1 Article

Risk interactions of coronavirus infection across age 2 groups after the peak of COVID-19 epidemic 3

4 Xinhua Yu 1*

5 ¹ xyu2@memphis.edu

6

7

* Correspondence: xyu2@memphis.edu Tel.: (01) 901-678-3433. Division of Epidemiology, Biostatistics & 8 Environmental Health, School of Public Health, University of Memphis.

9 Received: date; Accepted: date; Published: date

10 Abstract: Background: the COVID-19 pandemic has incurred significant disease burden 11 worldwide, particularly on elderly population. This study aims to explore how risks of infection 12 interact across age groups using data from South Korea. Methods: Daily new COVID-19 cases 13 from March 10 to April 30, 2020 were scraped from online open sources. A multivariate vector 14 autoregressive model for time series count data was used to examine the risk interactions across 15 age groups. Case counts from previous days were included as predictors to dynamically examine 16 the change of risk patterns. Results: In South Korea, the risk of coronavirus infection among 17 elderly people was significantly affected by other age groups. An increase of virus infection 18 among people aged 20-39 was associated with a double risk of infection among elderly people. 19 Meanwhile, an increase in virus infection among elderly people was also significantly associated 20 with risks of infection among other age groups. The risks of infection among younger people were 21 relatively unaffected by that of other age groups. Conclusions: Protecting elderly people from 22 coronavirus infection could not only reduce the risk of infection among themselves but also 23 ameliorate the risks of virus infection among other age groups. Such interventions should be 24 effective and for long term.

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

26

27 1. Introduction

28 The coronavirus disease 2019 (COVID-19) is caused by the infection of a novel Severe Acute 29 Respiratory Syndrome associated coronavirus (SARS-CoV-2) [1]. Since December 2019, over 11 30 million people have been infected with SARS-CoV-2 and over 528,000 people died of coronavirus 31 infection (https://coronavirus.jhu.edu/map.html, accessed on July 4, 2020). Of them, elderly people 32 and people with underlying chronic conditions suffered the heaviest disease burden [2-4]. For 33 example, about 80% deaths or of were people aged 65 above 34 (https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html), and 43.4% of 35 hospitalizations aged 65 or above [5]. In the state of Florida, US, people aged 65 or above accounted 36 for 54% of hospitalizations, and the mortality rate was 14% if infected with virus [6].

37 The reasons for the disproportional burden among elderly people were unclear [7]. Elderly 38 people generally have weaker immune system than younger people due to aging, and they are also 39 more likely to have multiple chronic conditions [8,9]. Thus, elderly people may have severe 40 symptoms if infected with coronavirus [10,11]. On the other hand, elderly people may have 41 exposed myriads of infections over their lifetime which may provide immunity against new virus 42 infection. Although cross-reaction of antibodies between SARS-CoV and SARS-CoV-2 was 43 observed, cross-neutralization was rare [12]. Thus, it was unlikely elderly people might have any 44 effective immunity against SARS-CoV-2.

45 Meanwhile, the COVID-19 pandemic is waning down in some countries such as South Korea 46 since March 10, 2020 (see Figure 1) [13,14], and society is gradually returning to normalcy [15]. A

2 of 13

47 potential rebound of new cases has been warned by many public health experts [16]. This is 48 reflected in an epidemic curve with a long tail and occasional spikes, which is demonstrated in the 49 epidemic process in South Korea (https://www.kcdc.info/covid-19/) [17,18]. In addition, if the 50 seasonality, immunity and cross-immunity of SARS-CoV-2 behave like previous coronaviruses, a 51 recent study predicted a long lasting and multi-wave epidemic was possible in the US [19]. 52 Therefore, it is imperative to examine risk patterns of coronavirus infection among elderly people 53 after the peak of epidemic.

54 Unfortunately, due to lack of testing kits and heterogeneous diagnosis criteria, epidemiological 55 data on COVID-19 among different countries (and even within a country) were often 56 noncomparable [20]. One notable exception is South Korea where extensive contact tracing and 57 mass testing not only curtailed the epidemic but also generated high quality data. In South Korea, 58 both asymptomatic and symptomatic cases were identified promptly [17,18]. Thus, a complete 59 picture of the epidemic process was possible to depict.

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 postpeak period of the epidemic process to evaluate the risk of infection among elderly people during
the period of society re-opening.

64

65 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 Jan 20, 2020, and the major epidemic started on Feb 19, 2020. Since the peak of first epidemic wave in South Korea ended around March 10 [21] (also Figure 1), we limited the time series of new cases between March 10 to April 30, 2020 for South Korea. 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 as elderly people.

The observed epidemic curves by age groups from March 10 to April 30, 2020 were plotted, and the predicted daily cases were obtained with a generalized additive model (GAM) [22] assuming daily new cases follow Poisson distributions. The smoothness of predicted values was achieved with thin plate regression splines with 16 knots using R *mgcv* package (see Appendix A).

We developed a vector autoregressive (VAR) model to examine the associations of the infection risks across age groups simultaneously [23]. 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) [24]. 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.

$$y_{j,t} \mid \alpha_{j}, \beta_{j,k}, b_{j,t} \sim generalized Poisson(\mu_{j,t}, \xi_j)$$
$$\mu_{j,t} = \exp\left(\alpha_j + \sum_{j=1}^{J} \sum_{k=1}^{K} \beta_{j,k} \ln(y_{j,t-k}) + b_{j,t}\right)$$
$$b_{j,t} \mid \Sigma \sim MultiNormal(0, \Sigma)$$

85 Where j = 1,...,J represented age groups, t=1,...,T represented days, and k = 1,...,K represented 86 the number of time lags. Because the typical incubation period of COVID-19 is five days [25], we 87 reported results from five-lag models. Three-lag and seven lag models were also explored, and 88 results from all models were consistent (see online codes and results). The scale parameter ξ in the 89 generalized Poisson distribution controls the magnitude of dispersion, that is, $\xi = 0$ corresponding 80 to a standard Poisson (mean = variance), $\xi < 0$ suggesting under-dispersion (mean > variance), and 0 91 < $\xi < 1$ indicating overdispersion (mean < variance). The $b_{j,i}$ could be viewed as a random effect to

3 of 13

92 account for the correlation of daily counts between age groups. The $b_{j,t}$ was assumed a multivariate 93 normal distribution.

94 The above model framework was similar to the common log-linear relative risk models in
 95 epidemiological studies which assume multiplicative associations between predictors and outcomes
 96 [26]. The coefficients βs could be interpreted as natural logarithms of risk ratios per one unit change
 97 of natural logarithms of case counts.

98 We fit the above models with Bayesian software stan through Rstan interface (http://mc-99 stan.org) [27]. To keep the model simple, we assumed weakly informative priors of student t 100 distributions for all α s and β s, and an LKJ prior with modal density around diagonals for 101 correlations between case series (see Appendix A). Hamiltonian Monte Carlo was used to obtain 102 posterior distributions of parameters. Diagnostic plots showed all chains mixed satisfactorily and 103 were converged. In addition, negative binomial models were also fitted, and results were similar to 104 those reported here except for wider confidence intervals (Appendix B, Tables 1). Noticed that the 105 dispersion factors estimated from the generalized Poisson models were 2.66 (1.58 - 5.13), 1.44 (0.96 -106 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, 107 respectively. These estimates were of moderate magnitude and two of them (age group 40-59 and 0-108 19) were not statistically different from 1. This also suggested that negative binomial models might 109 overestimate the dispersion factors which led to wider confidence intervals. The data, replicable 110 codes and other results were available online (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 Institutional Review
Board. No informed consent was needed. All authors declared no conflict of interest in conducting
this study.

115

116 3. Results

117 In South Korea, there were 3,383 COVID-19 cases between March 10 and April 30, 2020. Of 118 them, 283 cases aged 0-19 (8.4%), 1,141 aged 20-39 (50.0%), 987 aged 40-59 (29.2%), and 972 aged 60 119 or above (28.7%).

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





Model				Lags of predictors		
Outcomes 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 - 5	9					
	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 - 3	9					
	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)

Table 1: Risk interactions in coronavirus infection across age groups during the COVID-19 pandemic, South Korea, March 10 to April 30, 2020
 130

131 Note: * and # for p < 0.05

Table 1 described associations of risks of infection across age groups in South Korea. In addition to tracking effect from the first lag day (yesterday) among elderly people, the current risk of infection among elderly people was associated with double risk by one unit increase of infection risk among those aged 20-39 in the second lag day. Additionally, an increase of 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 of 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 less extent. Risks of infection among people aged 20-39 or 0-19 were less likely affected by other age groups.

144

145 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.

152 Our results were consistent with the current COVID-19 epidemic process, in which risk of 153 infections among elderly people might be affected by other age groups [28]. Although virus 154 transmission might differ among age groups [7,29], the risk interactions were likely due to personal 155 interactions between people of different age groups. Respiratory infectious diseases often spread 156 through personal contacts [30]. Previous studies showed that contacts were more frequent in young 157 age groups than older age groups, and interactions across age groups were less frequent than 158 within each age group [31]. During the emerging pandemic like COVID-19, stringent control 159 measures such as lock-down, strict social distancing and stay-at-home rules, were often 160 implemented promptly, leading to a significantly abrupt change of contact patterns within and 161 between age groups. Modern techniques such as contract tracing app, infection risk ID, and instant 162 notification of cases, also allowed us to efficiently isolate cases and quarantine high risk people. The 163 observed risk interactions between age groups in South Korea might be largely due to the change of 164 contact patterns during the epidemic period. As shown in our study, there were 2-5 lagging days in 165 the risk interactions across age groups, especial between old and young people. On the other hand, 166 the infection among elderly people may still be affected by and also affect the risks of infection 167 among other age groups. Passive community interactions such as grocery shopping might play an 168 important role in sustaining the epidemic.

Our results highlighted the importance of implementing and enforcing effective interventions in the whole society [32-34], and the highest priority of protecting elderly people [29]. Furthermore, we showed that an increase of 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 spillover effect in the whole society. This was consistent with our previous simulation study in which reducing contacts among elderly could reduce the virus infection and hospitalizations in the whole society [35].

176 There were some limitations in this study. The most important one was that we relied on 177 reported cases. The data from South Korea were more likely complete due to extensive contact 178 tracing and mass testing. Furthermore, the case reporting date (or virus infection detection/lab 179 confirmation date) was different from the virus infection date, and the average incubation period 180 for SARS-CoV-2 was about 5 days [25]. The laudable efforts of extensive contact tracing and mass 181 testing implemented by the South Korea government at the beginning of COVID-19 epidemic 182 significantly reduced the reporting delays, and likely identified many cases before symptom onsets 183 [17]. Therefore, the interval between virus infection and case reporting might be small. In addition,

2 of 13

184 there were other factors such as gender, socio-economic status and neighborhood environment 185 might also affect 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 virus [36-38]. 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.

193 Our study has several strengths. Firstly, data from South Korea were more likely complete 194 which would provide information about underlying epidemic mechanisms. Although different 195 social norms and health care systems might explain some differences in risk patterns between South 196 Korea and other regions, results from South Korea provided a baseline picture of risk interactions 197 among age groups under a well-controlled, ideal epidemic process. Different patterns might be due 198 to differences in population structures, magnitudes of control measures and contact patterns in the 199 society, while similar patterns in risk interactions between regions allowed us to infer the possible 200 paths of infections.

201 Secondly, we proposed a novel multivariate autoregressive model for time series of counts to 202 examine the risk of virus infection across age groups simultaneously. A flexible generalized Poisson 203 model fitted with Bayesian methods was used to account for overdispersion of count data [24]. 204 Unlike many other studies that used mechanistic epidemic models which was useful to describe the 205 epidemic process [39], our statistical models extended traditional relative risk models to time series 206 of count data. It should be note that this type of model likely overfit the data and collinearity 207 among lag variables also exist. Thus, having a priori hypotheses and choosing biologically relevant 208 lags are critical in building correct models and interpreting the results. Our lag models were based 209 on observed incubation period of COVID-19 and for testing pre-specific hypotheses. The principle 210 of our methods was similar to that of Institute for Health Metrics and Evaluation (IHME) [40] and 211 University of Texas-Austin models [41], all of which relied on time series analysis of count data. 212 However, we did not attempt to predict future cases. Rather, we focused on disentangling risk 213 interactions of infection across age groups, which was more important and relevant in disease 214 preventions.

Finally, during the process of re-opening the economy and society, the number of new cases may rebound, 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 epidemic. Therefore, we limited our study period to the post-peak of 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.

221

222 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 protections until the pandemic is completely over.

227 6. Patents

228 N/A

229 Supplementary Materials: N/A

Author Contributions: Dr. Xinhua Yu has full access to research data and conducted data analysis and report

231 writing.

3 of 13

- **Funding**: This research was funded by seed grant for data science from FedEx Institute of Technology at the
- 233 University of Memphis.
- 234 **Conflicts of Interest**: The authors declare no conflict of interest.
- 235

236 Abbreviations:

- 237 The following abbreviations are used in this manuscript:
- 238 SARS-CoV-2: severe acute respiratory syndrome coronavirus 2
- 239 COVID-19: coronavirus infectious disease 2019
- 240 GAM: generalized additive model
- 241 VAR: vector autoregressive regression
- 242 NB: negative binomial model
- 243 REML: restricted maximum likelihood
- 244 PMF: probability mass function
- 245

246 Appendix A:

247 I. Obtained Smoothed predicted daily cases with Generalized Additive Model (GAM)

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

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

where Y_{ij} represents the observed case counts of day *i* and group *j*, and E(Y_{ij}) is expected (predicted) value. The variable time represents day (1,...,I), b_k () represents a basis function for the *k*th term to smooth temporal trend, and $\beta_{j,kS}$ are regression coefficients for smooth term *k* and group *j*. Restricted maximum likelihood (REML) approach was used in parameter estimation. R *mgcv* package was used [22], and smooth terms were fitted using thin plate regression spline with 16 knots.

257 II. Model setups and comparisons

258 The standard Poisson distribution describes the distribution of y events occurring at a constant 259 rate of λ . The Probability mass function (PMF) is:

$$p(Y = y|\lambda) = \frac{\lambda^{y}}{y!}e^{-\lambda}, \quad \lambda > 0$$

$$E(Y) \stackrel{\text{def}}{=} \mu = \lambda = Var(Y)$$

$$Loglik = ylog(\lambda) - \lambda - log(y!)$$

In standard Poisson distribution, expected variance equals 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.

263 The generalized Poisson distribution introduces an additional scale parameter ξ [42] as quoted 264 in Hilbe JM. 2014 [26]. The PMF is:

$$p(Y = y | \lambda, \xi) = \frac{\lambda}{y!} (\lambda + \xi y)^{y-1} e^{-(\lambda + \xi y)}, \qquad \lambda > 0$$

$$E(Y) \stackrel{\text{def}}{=} \mu = \frac{\lambda}{1 - \xi}$$

$$Var(Y) = \frac{\lambda}{(1 - \xi)^3} = \frac{\mu}{(1 - \xi)^2}$$

265 Thus, $\varphi \stackrel{\text{def}}{=} \frac{1}{(1-\xi)^2}$ is the dispersion factor indicating how variance changes with the mean. 266 Therefore, if $\xi = 0$, then $\phi = 1$, corresponds to a standard Poisson (mean = variance); $0 < \xi < 1$, then ϕ

4 of 13

- 267 >1, models overdispersion (mean < variance); and $\xi < 0$, then $\phi < 1$, models under-dispersion (mean
- 268 > variance).
- 269 Reparametrize the PMF of generalized Poisson distribution with μ and ξ [24]:

$$p(Y = y | \mu, \xi) = \frac{\mu(1 - \xi)}{y!} \left(\mu - \xi(\mu - y)\right)^{y-1} e^{-(\mu - \xi(\mu - y))}$$

 $LogLik = log(\mu(1-\xi)) + (y-1)log(\mu - \xi(\mu - y)) - (\mu - \xi(\mu - y)) - log(y!)$

- 270 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:
 - $p(Y = y) = {y + r 1 \choose y} p^{y} (1 p)^{r}$ $E(Y) \stackrel{\text{def}}{=} \mu = \frac{rp}{1 - p}$ $Var(Y) = \frac{rp}{(1 - p)^{2}} = \frac{\mu}{1 - p} = \mu + \frac{\mu^{2}}{r}$
- 273 Thus, the overdispersion of Y is controlled by the shape parameter r.
- 274 Reparametrize the PMF with μ and r_{r}

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

$$p_{0}q(\mu r) = 2\log(\mu+r) + \log\left(\Gamma(y+r)\right) - \log\left(\Gamma(y+1)\right)^r$$

$$Loglik = yr(log(\mu r) - 2log(\mu + r)) + log(\Gamma(y + r)) - log(\Gamma(y + 1)) - log(\Gamma(r))$$

Where gamma function $\Gamma(x+1) = x!$ (x factorial) for an integer x, and the parameter r can be any

275 Where gamma function $\Gamma(x+1) = x!$ (x factor 276 positive real value.

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

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

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*.

284 However, although negative binomial distribution is often used to model new case counts 285 during disease outbreaks, it models only overdispersion and assumes a quadratic relationship 286 between variance and mean, while the generalized Poisson model is more flexible and assumes a 287 simpler first order association between variance and mean. Therefore, we chose to report results 288 from generalized Poisson models. Results from negative binomial models were included in the 289 appendix. In addition, it is also of note that there are extensions of negative binomial models in 290 which the association between mean and variance can be estimated from data, leading to a more 291 flexible model and also permitting the exploration of determinants of overdispersion [26].

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

$$y_{j,t} | \alpha_j, \beta_{j,k}, b_{j,t} \sim PMF(\mu_{j,t})$$

$$\mu_{j,t} = \exp\left(\alpha_j + \sum_{j=1}^J \sum_{k=1}^K \beta_{j,k} \ln(y_{j,t-k}) + b_{j,t}\right)$$

$$b_{j,t} | \Sigma \sim Mult iNormal(0, \Sigma)$$

294 Where j = 1,...,J represents age groups, t=1,...,T represents days, and k = 1,...,K represents the 295 number of lags. The PMF of Y can be either standard Poisson (λ), generalized Poisson (μ , ξ), or 296 negative binomial (μ , r) distribution.

297 The above VAR model included new case counts from previous days (lags) across age groups 298 as predictors[23], thus examining associations of the infection risks across age groups

5 of 13

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.

301 The correlation of daily counts between age groups was modeled through $b_{j,t}$ that can be 302 viewed as a random effect. The $b_{j,t}$ was assumed a multivariate normal distribution.

303 The exponential link between dynamic predictors and μ is equivalent to common relative risk

304 models in epidemiological studies, i.e., log-linear models for count data. Under this multiplicative

scale framework, the interpretation of β s are relative risks given one unit increase of predictors.

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

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

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

The models were fit with Bayesian software stan through Rstan interface [27]. A customized stan function was constructed for fitting generalized Poisson model. We employed Hamiltonian Monte Carlo with 5 Markov chains, each with 50,000 iterations plus 2000 warmups, to obtain posterior distributions of parameters. Diagnostic plots through 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, were available online (www.github.com/xinhuayu/riskinteractions/).

317 Appendix B: Additional tables

318 **Table 1:** Risk interactions in coronavirus infection across age groups based on negative binomial models, COVID-19, South Korea

Model				Lags of predictors	6	320
Outcomes	Predictors	Lag 1	Lag 2	Lag 3	Lag4	Lag 5 321
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 - 2.337)
	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 .33 5
Aged 40 - 5	9					326
-	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 - 3	9					330
-	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. 33)
	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 - 1.33)
	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 - 1 333
	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 - 1.14)
Aged 0 - 19)					335
0	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 - 1.337)
	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 - 23438

340 <

Note: * and # for p

0.05

341

342 References

343	1.	Zhu, N.; Zhang, D.; Wang, W.; Li, X.; Yang, B.; Song, J.; Zhao, X.; Huang, B.; Shi, W.; Lu, R.,
344		et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med 2020,
345		382, 727-733, doi:10.1056/NEJMoa2001017.
346	2.	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,
347		D.S.C., et al. Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med
348		2020 , 10.1056/NEJMoa2002032, doi:10.1056/NEJMoa2002032.
349	3.	Richardson, S.; Hirsch, J.S.; Narasimhan, M.; Crawford, J.M.; McGinn, T.; Davidson, K.W.;
350		and the Northwell, CR.C.; Barnaby, D.P.; Becker, L.B.; Chelico, J.D., et al. Presenting
351		Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With
352		COVID-19 in the New York City Area. <i>JAMA</i> 2020 , 10.1001/jama.2020.6775,
353		doi:10.1001/jama.2020.6775.
354	4.	Wu, Z.; McGoogan, J.M. Characteristics of and Important Lessons From the Coronavirus
355		Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72314 Cases From the
356		Chinese Center for Disease Control and Prevention. JAMA 2020, 10.1001/jama.2020.2648,
357		doi:10.1001/jama.2020.2648.
358	5.	Garg, S.; Kim, L.; Whitaker, M.; O'Halloran, A.; Cummings, C.; Holstein, R.; Prill, M.; Chai,
359		S.J.; Kirley, P.D.; Alden, N.B., et al. Hospitalization Rates and Characteristics of Patients
360		Hospitalized with Laboratory-Confirmed Coronavirus Disease 2019 - COVID-NET, 14
361		States, March 1-30, 2020. MMWR Morb Mortal Wkly Rep 2020, 69, 458-464,
362		doi:10.15585/mmwr.mm6915e3.
363	6.	Yu, X. Urban-rural inequalities during the COVID-19 pandemic among elderly people in
364		Florida, US. <i>med Rxiv</i> 2020, 10.1101/2020.05.01.20087791, 2020.2005.2001.20087791,
365		doi:10.1101/2020.05.01.20087791.
366	7.	Hay, J.A.; Haw, D.J.; Hanage, W.P.; Metcalf, C.J.E.; Mina, M.J. Implications of the Age
367		Profile of the Novel Coronavirus PNAS 2020.
368	8.	Chavan, P.; Kedia, S.; Yu, X. Physical and Functional Limitations in US Older Cancer
369		Survivors, J of Palliative Care & Medicine, 2017. 201 7, 10.4172/2165-7386.1000312,
370		doi:10.4172/2165-7386.1000312.
371	9.	Ward, B.W.; Schiller, J.S.; Goodman, R.A. Multiple chronic conditions among US adults: a
372		2012 update. Prev Chronic Dis 2014, 11, E62, doi:10.5888/pcd11.130389.
373	10.	CDC. Severe Outcomes Among Patients with Coronavirus Disease 2019 (COVID-19) $-$
374		United States, February 12–March 16, 2020. MMWR Morb Mortal Wkly Rep 2020, 343-346.
375	11.	Imai, N.; Cori, A.; Dorigatti, I.; Baguelin, M.; Donnelly, C.A.; Riley, S.; Ferguson, N.M.
376		Report 3: Transmissibility of 2019-nCoV Imperial College of London: London, OK, 2020.
377	12.	Lv, H.; Wu, N.C.; Tsang, O.TY.; Yuan, M.; Perera, R.A.P.M.; Leung, W.S.; So, R.T.Y.; Chan,
378		J.M.C.; Yip, G.K.; Chik, T.S.H., et al. Cross-reactive Antibody Response between SARS-CoV-
379		2 and SARS-CoV Infections. Cell Reports 2020, 31, 107725,
380		doi: <u>https://doi.org/10.1016/j.celrep.2020.107725</u> .
381	13.	Kim, S.; Castro, M.C. Spatiotemporal pattern of COVID-19 and government response in
382		South Korea (as of May 31, 2020). International Journal of Infectious Diseases 2020,
383		https://doi.org/10.1016/j.ijid.2020.07.004, doi:https://doi.org/10.1016/j.ijid.2020.07.004.

2 of 13

384	14.	Yu, X.; Duan, J.; Jiang, Y.; Zhang, H. Distinctive trajectories of COVID-19 epidemic by age
385		and gender: a retrospective modeling of the epidemic in South Korea. <i>International Journal of</i>
386		Infectious Diseases 2020 , https://doi.org/10.1016/j.ijid.2020.06.101,
387		doi:https://doi.org/10.1016/j.ijid.2020.06.101.
388	15.	Anderson, R.M.; Heesterbeek, H.; Klinkenberg, D.; Hollingsworth, T.D. How will country-
389		based mitigation measures influence the course of the COVID-19 epidemic? <i>Lancet</i> 2020 , 395,
390		931-934, doi:10.1016/S0140-6736(20)30567-5.
391	16.	Chowell, G.; Mizumoto, K. The COVID-19 pandemic in the USA: what might we expect?
392		Lancet 2020, 395, 1093-1094, doi:10.1016/S0140-6736(20)30743-1.
393	17.	Shim, E.; Tariq, A.; Choi, W.; Lee, Y.; Chowell, G. Transmission potential and severity of
394		COVID-19 in South Korea. Int J Infect Dis 2020, 93, 339-344, doi:10.1016/j.ijid.2020.03.031.
395	18.	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,
396		Y., et al. Coronavirus Disease Outbreak in Call Center, South Korea. <i>Emerg Infect Dis</i> 2020,
397		26, doi:10.3201/eid2608.201274.
398	19.	Kissler, S.M.; Tedijanto, C.; Goldstein, E.; Grad, Y.H.; Lipsitch, M. Projecting the
399		transmission dynamics of SARS-CoV-2 through the postpandemic period. Science 2020,
400		10.1126/science.abb5793, doi:10.1126/science.abb5793.
401	20.	Lipsitch, M.; Swerdlow, D.L.; Finelli, L. Defining the Epidemiology of Covid-19 - Studies
402		Needed. N Engl J Med 2020, 382, 1194-1196, doi:10.1056/NEJMp2002125.
403	21.	Kim, Y.J.; Seo, M.H.; Yeom, H.E. Estimating a breakpoint in the spread pattern of COVID-
404		19 in South Korea. Int J Infect Dis 2020, 10.1016/j.ijid.2020.06.055,
405		doi:10.1016/j.ijid.2020.06.055.
406	22.	Wood, S.N. Generalized Additive Models: An Introduction with R, 2nd ed.; Chapman and
407		Hall/CRC: 2017.
408	23.	Brandt, P.T.; Sandler, T. A bayesian Poisson vector autoregressive model. Political Analysis
409		2012 , 20, 23.
410	24.	Yang, Z.; Hardin, J.W.; Addy, C.L.; Vuong, Q.H. Testing approaches for overdispersion in
411		poisson regression versus the generalized poisson model. <i>Biom J</i> 2007, 49, 565-584,
412		doi:10.1002/bimj.200610340.
413	25.	Lauer, S.A.; Grantz, K.H.; Bi, Q.; Jones, F.K.; Zheng, Q.; Meredith, H.R.; Azman, A.S.; Reich,
414		N.G.; Lessler, J. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From
415		Publicly Reported Confirmed Cases: Estimation and Application. Ann Intern Med 2020,
416		10.7326/M20-0504, doi:10.7326/M20-0504.
417	26.	Hilbe, J.M. Modeling Count Data; Cambridge University Press: 2014.
418	27.	Carpenter, B.; Gelman, A.; Hoffman, M.D.; Lee, D.; Goodrich, B.; Betancourt, M.; Brubaker,
419		M.; Guo, J.; Li, P.; Riddell, A. Stan: A probabilistic programming language. Journal of
420		<i>Statistical Software</i> 2017 , <i>76</i> (1), doi:10.18637/jss.v076.i01.
421	28.	Li, Q.; Guan, X.; Wu, P.; Wang, X.; Zhou, L.; Tong, Y.; Ren, R.; Leung, K.S.M.; Lau, E.H.Y.;
422		Wong, J.Y., et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-
423		Infected Pneumonia. N Engl J Med 2020, 382, 1199-1207, doi:10.1056/NEJMoa2001316.
424	29.	Davies, N.G.; Klepac, P.; Liu, Y.; Prem, K.; Jit, M.; Eggo, R.M. Age-dependent effects in the
425		transmission and control of COVID-19 epidemics. med Rxiv 2020,
426		10.1101/2020.03.24.20043018, 2020.2003.2024.20043018, doi:10.1101/2020.03.24.20043018.

3 of 13

427	30.	Zhang, L.: Litvinova, M.: Liang, Y.: Wang, Y.: Wang, W.: Zhao, S.: Wu, O.: Merler, S.:
428		Viboud C: Vespignani A et al Changes in contact patterns shape the dynamics of the
429		COVID-19 outbreak in China Science 2020 10 1126/science abb8001
430		doi:10.1126/science.abb8001.
431	31.	Mossong, L: Hens, N.; Iit, M.; Beutels, P.; Auranen, K.; Mikolaiczyk, R.; Massari, M.;
432		Salmaso S: Tomba GS: Wallinga L et al. Social contacts and mixing patterns relevant to
433		the spread of infectious diseases <i>PLoS Med</i> 2008 5, e74, doi:10.1371/journal.pmed.0050074
434	32	Cowling B1: Ho I M: Leung GM Effectiveness of control measures during the SARS
435	52.	epidemic in Beijing: a comparison of the Rt curve and the epidemic curve. <i>Epidemial Infect</i>
436		2008 136 562-566 doi:10.1017/S0950268807008722
437	33	Gostic K : Gomez A C : Mummah R O : Kucharski A L : Llovd-Smith LO Estimated
438	00.	effectiveness of symptom and risk screening to prevent the spread of COVID-19. <i>Flife</i> 2020
439		<i>9.</i> doi:10.7554/eLife.55570.
440	34.	Pan, A.; Liu, L.; Wang, C.; Guo, H.; Hao, X.; Wang, O.; Huang, J.; He, N.; Yu, H.; Lin, X., et
441		al. Association of Public Health Interventions With the Epidemiology of the COVID-19
442		Outbreak in Wuhan, China. JAMA 2020, 10.1001/jama.2020.6130, doi:10.1001/jama.2020.6130.
443	35.	Yu, X. Modeling Return of the Epidemic: Impact of Population Structure, Asymptomatic
444		Infection, Case Importation and Personal Contacts. <i>med Rxiv</i> 2020 ,
445		10.1101/2020.04.26.20081109, 2020.2004.2026.20081109, doi:10.1101/2020.04.26.20081109.
446	36.	Bai, Y.; Yao, L.; Wei, T.; Tian, F.; Jin, D.Y.; Chen, L.; Wang, M. Presumed Asymptomatic
447		Carrier Transmission of COVID-19. JAMA 2020, 10.1001/jama.2020.2565,
448		doi:10.1001/jama.2020.2565.
449	37.	Li, C.; Ji, F.; Wang, L.; Wang, L.; Hao, J.; Dai, M.; Liu, Y.; Pan, X.; Fu, J.; Li, L., et al.
450		Asymptomatic and Human-to-Human Transmission of SARS-CoV-2 in a 2-Family Cluster,
451		Xuzhou, China. Emerg Infect Dis 2020, 26, doi:10.3201/eid2607.200718.
452	38.	Li, G.; Li, W.; He, X.; Cao, Y. Asymptomatic and Presymptomatic Infectors: Hidden Sources
453		of COVID-19 Disease. <i>Clin Infect Dis</i> 2020 , 10.1093/cid/ciaa418, doi:10.1093/cid/ciaa418.
454	39.	Kucharski, A.J.; Russell, T.W.; Diamond, C.; Liu, Y.; Edmunds, J.; Funk, S.; Eggo, R.M.;
455		Centre for Mathematical Modelling of Infectious Diseases, Cw.g. Early dynamics of
456		transmission and control of COVID-19: a mathematical modelling study. Lancet Infect Dis
457		2020 , 10.1016/S1473-3099(20)30144-4, doi:10.1016/S1473-3099(20)30144-4.
458	40.	IHME. Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilatordays and deaths by
459		US state in the next 4 months. IHME COVID-19 health service utilization forecasting team; 2020.
460	41.	Woody, S.; Garcia Tec, M.; Dahan, M.; Gaither, K.; Lachmann, M.; Fox, S.; Meyers, L.A.;
461		Scott, J.G. Projections for first-wave COVID-19 deaths across the US using social-distancing
462		measures derived from mobile phones. <i>medRxiv</i> 2020, 10.1101/2020.04.16.20068163,
463		2020.2004.2016.20068163, doi:10.1101/2020.04.16.20068163.
464	42.	Consul, P. generalized poisson distribution: properties and applications; Marcel Decker: New
465		York, 1989.