



Article

# Risk Interactions of Coronavirus Infection across Age Groups after the Peak of COVID-19 Epidemic

Xinhua Yu

Division of Epidemiology, Biostatistics & Environmental Health, School of Public Health,  
University of Memphis, Memphis, TN 38152, USA; xyu2@memphis.edu; Tel.: +1-901-678-3433

Received: 18 May 2020; Accepted: 20 July 2020; Published: 21 July 2020



**Abstract:** Background: The COVID-19 pandemic has incurred significant disease burden worldwide, particularly on the elderly population. This study aims to explore how risks of coronavirus infection interact across age groups using data from South Korea. Methods: Daily new COVID-19 cases from 10 March to 30 April 2020 were scraped from online open sources. A multivariate vector autoregressive model for time series of count data was used to examine the risk interactions across age groups. Case counts from previous days were included as predictors to dynamically examine the change of risk patterns. Results: In South Korea, the risk of coronavirus infection among elderly people was significantly affected by other age groups. An increase in virus infection among people aged 20–39 was associated with a double risk of infection among elderly people. Meanwhile, an increase in virus infection among elderly people was also significantly associated with risks of infection among other age groups. The risks of infection among younger people were relatively unaffected by that of other age groups. Conclusions: Protecting elderly people from coronavirus infection could not only reduce the risk of infection among themselves but also ameliorate the risks of virus infection among other age groups. Such interventions should be effective and for the long term.

**Keywords:** COVID-19; elderly people; risk interaction; South Korea; virus infection; SARS-CoV-2

## 1. Introduction

Coronavirus disease 2019 (COVID-19) is caused by the infection of a novel severe acute respiratory syndrome-associated coronavirus (SARS-CoV-2) [1,2]. Since December 2019, over 11 million people have been infected with SARS-CoV-2 and over 528,000 people have died of coronavirus infection (<https://coronavirus.jhu.edu/map.html>, accessed on 4 July 2020). Of them, elderly people and people with underlying chronic conditions suffered the heaviest disease burden [3–5]. For example, in the US, as reported by the US Center for Disease Control and Intervention, about 80% of deaths were people aged 65 or above (<https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html>), and 43.4% of hospitalizations were people aged 65 or above [6,7]. On the other hand, the case fatality rate of COVID-19 in South Korea was about 2.1% as of 16 July 2020, according to the Korea Center for Disease Control report (<http://ncov.mohw.go.kr/en/>).

The reasons for the disproportional burden among elderly people were unclear [8]. Elderly people generally have weaker immune systems than younger people due to aging, and they are also more likely to have multiple chronic conditions [9,10]. Thus, elderly people may have severe symptoms if infected with coronavirus [11,12]. On the other hand, elderly people may have been exposed to myriads of infections over their lifetime which may provide partial immunity against new virus infection. Although cross-reaction of antibodies between SARS-CoV and SARS-CoV-2 was observed, cross-neutralization was rare [13]. Thus, it was unlikely elderly people might have any effective immunity against SARS-CoV-2.

Epidemic data of high quality are essential to understand the mechanism of an epidemic and compare the courses of epidemic and risk patterns of infection across different groups. Unfortunately, due to insufficient testing kits, heterogeneous diagnosis criteria, and varied implementation of different interventions, epidemiological data on COVID-19 pandemic among different countries (and even within a country) and during different periods were often noncomparable [14]. One notable exception is South Korea, where extensive contact tracing and mass testing not only curtailed the epidemic but also generated high-quality data. In South Korea, both asymptomatic and symptomatic cases were identified promptly from the very beginning of the epidemic [15,16]. Thus, depicting a complete picture of the epidemic process was possible in South Korea.

In South Korea, the first case of COVID-19 occurred on 20 January 2020 and the major outbreak started on 19 February 2020. The COVID-19 pandemic was waning down since 10 March 2020 (see Figure 1) [17,18]. However, potential rebounds of new cases have been warned by many public health experts [19]. This is reflected in an epidemic curve with a long tail and occasional spikes, which is demonstrated in the epidemic process in South Korea (<https://www.kcdc.info/covid-19/>) [15,16]. In addition, if the seasonality, immunity, and cross-immunity of SARS-CoV-2 behave like previous coronaviruses, a recent study predicted that a long-lasting and multi-wave epidemic was possible [20]. Furthermore, after the first epidemic peak is passed, which suggests mitigating interventions are likely effective, the society is expected to gradually return to normalcy [21]. Strategies for re-opening the society should be evaluated, and exploring risk patterns of infection during the post-peak period will provide important evidence for decision-making. In the COVID-19 epidemic, given the disproportional disease burden born by elderly people, it is imperative to examine risk patterns of coronavirus infection among elderly people after the peak of the epidemic. Epidemic data from South Korea will allow us to examine the risk interactions among different age groups, which can serve as an exemplary reference model for other countries.

In this study, we will examine how risks of coronavirus infection interact across age groups using time series analysis. Using the high-quality data from South Korea, we will focus on the post-peak period of the epidemic process to evaluate the risk of infection among elderly people during the period of society re-opening.

## 2. Materials and Methods

Daily new COVID-19 case counts from South Korea were obtained from the website (<https://www.kaggle.com/kimjihoo/coronavirusdataset>) which were scraped from the Korea Center for Disease Control website. The first COVID-19 case in South Korea appeared on 20 January 2020, and the major epidemic started on 19 February 2020. Since the peak of the first major epidemic wave in South Korea ended around 10 March [22] (also Figure 1), we limited the time series of new cases from 10 March to 30 April 2020. All daily cases were stratified by age groups (0–19, 20–39, 40–59, and 60 or above). Those aged 60 or above were referred to as elderly people.

The observed epidemic curves by age groups from 10 March to 30 April 2020 were plotted, and the predicted daily cases were obtained from a generalized additive model (GAM), [23] assuming daily new cases follow Poisson distributions. The smoothness of predicted values was achieved with thin plate regression splines with 16 knots using the R *mgcv* package (see Appendix A). The number of knots was determined empirically in which 16 knots best captured the non-linearity of the epidemic curve without being too smooth and too rough.

We developed a vector autoregressive (VAR) model to examine the associations of the infection risks across age groups simultaneously [24]. Specifically, we assumed daily new case counts ( $y_{j,t}$ ) followed a generalized Poisson distribution to account for over-dispersion of case counts (i.e., observed variance is larger than expected variance) [25]. The model also included case counts from previous days (lags) across age groups as predictors to form a dynamic model (see Appendix A for details).

Therefore, the current risk of infection in each age group was predicted not only by previous case counts in its own group but also by previous counts from other age groups.

$$\begin{aligned} y_{j,t} | \alpha_j, \beta_{j,k}, b_{j,t} &\sim \text{generalized Poisson}(\mu_{j,t}, \xi_j) \\ \mu_{j,t} &= \exp\left(\alpha_j + \sum_{j=1}^J \sum_{k=1}^K \beta_{j,k} \ln(y_{j,t-k}) + b_{j,t}\right) \\ b_{j,t} | \Sigma &\sim \text{MultiNormal}(0, \Sigma) \end{aligned} \quad (1)$$

where  $j = 1, \dots, J$  represents age groups,  $t = 1, \dots, T$  represents days, and  $k = 1, \dots, K$  represents the number of time lags. Since the typical incubation period of COVID-19 is five days [26], we reported results from five-lag models, that is, the daily counts from one prior day (yesterday), two prior days (the day before yesterday), and from up till five days before were included as predictors in the models. Three-lag and seven-lag models were also explored, and results from all models were consistent (see online codes and results). The scale parameter  $\xi$  in the generalized Poisson distribution controls the magnitude of dispersion, that is,  $\xi = 0$  corresponding to a standard Poisson (mean = variance),  $\xi < 0$  suggesting under-dispersion (mean > variance), and  $0 < \xi < 1$  indicating overdispersion (mean < variance). The  $b_{j,t}$  could be viewed as a random effect to account for the correlation of daily counts between age groups. The  $b_{j,t}$  was assumed a multivariate normal distribution.

The above model framework was similar to the common log-linear relative risk models in epidemiological studies which assume multiplicative associations between predictors and outcomes [27]. The coefficients  $\beta$ s could be interpreted as natural logarithms of risk ratios per one-unit change of natural logarithms of case counts.

We fit the above models with the Bayesian software stan through the Rstan interface (<http://mc-stan.org>) [28]. To keep the model simple, we assumed weakly informative priors of student t distributions for all  $\alpha$ s and  $\beta$ s, and an LKJ prior with modal density around diagonals for correlations between case series (see Appendix A). Hamiltonian Monte Carlo was used to obtain posterior distributions of parameters. Diagnostic plots showed all chains mixed satisfactorily and were converged. In addition, negative binomial models were also fitted, and results were similar to those reported here except for wider confidence intervals (Appendix B, Table A1). Notice that the dispersion factors ( $1/\xi$ ) estimated from the generalized Poisson models were 2.66 (1.58–5.13), 1.44 (0.96–2.46), 2.21 (1.37–3.82), and 0.90 (0.59–1.50) for those aged 60 or above, 40–59, 20–39, and 0–19, respectively. These estimates were of moderate magnitude and two of them (age group 40–59 and 0–19) were not statistically different from 1. This also suggested that negative binomial models might overestimate the dispersion factors which led to wider confidence intervals. The data, replicable codes, and other results are available online ([www.github.com/xinhuayu/riskinteractions/](http://www.github.com/xinhuayu/riskinteractions/)).

*Ethics statement:* This study was based on publicly available data. There was no direct involvement of human subjects. Therefore, it was exempted from the approval of the institutional review board. No informed consent was needed. All authors declared no conflict of interest in conducting this study.

### 3. Results

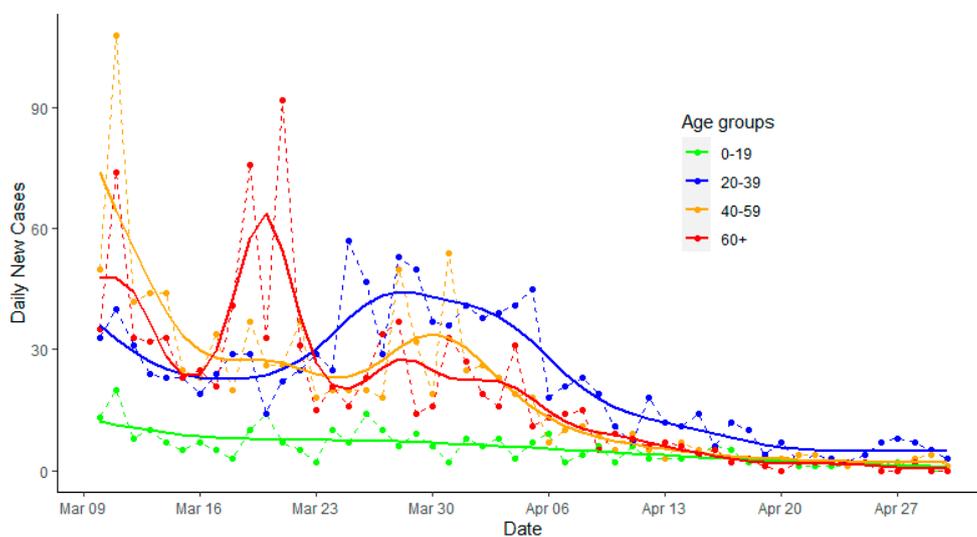
In South Korea, there were 3383 COVID-19 cases between 10 March and 30 April 2020. Of them, 283 cases were people aged 0–19 (8.4%), 1141 aged 20–39 (50.0%), 987 aged 40–59 (29.2%), and 972 aged 60 or above (28.7%).

Figure 1 presents the epidemic curves with fitted values by age groups for South Korea. After 10 March, there was a small spike among those aged 60 or above around 20 March 2020, and a small rebound among those aged 20–39 (e.g., around 30 March to 5 April 2020), followed by those aged 40–59 and aged 60 or older.

Table 1 describes associations of risks of infection across age groups in South Korea. In addition to tracking the effect from the first lag day (yesterday) among elderly people, the current daily counts of infection among elderly people were doubled (risk ratio = 2) with a one-unit increase in the natural

logarithm of case counts (or 2.8 times increase in daily counts) on the day before yesterday (lag 2) among those aged 20–39. Additionally, an increase in infection during the fourth prior day (lag 4) in those aged 20–39 was associated with a decrease in daily counts in elderly people, possibly due to a periodic effect. As the epidemic came in waves, an increase in the case counts in the young age group would have taken an additional four days to qualitatively impact the risk of infection among elderly people. Finally, an increase in infection in the youngest population during the fifth lag days was also associated with an increased risk of infection among elderly people by 64%.

More importantly, an increase in virus infection among elderly people was associated with increases in risks of infection among all other age groups, but with longer delays in younger populations. Furthermore, the risk of infection among people aged 40–59 was affected by both old and young people, but to a lesser extent. Risks of infection among people aged 20–39 or 0–19 were less likely affected by other age groups.



**Figure 1.** Observed daily new cases of COVID-19 and predicted epidemic curve from the generalized additive model, South Korea, from 10 March to 30 April 2020, by age groups.

**Table 1.** Risk ratios in the coronavirus infection across age groups during the COVID-19 pandemic, South Korea, 10 March to 30 April 2020.

Model Outcomes	Predictors	Lags of Predictors				
		Lag 1	Lag 2	Lag 3	Lag 4	Lag 5
Aged 60 or above						
	60 or above	2.09 (1.28–3.17) *	1.21 (0.78–1.80)	0.90 (0.60–1.39)	1.03 (0.60–1.67)	0.96 (0.63–1.53)
	40–59	0.95 (0.49–1.84)	1.81 (0.98–3.29)	0.59 (0.30–1.18)	1.37 (0.85–2.23)	1.03 (0.61–1.70)
	20–39	0.89 (0.53–1.46)	2.02 (1.12–3.47) *	1.28 (0.72–2.31)	0.41 (0.22–0.78) #	1.11 (0.58–2.22)
	0–19	0.90 (0.61–1.43)	1.36 (0.95–1.95)	0.82 (0.57–1.19)	0.61 (0.44–0.85) #	1.64 (1.03–2.58) *
Aged 40–59						
	60 or above	1.66 (1.19–2.29) *	0.79 (0.55–1.12)	0.89 (0.64–1.23)	0.98 (0.70–1.40)	1.02 (0.71–1.40)
	40–59	1.01 (0.62–1.63)	1.76 (1.16–2.66) *	1.23 (0.76–1.96)	0.89 (0.62–1.29)	0.88 (0.62–1.28)
	20–39	1.13 (0.75–1.73)	1.12 (0.76–1.67)	1.59 (1.03–2.50) *	0.60 (0.37–0.95)	0.88 (0.53–1.48)
	0–19	1.04 (0.79–1.37)	1.28 (0.97–1.68)	1.17 (0.88–1.57)	1.14 (0.88–1.49)	1.17 (0.84–1.62)
Aged 20–39						
	60 or above	0.95 (0.69–1.29)	0.89 (0.65–1.23)	1.01 (0.72–1.39)	1.54 (1.11–2.12) *	0.97 (0.69–1.31)
	40–59	1.17 (0.73–1.88)	1.18 (0.78–1.78)	0.98 (0.60–1.52)	0.90 (0.64–1.26)	1.04 (0.75–1.47)
	20–39	1.56 (1.05–2.36) *	1.02 (0.68–1.55)	1.45 (0.96–2.24)	1.06 (0.68–1.64)	0.85 (0.55–1.32)
	0–19	1.04 (0.79–1.37)	0.96 (0.75–1.27)	0.88 (0.66–1.20)	0.97 (0.75–1.26)	0.74 (0.54–1.00)
Aged 0–19						
	60 or above	1.78 (1.23–2.61) *	1.34 (0.86–2.06)	0.99 (0.67–1.51)	0.82 (0.52–1.25)	1.55 (1.02–2.30) *
	40–59	0.76 (0.41–1.38)	0.71 (0.42–1.18)	0.85 (0.47–1.56)	0.91 (0.55–1.57)	0.97 (0.62–1.54)
	20–39	1.51 (0.92–2.55)	1.50 (0.85–2.57)	1.04 (0.61–1.80)	0.78 (0.44–1.38)	0.91 (0.51–1.62)
	0–19	0.93 (0.64–1.36)	0.85 (0.60–1.21)	0.85 (0.60–1.27)	0.92 (0.62–1.38)	1.51 (0.99–2.31)

Note: \* and # for  $p < 0.05$ . Lag 1 of predictors means daily counts from previous day (yesterday), lag 2 means daily counts from the day before yesterday and so forth.

#### 4. Discussion

This was the first study to quantify risk interactions of SARS-CoV-2 infection across age groups based on vector autoregressive models using epidemic data of high quality from South Korea. We found that in South Korea, the risk of infection among elderly people was significantly affected by other age groups. An increase in virus infection among elderly people was also significantly associated with increased risks of infection among other age groups. Risks of infections among younger people were relatively unaffected by that of other age groups.

Our results were consistent with the current COVID-19 epidemic process, in which risk of infections among elderly people might be affected by other age groups [8]. Although virus transmission might differ among age groups [8,29], the risk interactions were likely due to personal interactions between people of different age groups. Respiratory infectious diseases often spread through personal contacts [30]. Previous studies showed that contacts were more frequent in young age groups than older age groups, and interactions across age groups were less frequent than within each age group [31]. During the emerging pandemic of COVID-19, stringent control measures such as lock-down, strict social distancing, and stay-at-home rules, were often implemented promptly, leading to an abrupt and significant change in contact patterns within and between age groups. Modern techniques such as contact tracing apps, infection risk IDs, and instant notifications of cases also allowed us to efficiently isolate cases and quarantine high-risk people. The observed risk interactions between age groups in South Korea might be largely due to the change in contact patterns during the pandemic period. As shown in our study, there were two–five lagging days in the risk interactions across age groups, especially between old and young people. On the other hand, the infection among elderly people may still be affected by and also affect the risks of infection among other age groups. In addition, in the US, half of the infected cases could not be traced back to known cases in a telephone survey conducted in outpatient and inpatient settings of academic medical centers [32]. Therefore, although elderly people may refrain from contacting known cases, they may be still at risk of infection by doing regular activities such as grocery shopping. Passive community interactions may play an important role in sustaining the epidemic.

Our results highlighted the importance of implementing and enforcing effective interventions on the whole society [33–35], and the higher priority of protecting elderly people [29]. Furthermore, we showed that an increase in coronavirus infection among elderly people was associated with increased risks of infection among other age groups, suggesting protecting elderly people and reducing the risk of infection among elderly people had a spillover effect on the whole society. This was consistent with our previous simulation study in which reducing contacts among the elderly could reduce the virus infection and hospitalizations in the whole society [36].

There were some limitations in this study. The most important one was that we relied on reported cases. The data from South Korea were more likely complete due to extensive contact tracing and mass testing. Furthermore, the case reporting date (or virus infection detection/lab confirmation date) was different from the virus infection date, and the average incubation period for SARS-CoV-2 was about five days [26]. The laudable efforts of extensive contact tracing and mass testing implemented by the South Korea government since the beginning of COVID-19 epidemic significantly reduced reporting delays, and likely identified many cases before symptom onsets [15]. Therefore, the interval between virus infection and case reporting might be small. In addition, there were other factors such as gender, socio-economic status, and neighborhood environment which might have also affected the risk of infection.

Moreover, although we interpreted the results with action terms, they had no explicit causative meanings. For example, younger people tended to have milder or no symptoms (i.e., subclinical cases) if infected with the coronavirus [37–39]. Thus, it was possible that an increased number of detected cases among young people implied the existence of an increase in subclinical cases in the community who might unknowingly infect other people, including elderly people. Subclinical cases could only be

identified through extensive contact tracing and mass testing. Without this information, it is impossible to examine the route of infections in the community.

Our study has several strengths. Firstly, data from South Korea were more likely complete which provided instrumental information about underlying epidemic mechanisms. Although different social norms and health care systems might explain some differences in risk patterns between South Korea and other regions, results from South Korea provided a baseline picture of risk interactions among age groups under a well-controlled, ideal epidemic process. Different patterns might be due to differences in population structures, magnitudes of control measures, and contact patterns in the society, while similar patterns in risk interactions between regions allowed us to infer the possible paths of infections.

Secondly, we proposed a novel multivariate autoregressive model for time series of counts to examine the risk of virus infection across age groups simultaneously. A flexible generalized Poisson model fitted with Bayesian methods was used to account for overdispersion of count data [25]. Unlike many other studies that used mechanistic epidemic models which were useful to describe the epidemic process [40], our statistical models extended traditional relative risk models to time series of count data. It should be noted that this type of model likely overfits the data, and collinearity among lag variables also exists. Thus, having a priori hypotheses and choosing biologically relevant lags are critical in building correct models and interpreting the results. Our lag models were based on the observed incubation period of COVID-19 and for testing pre-specified hypotheses. The principle of our methods was similar to that of the Institute for Health Metrics and Evaluation (IHME) [41] and University of Texas-Austin models [42], all of which relied on time series analysis of count data. However, we did not attempt to predict future cases. Rather, we focused on disentangling risk interactions of infection across age groups, which is more important and relevant in disease prevention.

Finally, during the process of re-opening the economy and society, the number of new cases may rebound, and multiple small waves or a second big wave of epidemic are possible. A contentious issue was whether and how to protect high-risk populations such as elderly people during the return of the epidemic. Therefore, we limited our study period to the post-peak of the epidemic to answer this imminent question. Our study strongly supported that high-risk populations such as elderly people should still take serious precautions during the post-epidemic period.

## 5. Conclusions

In summary, protecting elderly people from coronavirus infection might not only be associated with a reduced risk of infection among themselves but also related to lower risks of virus infection among other age groups. Therefore, elderly people should keep on practicing social distancing and maintaining effective personal protection while still maintaining a good care of their comorbidities.

**Funding:** This research was funded by seed grant for data science from FedEx Institute of Technology at the University of Memphis.

**Acknowledgments:** We sincerely appreciate two anonymous reviewers for their insightful comments and suggestions that significantly improved the manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

SARS-CoV-2	severe acute respiratory syndrome coronavirus 2
COVID-19	coronavirus infectious disease 2019
GAM	generalized additive model
VAR	vector autoregressive regression
NB	negative binomial model
REML	restricted maximum likelihood
PMF	probability mass function

## Appendix A

### Appendix A.1. Obtained Smoothed Predicted Daily Cases with Generalized Additive Model (GAM)

Assuming daily new cases follow a Poisson distribution or negative binomial (NB) distribution (see below), the GAM is a linear regression with a smoothed time term. For simplicity, a separate GAM regression was fitted for each age group:

$$\log_e(E(Y_{ij})) = \beta_0 + \sum_{k=1}^K \beta_{j,k} \cdot b_k(\text{time}_i) \quad (\text{A1})$$

where  $Y_{ij}$  represents the observed case counts of day  $i$  and group  $j$ , and  $E(Y_{ij})$  is the expected (predicted) value. The variable  $\text{time}_i$  represents day  $(1, \dots, I)$ ,  $b_k(\cdot)$  represents a basis function for the  $k$ th term to the smooth temporal trend, and  $\beta_{j,k}$ s are regression coefficients for the smooth term  $k$  and group  $j$ . The restricted maximum likelihood (REML) approach was used in the parameter estimation. The R *mgcv* package was used [23], and smooth terms were fitted using a thin plate regression spline with 16 knots. The number of knots is typically determined empirically, commonly with 8–20 knots. A smaller number of knots yields a smoother curve while a larger number of knots results in a rougher curve. We observed that 16 knots are reasonable to capture the nonlinearity of the epidemic curve without introducing too many spikes.

### Appendix A.2. Model Setups and Comparisons

The standard Poisson distribution describes the distribution of  $y$  events occurring at a constant rate of  $\lambda$ . The probability mass function (PMF) is

$$\begin{aligned} p(Y = y|\lambda) &= \frac{\lambda^y}{y!} e^{-\lambda}, \lambda > 0 \\ E(Y) &\stackrel{\text{def}}{=} \mu = \lambda = \text{Var}(Y) \\ \text{Loglik} &= y \log(\lambda) - \lambda - \log(y!) \end{aligned} \quad (\text{A2})$$

In standard Poisson distribution, expected variance equals the mean. If observed variance is larger than expected variance (i.e., the mean), then overdispersion exists. This often occurs when outcomes are correlated, such as daily new case counts during a disease outbreak.

The generalized Poisson distribution introduces an additional scale parameter  $\xi$  [43] as quoted in Hilbe JM. 2014 [27]. The PMF is

$$\begin{aligned} p(Y = y|\lambda, \xi) &= \frac{\lambda}{y!} (\lambda + \xi y)^{y-1} e^{-(\lambda + \xi y)}, \lambda > 0 \\ E(Y) &\stackrel{\text{def}}{=} \mu = \frac{\lambda}{1-\xi} \\ \text{Var}(Y) &= \frac{\lambda}{(1-\xi)^3} = \frac{\mu}{(1-\xi)^2} \end{aligned} \quad (\text{A3})$$

Thus,  $\varphi \stackrel{\text{def}}{=} \frac{1}{(1-\xi)^2}$  is the dispersion factor indicating how variance changes with the mean. Therefore, if  $\xi = 0$ , then  $\varphi = 1$  corresponds to a standard Poisson (mean = variance);  $0 < \xi < 1$ , then  $\varphi > 1$  models overdispersion (mean < variance); and  $\xi < 0$ , then  $\varphi < 1$  models under-dispersion (mean > variance).

Reparametrize the PMF of generalized Poisson distribution with  $\mu$  and  $\xi$  [25]:

$$\begin{aligned} p(Y = y|\mu, \xi) &= \frac{\mu(1-\xi)}{y!} (\mu - \xi(\mu - y))^{y-1} e^{-(\mu - \xi(\mu - y))} \\ \text{LogLik} &= \log(\mu(1-\xi)) + (y-1) \log(\mu - \xi(\mu - y)) - (\mu - \xi(\mu - y)) - \log(y!) \end{aligned} \quad (\text{A4})$$

On the other hand, the negative binomial distribution describes the distribution of the number of successes given a predefined  $r$  number of failures during a sequence of independent Bernoulli trials with a success probability  $p$ :

$$\begin{aligned}
 p(Y = y) &= \binom{y + r - 1}{y} p^y (1 - p)^r \\
 E(Y) &\stackrel{\text{def}}{=} \mu = \frac{rp}{1-p} \\
 Var(Y) &= \frac{rp}{(1-p)^2} = \frac{\mu}{1-p} = \mu + \frac{\mu^2}{r}
 \end{aligned}
 \tag{A5}$$

Thus, the overdispersion of  $Y$  is controlled by the shape parameter  $r$ . Reparametrize the PMF with  $\mu$  and  $r$ :

$$p(Y = y) = \binom{y + r - 1}{y} \left(\frac{\mu}{\mu + r}\right)^y \left(\frac{r}{\mu + r}\right)^r
 \tag{A6}$$

$$\text{Loglik} = yr(\log(\mu r) - 2\log(\mu + r)) + \log(\Gamma(y + r)) - \log(\Gamma(y + 1)) - \log(\Gamma(r))
 \tag{A7}$$

where gamma function  $\Gamma(x+1) = x!$  ( $x$  factorial) is for an integer  $x$ , and the parameter  $r$  can be any positive real value.

Note that negative binomial distribution can be viewed as a gamma–Poisson mixture distribution in which  $Y \sim \text{Poisson}(\lambda)$  and  $\lambda \sim \text{gamma}(r, \lambda/r)$ . That is, the negative binomial distribution  $(r, p)$  is the posterior distribution of  $\text{Poisson}(\lambda)$  with  $\text{gamma}(r, \lambda/r)$  as the conjugate prior of  $\lambda$ , where  $\lambda \stackrel{\text{def}}{=} \mu = rp/(1 - p)$ . Rewriting  $\text{gamma}(r, \lambda/r)$  as  $\text{gamma}(r, p/(1 - p))$  and using  $\Gamma$  functions to represent factorials:

$$\int_0^\infty \frac{\lambda^y e^{-\lambda}}{\Gamma(y + 1)} \cdot \frac{\lambda^{r-1} e^{-\lambda(\frac{1-p}{p})}}{\left(\frac{p}{1-p}\right)^r \Gamma(r)} d\lambda = \frac{\Gamma(r + y)}{\Gamma(r)\Gamma(y + 1)} p^y (1 - p)^r
 \tag{A8}$$

Under this framework, negative binomial distribution is appealing as a natural extension of Poisson distribution to allow for overdispersion that is controlled by the shape parameter  $r$ .

However, although negative binomial distribution is often used to model new case counts during disease outbreaks, it models only overdispersion and assumes a quadratic relationship between variance and mean, while the generalized Poisson model is more flexible and assumes a simpler first order association between variance and mean. Therefore, we chose to report results from generalized Poisson models. Results from negative binomial models are included in the Appendix B. In addition, it is also of note that there are extensions of negative binomial models in which the association between mean and variance can be estimated from data, leading to a more flexible model and also permitting the exploration of determinants of overdispersion [27].

In this study, we proposed the following hierarchical vector autoregressive model (VAR) for count data:

$$\begin{aligned}
 &y_{j,t} \mid \alpha_j, \beta_{j,k}, b_{j,t} \sim \text{PMF}(\mu_{j,t}) \\
 \mu_{j,t} &= \exp\left(\alpha_j + \sum_{j=1}^J \sum_{k=1}^K \beta_{j,k} \ln(y_{j,t-k}) + b_{j,t}\right) \\
 &b_{j,t} \mid \Sigma \sim \text{MultiNormal}(0, \Sigma)
 \end{aligned}
 \tag{A9}$$

where  $j = 1, \dots, J$  represents age groups,  $t = 1, \dots, T$  represents days, and  $k = 1, \dots, K$  represents the number of lags. The PMF of  $Y$  can be either standard Poisson ( $\lambda$ ), generalized Poisson ( $\mu, \xi$ ), or negative binomial ( $\mu, r$ ) distribution.

The above VAR model included new case counts from previous days (lags) across age groups as predictors [24], thus examining associations of the infection risks across age groups simultaneously.

That is, the current risk of infection in each age group was predicted not only by previous case counts in its own group but also by previous counts from other age groups.

The correlation of daily counts between age groups was modeled through  $b_{j,t}$  that can be viewed as a random effect. The  $b_{j,t}$  was assumed a multivariate normal distribution.

The exponential link between dynamic predictors and  $\mu$  is equivalent to common relative risk models in epidemiological studies, i.e., log-linear models for count data. Under this multiplicative scale framework, the interpretations of  $\beta$ s are relative risks given a one-unit increase in predictors.

During the model fitting, we assumed some weakly informative priors for all parameters:

$$\begin{aligned}
 \alpha_j &\sim t(df = 5, location = 0, scale = 2.5) \\
 \beta_{j,k} &\sim t(df = 5, location = 0, scale = 2.5) \\
 \xi_j &\sim normal(0, 0.3), \text{ and } -0.9 \leq \xi_j \leq 0.9 \\
 \frac{1}{\sqrt{r}} &\stackrel{\text{def}}{=} \varphi \sim \text{positive half-normal}(0, 1) \\
 \Sigma &= D'RD, \\
 \text{where standard deviation } D &= \text{diag}(sd_j), \text{ } sd_j \sim \text{cauchy}(0, 5) \\
 \text{and } R &= \text{correlation matrix}, R \sim \text{LKJ}(2)
 \end{aligned} \tag{A10}$$

The LKJ prior is a special prior most suitable for correlations. The LKJ(2) assumes a modal density surrounding diagonals.

The models were fit with Bayesian software stan through the Rstan interface [28]. A customized stan function was constructed for fitting the generalized Poisson model. We employed Hamiltonian Monte Carlo with five Markov chains, each with 50,000 iterations plus 2000 warmups, to obtain posterior distributions of parameters. Diagnostic plots through the shinestan package showed all chains mixed well and were converged. The replicable data and codes, including models with daily case counts as standard Poisson, generalized Poisson, or negative binomial distributions, are available online ([www.github.com/xinhuayu/riskinteractions/](http://www.github.com/xinhuayu/riskinteractions/)).

**Appendix B. Additional Table**

**Table A1.** Risk interactions in coronavirus infection across age groups based on negative binomial models, COVID-19, South Korea.

Model Outcomes	Predictors	Lags of Predictors				
		Lag 1	Lag 2	Lag 3	Lag 4	Lag 5
Aged 60 or above						
	60 or above	1.93 (1.14–3.19) *	1.09 (0.66–1.80)	1.02 (0.61–1.82)	0.99 (0.57–1.67)	1.11 (0.68–1.87)
	40–59	1.03 (0.53–2.02)	1.66 (0.83–3.01)	0.69 (0.36–1.46)	1.26 (0.72–2.25)	1.00 (0.57–1.71)
	20–39	0.94 (0.54–1.69)	1.30 (0.66–2.42)	1.42 (0.73–2.68)	0.50 (0.25–1.01)	1.21 (0.63–2.37)
	0–19	1.10 (0.72–1.78)	1.29 (0.79–1.98)	0.85 (0.51–1.35)	0.62 (0.40–0.98) #	1.52 (0.88–2.55)
Aged 40–59						
	60 or above	1.64 (1.12–2.41) *	0.75 (0.50–1.09)	0.97 (0.67–1.49)	1.01 (0.69–1.53)	1.02 (0.69–1.53)
	40–59	1.15 (0.68–1.96)	1.55 (0.92–2.46)	1.14 (0.64–2.24)	0.90 (0.54–1.39)	0.95 (0.63–1.47)
	20–39	1.10 (0.59–1.73)	1.13 (0.70–1.81)	1.54 (0.95–2.47)	0.63 (0.37–1.27)	0.89 (0.44–1.56)
	0–19	1.09 (0.77–1.53)	1.23 (0.87–1.78)	1.16 (0.82–1.64)	1.16 (0.82–1.60)	1.15 (0.79–1.68)
Aged 20–39						
	60 or above	1.00 (0.71–1.44)	0.93 (0.65–1.39)	1.00 (0.69–1.48)	1.58 (1.10–2.26) *	0.90 (0.64–1.29)
	40–59	1.12 (0.69–1.86)	1.09 (0.67–1.72)	0.96 (0.58–1.53)	0.97 (0.65–1.44)	1.08 (0.72–1.61)
	20–39	1.63 (1.09–2.48) *	1.04 (0.66–1.64)	1.35 (0.86–2.11)	0.99 (0.58–1.56)	0.96 (0.60–1.56)
	0–19	1.04 (0.74–1.44)	0.95 (0.69–1.29)	0.89 (0.63–1.23)	0.95 (0.67–1.40)	0.73 (0.51–1.04)
Aged 0–19						
	60 or above	1.77 (1.17–2.76) *	1.35 (0.82–2.14)	1.00 (0.62–1.61)	0.81 (0.50–1.32)	1.52 (0.97–2.44)
	40–59	0.76 (0.39–1.42)	0.71 (0.39–1.27)	0.86 (0.43–1.64)	0.94 (0.53–1.66)	0.96 (0.58–1.60)
	20–39	1.53 (0.87–2.72)	1.46 (0.82–2.69)	1.05 (0.57–1.92)	0.78 (0.41–1.41)	0.90 (0.49–1.74)
	0–19	0.94 (0.63–1.44)	0.83 (0.57–1.25)	0.88 (0.58–1.33)	0.92 (0.59–1.44)	1.52 (0.95–2.42)

Note: \* and # for  $p < 0.05$ .

## References

1. Zhu, N.; Zhang, D.; Wang, W.; Li, X.; Yang, B.; Song, J.; Zhao, X.; Huang, B.; Shi, W.; Lu, R.; et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N. Engl. J. Med.* **2020**, *382*, 727–733. [[CrossRef](#)] [[PubMed](#)]
2. Li, Q.; Guan, X.; Wu, P.; Wang, X.; Zhou, L.; Tong, Y.; Ren, R.; Leung, K.S.M.; Lau, E.H.Y.; Wong, J.Y.; et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *N. Engl. J. Med.* **2020**, *382*, 1199–1207. [[CrossRef](#)] [[PubMed](#)]
3. Guan, W.J.; Ni, Z.Y.; Hu, Y.; Liang, W.H.; Ou, C.Q.; He, J.X.; Liu, L.; Shan, H.; Lei, C.L.; Hui, D.S.C.; et al. Clinical Characteristics of Coronavirus Disease 2019 in China. *N. Engl. J. Med.* **2020**. [[CrossRef](#)] [[PubMed](#)]
4. Richardson, S.; Hirsch, J.S.; Narasimhan, M.; Crawford, J.M.; McGinn, T.; Davidson, K.W.; Barnaby, D.P.; Becker, L.B.; Chelico, J.D.; Cohen, S.L.; et al. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. *JAMA* **2020**. [[CrossRef](#)] [[PubMed](#)]
5. Wu, Z.; McGoogan, J.M. 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**. [[CrossRef](#)]
6. Garg, S.; Kim, L.; Whitaker, M.; O'Halloran, A.; Cummings, C.; Holstein, R.; Prill, M.; Chai, S.J.; Kirley, P.D.; Alden, N.B.; 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*, 458–464. [[CrossRef](#)]
7. Yu, X. Urban-rural inequalities during the COVID-19 pandemic among elderly people in Florida, US. *medRxiv* **2020**. [[CrossRef](#)]
8. Hay, J.A.; Haw, D.J.; Hanage, W.P.; Metcalf, C.J.E.; Mina, M.J. Implications of the Age Profile of the Novel Coronavirus. Available online: <https://dash.harvard.edu/handle/1/42639493> (accessed on 4 July 2020).
9. Chavan, P.; Kedia, S.; Yu, X. Physical and Functional Limitations in US Older Cancer Survivors. *J. Palliat. Care Med.* **2017**. [[CrossRef](#)]
10. Ward, B.W.; Schiller, J.S.; Goodman, R.A. Multiple chronic conditions among US adults: A 2012 update. *Prev. Chronic Dis.* **2014**, *11*, E62. [[CrossRef](#)]
11. 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*–346.
12. Imai, N.; Cori, A.; Dorigatti, I.; Baguelin, M.; Donnelly, C.A.; Riley, S.; Ferguson, N.M. *Report 3: Transmissibility of 2019-nCoV*; Imperial College of London: London, UK, 2020.
13. Lv, H.; Wu, N.C.; Tsang, O.T.-Y.; Yuan, M.; Perera, R.A.P.M.; Leung, W.S.; So, R.T.Y.; Chan, J.M.C.; Yip, G.K.; Chik, T.S.H.; et al. Cross-reactive Antibody Response between SARS-CoV-2 and SARS-CoV Infections. *Cell Rep.* **2020**, *31*, 107725. [[CrossRef](#)] [[PubMed](#)]
14. Lipsitch, M.; Swerdlow, D.L.; Finelli, L. Defining the Epidemiology of Covid-19—Studies Needed. *N. Engl. J. Med.* **2020**, *382*, 1194–1196. [[CrossRef](#)]
15. 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*, 339–344. [[CrossRef](#)]
16. Park, S.Y.; Kim, Y.M.; Yi, S.; Lee, S.; Na, B.J.; Kim, C.B.; Kim, J.I.; Kim, H.S.; Kim, Y.B.; Park, Y.; et al. Coronavirus Disease Outbreak in Call Center, South Korea. *Emerg. Infect. Dis.* **2020**, *26*. [[CrossRef](#)] [[PubMed](#)]
17. Kim, S.; Castro, M.C. Spatiotemporal pattern of COVID-19 and government response in South Korea (as of May 31, 2020). *Int. Ernational J. Infect. Dis.* **2020**. [[CrossRef](#)] [[PubMed](#)]
18. Yu, X.; Duan, J.; Jiang, Y.; Zhang, H. Distinctive trajectories of COVID-19 epidemic by age and gender: A retrospective modeling of the epidemic in South Korea. *Int. Ernational J. Infect. Dis.* **2020**. [[CrossRef](#)]
19. Chowell, G.; Mizumoto, K. The COVID-19 pandemic in the USA: What might we expect? *Lancet* **2020**, *395*, 1093–1094. [[CrossRef](#)]
20. Kissler, S.M.; Tedijanto, C.; Goldstein, E.; Grad, Y.H.; Lipsitch, M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science* **2020**. [[CrossRef](#)]

21. Anderson, R.M.; Heesterbeek, H.; Klinkenberg, D.; Hollingsworth, T.D. How will country-based mitigation measures influence the course of the COVID-19 epidemic? *Lancet* **2020**, *395*, 931–934. [[CrossRef](#)]
22. Kim, Y.J.; Seo, M.H.; Yeom, H.E. Estimating a breakpoint in the spread pattern of COVID-19 in South Korea. *Int. J. Infect. Dis.* **2020**. [[CrossRef](#)]
23. Wood, S.N. *Generalized Additive Models: An Introduction with R*, 2nd ed.; Chapman and Hall/CRC: Boca Raton, FL, USA, 2017.
24. Brandt, P.T.; Sandler, T. A bayesian Poisson vector autoregressive model. *Political Anal.* **2012**, *20*, 23. [[CrossRef](#)]
25. Yang, Z.; Hardin, J.W.; Addy, C.L.; Vuong, Q.H. Testing approaches for overdispersion in poisson regression versus the generalized poisson model. *Biom. J.* **2007**, *49*, 565–584. [[CrossRef](#)] [[PubMed](#)]
26. Lauer, S.A.; Grantz, K.H.; Bi, Q.; Jones, F.K.; Zheng, Q.; Meredith, H.R.; Azman, A.S.; Reich, N.G.; Lessler, J. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Ann. Int. Ern. Med.* **2020**. [[CrossRef](#)] [[PubMed](#)]
27. Hilbe, J.M. *Modeling Count Data*; Cambridge University Press: Cambridge, UK, 2014.
28. Carpenter, B.; Gelman, A.; Hoffman, M.D.; Lee, D.; Goodrich, B.; Betancourt, M.; Brubaker, M.; Guo, J.; Li, P.; Riddell, A. Stan: A probabilistic programming language. *J. Stat. Softw.* **2017**, *76*. [[CrossRef](#)]
29. Davies, N.G.; Klepac, P.; Liu, Y.; Prem, K.; Jit, M.; Eggo, R.M. Age-dependent effects in the transmission and control of COVID-19 epidemics. *medRxiv* **2020**. [[CrossRef](#)]
30. Zhang, J.; Litvinova, M.; Liang, Y.; Wang, Y.; Wang, W.; Zhao, S.; Wu, Q.; Merler, S.; Viboud, C.; Vespignani, A.; et al. Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. *Science* **2020**. [[CrossRef](#)]
31. Mossong, J.; Hens, N.; Jit, M.; Beutels, P.; Auranen, K.; Mikolajczyk, R.; Massari, M.; Salmaso, S.; Tomba, G.S.; Wallinga, J.; et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med.* **2008**, *5*, e74. [[CrossRef](#)]
32. Tenforde, M.W.; Billig Rose, E.; Lindsell, C.J.; Shapiro, N.I.; Files, D.C.; Gibbs, K.W.; Prekker, M.E.; Steingrub, J.S.; Smithline, H.A.; Gong, M.N.; et al. Characteristics of Adult Outpatients and Inpatients with COVID-19-11 Academic Medical Centers, United States, March-May 2020. *MMWR Morb Mortal Wkly. Rep.* **2020**, *69*, 841–846. [[CrossRef](#)]
33. Cowling, B.J.; Ho, L.M.; Leung, G.M. Effectiveness of control measures during the SARS epidemic in Beijing: A comparison of the Rt curve and the epidemic curve. *Epidemiol Infect.* **2008**, *136*, 562–566. [[CrossRef](#)]
34. Gostic, K.; Gomez, A.C.; Mummah, R.O.; Kucharski, A.J.; Lloyd-Smith, J.O. Estimated effectiveness of symptom and risk screening to prevent the spread of COVID-19. *Elife* **2020**, *9*. [[CrossRef](#)]
35. Pan, A.; Liu, L.; Wang, C.; Guo, H.; Hao, X.; Wang, Q.; Huang, J.; He, N.; Yu, H.; Lin, X.; et al. Association of Public Health Interventions With the Epidemiology of the COVID-19 Outbreak in Wuhan, China. *JAMA* **2020**. [[CrossRef](#)] [[PubMed](#)]
36. Yu, X. Modeling Return of the Epidemic: Impact of Population Structure, Asymptomatic Infection, Case Importation and Personal Contacts. *medRxiv* **2020**. [[CrossRef](#)]
37. Bai, Y.; Yao, L.; Wei, T.; Tian, F.; Jin, D.Y.; Chen, L.; Wang, M. Presumed Asymptomatic Carrier Transmission of COVID-19. *JAMA* **2020**. [[CrossRef](#)]
38. Li, C.; Ji, F.; Wang, L.; Wang, L.; Hao, J.; Dai, M.; Liu, Y.; Pan, X.; Fu, J.; Li, L.; et al. Asymptomatic and Human-to-Human Transmission of SARS-CoV-2 in a 2-Family Cluster, Xuzhou, China. *Emerg. Infect. Dis.* **2020**, *26*. [[CrossRef](#)]
39. Li, G.; Li, W.; He, X.; Cao, Y. Asymptomatic and Presymptomatic Infectors: Hidden Sources of COVID-19 Disease. *Clin. Infect. Dis.* **2020**. [[CrossRef](#)]
40. Kucharski, A.J.; Russell, T.W.; Diamond, C.; Liu, Y.; Edmunds, J.; Funk, S.; Eggo, R.M. Centre for Mathematical Modelling of Infectious Diseases, C.-w.g. Early dynamics of transmission and control of COVID-19: A mathematical modelling study. *Lancet Infect. Dis.* **2020**. [[CrossRef](#)]
41. 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. *MMWR Morb Mortal Wkly. Rep.* **2020**, *69*, 343–346.

42. Woody, S.; Garcia Tec, M.; Dahan, M.; Gaither, K.; Lachmann, M.; Fox, S.; Meyers, L.A.; Scott, J.G. Projections for first-wave COVID-19 deaths across the US using social-distancing measures derived from mobile phones. *medRxiv* 2020. [[CrossRef](#)]
43. Consul, P. *Generalized Poisson Distribution: Properties and Applications*; Marcel Decker: New York, NY, USA, 1989.



© 2020 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).