves depicted significantly different patterns across age groups ( $p$  for interaction < 0.001) (Figure 2b). The epidemic curve for the youngest group (aged 0–19) showed the lowest daily case counts and was largely stable over the whole period, while there were two peaks in the epidemic process for people aged 20–39 years. Moreover, the epidemic among people aged 20–39 led the whole epidemic process in the total population, such that not only did young adults have more daily new cases than other age groups, but also the epidemic processes for those aged 40–59 and 60+ years trailed 1–3 days behind that for the 20–39 years group.

To further explore age and gender effects on the epidemic process, we plotted the fitted epidemic curves, by age group, for males and females separately (Figure 3a–b). Among males, those aged 20–39 had the highest predicted daily counts and experienced two peaks over time, while those aged 60 or older had much lower daily case counts, with the curve decreasing consistently over time despite large fluctuations in the epidemic for young adults. Those aged 40–59 also experienced two peaks in the epidemic, but these were at a smaller scale compared with those for young adults.

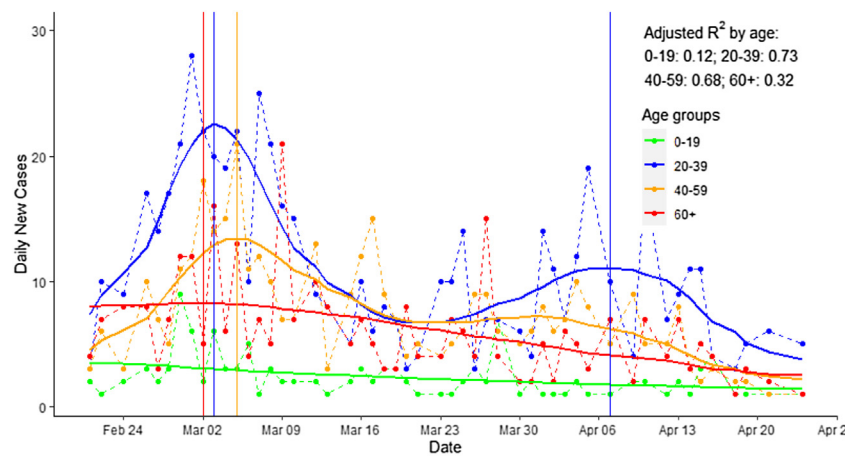


(a)

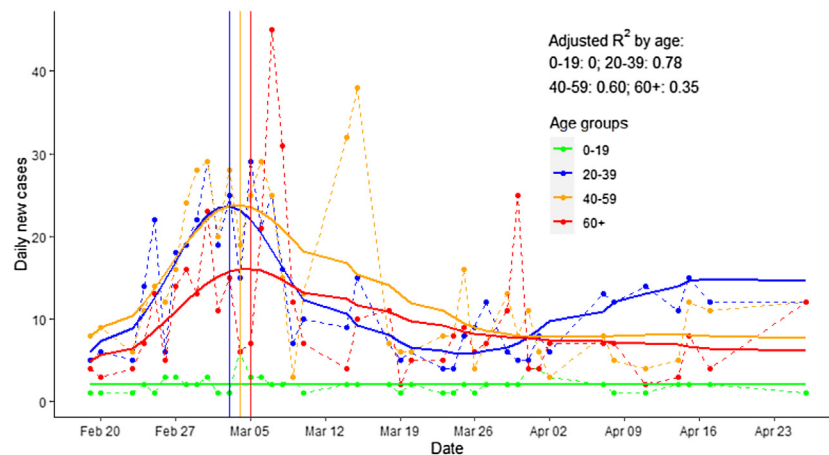


(b)

**Figure 2.** Trajectories of the COVID-19 epidemic process by (a) gender and (b) age group, South Korea.



(a)



(b)

**Figure 3.** Trajectories of the COVID-19 epidemic by age group, South Korea, for (a) males and (b) females.

The patterns of epidemic processes, by age group, among females were different from those of males. Those aged 40–59 and 20–39 showed similar epidemic processes during the first peak. The daily case counts among females aged 20–39 also increased after April 1, 2020. Females aged 60 or above had a less pronounced epidemic but, overall, it was similar to that for females aged 40–59.

## Discussion

Our study demonstrated different trajectories of the COVID-19 epidemic according to gender and age groups, based on South Korean data. First, based on case reporting date and assuming similar incubation periods and reporting delays across all groups and over the whole study period, young people aged 20–39 years led the epidemic processes across society, experiencing two peaks about 1 month apart — one major peak around March 1 and a smaller peak around April 7, 2020. Second, those aged 0–19 years experienced a much smaller magnitude of epidemic overall. Finally, the epidemic process among people aged 60 or above trailed behind that of younger people, with the magnitude of the epidemic smaller than that for people aged 20–39 or 40–59. After March 15, there was a steady decline in daily new cases among people aged 60 or above, despite large fluctuations in case counts among young adults.

Our findings were consistent with other reports, in which younger people accounted for most confirmed COVID-19 cases (Guan et al., 2020; Wu and McGoogan, 2020; Zhang et al., 2020). Our empirical evidence from high-quality data supported the idea that the COVID-19 epidemic was driven by infection among young adults. In addition, children had the lowest burden of disease, possibly due to early school closure and vacation breaks during that period. This pattern was different from that for typical respiratory infection diseases such as seasonal flu, in which most cases tend to be school-age children.

Worldwide, people aged 60 or above endured a disproportionate burden of COVID-19 disease (Wu and McGoogan, 2020). They had a higher risk of hospitalization, with around 80% of deaths occurring in this age group (Garg et al., 2020). However, it was unclear whether elderly people were more likely to get infected, whether virus transmissibility was higher among the elderly, or whether elderly people were merely more likely to have severe diseases than younger people (Hay et al., 2020; Zhang et al., 2020). Elderly people generally have weaker immune systems than younger people. On the other hand, they have been exposed to many viruses over their lifetime, which may help protect them from infection by a new virus, but there was no evidence for any prior immunity to SARS-CoV2. Nonetheless, our findings provided some hope for mitigating the impact of the epidemic on this

vulnerable population. As demonstrated in [Figures 2b and 3a and b](#), fitted daily case counts for those aged 60 or older declined consistently after March 15, 2020, despite a second peak occurring in early April among people aged 20–39. Although we did not have access to detailed information on health conditions and behavioral changes among elderly South Korean people during the COVID-19 pandemic, we believe that by promptly isolating cases, applying extensive contact tracing, and placing at-risk people in quarantine, early and efficiently, together with social distancing, avoiding contact with young cases, and proper personal protection ([Anderson et al., 2020](#); [Shim et al., 2020](#)), elderly people could be effectively protected from viral infection, despite a second rebound in young adults. South Korea has already set an excellent model for other countries to consider. For example, so far there have been only around 10,000 cases and 282 deaths during the COVID-19 epidemic in South Korea ([http://www.cdc.gov/cdc\\_eng/](http://www.cdc.gov/cdc_eng/), accessed on June 25, 2020).

Although overall gender differences in the COVID-19 epidemic were moderate, our age- and gender-specific analyses suggested that females (and to a lesser extent, males) aged 40–59 had a similar experience of the epidemic to that of people aged 20–39. This might be because this age group had close and frequent contacts with younger people in work or within households. Though the risks of hospitalization and death were low among this population, they were higher than for regular respiratory infectious diseases such as seasonal flu. Thus, the disease burden in this middle-age group should not be neglected.

There were some limitations in this study. First, we excluded cases from the city of Daegu (over 6000 cases) because detailed information on cases in that city was not released to the public. Although it was unlikely to bias our results, information from such a large outbreak could have provided additional insights on how the epidemic unfolded among people of different ages and gender. However, during the early stage of the epidemic, few gender- and age-stratified data were publicly available, while most individual-level data from other regions were also incomplete.

Second, we employed statistical methods to examine the epidemic trajectories. There were two perspectives for modeling the epidemic process ([Hethcote, 2000](#); [Unkel et al., 2012](#)). One common approach was to model the process based on the mechanisms of the epidemic. For example, the susceptible-exposed-infectious-removed (SEIR) model and its variants had been used to assess the dynamics of the epidemic, obtain epidemic parameters, and evaluate the impact of various control measures ([Kucharski et al., 2020](#); [Peak et al., 2017](#); [Prem et al., 2020](#); [Yu, 2020b](#)). Agent-based models had also been used to simulate the epidemic process and assess the effects of various interventions ([Ferguson et al., 2020](#); [Wu et al., 2020](#)). The other perspective was based on traditional statistical models. Non-linear models, such as the generalized logistic growth model ([Chowell, 2017](#)), had been used to model the growth of the epidemic and estimate the growth rate of cases over time. In addition, some researchers had directly modeled the epidemic curve with regression techniques, assuming that daily counts follow certain distributions, such as Poisson or negative binomial distributions. For example, models based on time series of count data were adopted to predict COVID-19 deaths in the US, including models from the Institute of Health Metrics and Evaluation (IHME) ([IHME, 2020](#)) and the University of Texas-Austin ([Woody et al., 2020](#)). Our previous research had also used vector autoregressive models to examine the risk interactions across age groups after the peak of the COVID-19 epidemic ([Yu, 2020c](#)). While there were many uncertainties among different gender and age groups about contact patterns, virus transmissibility, and behavioral changes during the epidemic, since the epidemic data from South Korea were more

likely to be complete, it was possible to directly model the daily counts with regression models, assuming a common distribution for the count data. We believe that our models avoided the many unfounded assumptions found in the more complicated epidemic process models.

Third, we used only case reporting or lab confirmation dates in this study, which were likely to be 3–5 days away from the actual virus infection date. The average incubation time for COVID-19 was reported as being about 5 days ([Lauer et al., 2020](#)) and the report delay in South Korea was unknown, but likely to have been very short due to extensive testing. Thus, we made some untestable assumptions in comparing epidemic trajectories between age and gender groups. The incubation periods and reporting delays were assumed to be the same across all groups and over the whole study period. This should have been pertinent in South Korea because mass testing and contact tracing started at the beginning of the epidemic ([Shim et al., 2020](#)), but may not be appropriate for regions where testing is severely limited and delayed.

Finally, we only analyzed data from South Korea. The epidemic processes for COVID-19 in different countries are likely to vary due to differences in population structure and in interventions used to mitigate the epidemic ([Anderson et al., 2020](#); [Chowell and Mizumoto, 2020](#); [Hay et al., 2020](#); [Lipsitch et al., 2020](#)). However, we still expect our findings to provide a general picture of the epidemic trajectories of COVID-19, and to serve as a reference for other regions. In addition, as witnessed elsewhere in this COVID-19 epidemic, politics and ideology can often overtake science and public health, so that effective interventions are sometimes implemented too late or are incomplete, leaving the public confused and public health practitioners in a conundrum.

The main strength of our study was the straightforward nature of our analyses in exploring different epidemic processes, based on high-quality data. Insights often emerge through such modeling exercises. We stratified the models by age and gender and discovered their different trajectories during the epidemic. Recent studies have predicted a long-lasting epidemic for COVID-19 and possible multiple waves of outbreaks after societal reopening ([Kissler et al., 2020](#)). Our findings were unique in providing empirical evidence for designing effective public health strategies to mitigate the impact of recurrent COVID-19 epidemics, and to protect vulnerable populations.

## Conclusions

In South Korea, and probably in other countries, COVID-19 epidemic processes have had distinctive dynamic patterns among age and gender groups. The epidemic among young adults led the epidemic process across the whole population, with a second peak occurring in people aged 20–39 years. Most importantly, during the post-peak period of the COVID-19 epidemic and in the process of gradually returning society and the economy to normalcy, elderly people could be protected effectively through case isolation, contact tracing, mass testing, and proper personal protection, as exemplified in South Korea.

## Funding

Dr. Xinhua Yu and Mr. Jiasong Duan were supported by the FedEx Institute of Technology, University of Memphis in conducting this research.

## Ethics statement

This study used only publicly available data and no human subjects were directly involved. It was thus deemed to be exempt

from approval by the Institutional Review Board. No informed consent was needed. All authors declared no conflicts of interest in conducting this study.

### Conflict of interest

None declared.

### References

- Anderson RM, Heesterbeek H, Klinkenberg D, Hollingsworth TD. How will country-based mitigation measures influence the course of the COVID-19 epidemic? *Lancet* 2020;395:931–4, doi:http://dx.doi.org/10.1016/S0140-6736(20)30567-5.
- CDC. Severe outcomes among patients with coronavirus disease 2019 (COVID-19) – United States, February 12–March 16, 2020. *MMWR Morb Mortal Wkly Rep* 2020;343–6.
- Chowell G. Fitting dynamic models to epidemic outbreaks with quantified uncertainty: a Primer for parameter uncertainty, identifiability, and forecasts. *Infect Dis Model* 2017;2:379–98, doi:http://dx.doi.org/10.1016/j.idm.2017.08.001.
- Chowell G, Mizumoto K. The COVID-19 pandemic in the USA: what might we expect? *Lancet* 2020;395:1093–4, doi:http://dx.doi.org/10.1016/S0140-6736(20)30743-1.
- Ferguson N, Laydon D, Nedjati Gilani G, Imai N, Ainslie K, Baguelin M, et al. Report 9: impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. Imperial College, UK; 2020, doi:http://dx.doi.org/10.25561/77482.
- Garg S, Kim L, Whitaker M, O'Halloran A, Cummings C, Holstein R, et al. Hospitalization rates and characteristics of patients hospitalized with laboratory-confirmed coronavirus disease 2019 – COVID-NET, 14 States, March 1–30, 2020. *MMWR Morb Mortal Wkly Rep* 2020;69:458–64, doi:http://dx.doi.org/10.15585/mmwr.mm6915e3.
- Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, et al. Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med* 2020;, doi:http://dx.doi.org/10.1056/NEJMoa2002032.
- Hay JA, Haw DJ, Hanage WP, Metcalf CJE, Mina MJ. Implications of the age profile of the novel coronavirus. *Proc Natl Acad Sci U S A* 2020;.
- Hethcote HW. The mathematics of infectious diseases. *SIAM Rev* 2000;42:54.
- Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020;395:497–506, doi:http://dx.doi.org/10.1016/S0140-6736(20)30183-5.
- IHME. Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator-days and deaths by US state in the next 4 months. IHME COVID-19 Health Service Utilization Forecasting Team; 2020.
- Kissler SM, Tedijanto C, Goldstein E, Grad YH, Lipsitch M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science* 2020;, doi:http://dx.doi.org/10.1126/science.abb5793.
- Kucharski AJ, Russell TW, Diamond C, Liu Y, Edmunds J, Funk S, et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect Dis* 2020;, doi:http://dx.doi.org/10.1016/S1473-3099(20)30144-4.
- Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Ann Intern Med* 2020;, doi:http://dx.doi.org/10.7326/M20-0504.
- Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N Engl J Med* 2020a;382:1199–207, doi:http://dx.doi.org/10.1056/NEJMoa2001316.
- Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* 2020b;368:489–93, doi:http://dx.doi.org/10.1126/science.abb3221.
- Lipsitch M, Swerdlow DL, Finelli L. Defining the epidemiology of Covid-19 – studies needed. *N Engl J Med* 2020;382:1194–6, doi:http://dx.doi.org/10.1056/NEJMp2002125.
- Nishiura H, Linton NM, Akhmetzhanov AR. Serial interval of novel coronavirus (COVID-19) infections. *Int J Infect Dis* 2020;93:284–6, doi:http://dx.doi.org/10.1016/j.ijid.2020.02.060.
- Park SY, Kim YM, Yi S, Lee S, Na BJ, Kim CB, et al. Coronavirus disease outbreak in call center, South Korea. *Emerg Infect Dis* 2020;26;, doi:http://dx.doi.org/10.3201/eid2608.201274.
- Peak CM, Childs LM, Grad YH, Buckee CO. Comparing nonpharmaceutical interventions for containing emerging epidemics. *Proc Natl Acad Sci U S A* 2017;114:4023–8, doi:http://dx.doi.org/10.1073/pnas.1616438114.
- Pew Research Center. Most Americans say coronavirus outbreak has impacted their lives. 2020 <https://www.pewsocialtrends.org/2020/03/30/most-americans-say-coronavirus-outbreak-has-impacted-their-lives/> [Accessed 18 June 2020].
- Prem K, Liu Y, Russell TW, Kucharski AJ, Eggo RM, Davies N, et al. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *Lancet Public Health* 2020;, doi:http://dx.doi.org/10.1016/S2468-2667(20)30073-6.
- Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW, et al. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City area. *JAMA* 2020;, doi:http://dx.doi.org/10.1001/jama.2020.6775.
- Shim E, Tariq A, Choi W, Lee Y, Chowell G. Transmission potential and severity of COVID-19 in South Korea. *Int J Infect Dis* 2020;93:339–44, doi:http://dx.doi.org/10.1016/j.ijid.2020.03.031.
- Unkel S, Farrington CP, Garthwaite PH, Robertson C, Andrews N. Statistical methods for the prospective detection of infectious disease outbreaks: a review. *J R Stat Soc A* 2012;175:33.
- Wood SN. Generalized additive models: an introduction with R. 2nd ed. Chapman and Hall/CRC; 2017.
- Woody S, Garcia Tec M, Dahan M, Gaither K, Lachmann M, Fox S, et al. Projections for first-wave COVID-19 deaths across the US using social-distancing measures derived from mobile phones. *medRxiv* 2020;, doi:http://dx.doi.org/10.1101/2020.04.16.20068163.
- Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet* 2020;395:689–97, doi:http://dx.doi.org/10.1016/S0140-6736(20)30260-9.
- Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72314 cases from the Chinese Center for Disease Control and Prevention. *JAMA* 2020;, doi:http://dx.doi.org/10.1001/jama.2020.2648.
- Yu X. Did elderly people living in small towns or rural areas suffer heavier disease burden during the COVID-19 epidemic? *medRxiv* 2020a;, doi:http://dx.doi.org/10.1101/2020.05.01.20087791.
- Yu X. Modeling return of the epidemic: impact of population structure, asymptomatic infection, case importation and personal contacts. *medRxiv* 2020b;, doi:http://dx.doi.org/10.1101/2020.04.26.20081109.
- Yu X. Risk interactions of coronavirus infection across age groups after the peak of COVID-19 epidemic. *medRxiv* 2020c;, doi:http://dx.doi.org/10.1101/2020.05.17.20105049.
- Zhang J, Litvinova M, Liang Y, Wang Y, Wang W, Zhao S, et al. Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. *Science* 2020;, doi:http://dx.doi.org/10.1126/science.abb8001.
- Zhao S, Cao P, Gao D, Zhuang Z, Cai Y, Ran J, et al. Serial interval in determining the estimation of reproduction number of the novel coronavirus disease (COVID-19) during the early outbreak. *J Travel Med* 2020a;, doi:http://dx.doi.org/10.1093/jtm/taaa033.
- Zhao S, Lin Q, Ran J, Musa SS, Yang G, Wang W, et al. The basic reproduction number of novel coronavirus (2019-nCoV) estimation based on exponential growth in the early outbreak in China from 2019 to 2020: a reply to Dhungana. *Int J Infect Dis* 2020b;, doi:http://dx.doi.org/10.1016/j.ijid.2020.02.025.
- Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med* 2020;382:727–33, doi:http://dx.doi.org/10.1056/NEJMoa2001017.