medRxiv preprint doi: https://doi.org/10.1101/2020.05.17.20105049; this version posted May 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

Xinhua Yu

Author affiliation:

Xinhua Yu, MD, PhD

Division of Epidemiology, Biostatistics & Environmental Health

School of Public Health

University of Memphis

Email: xyu2@memphis.edu

Phone: 901-678-3433

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.17.20105049; this version posted May 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Abstract

Background: the COVID-19 epidemic has incurred significant disease burden worldwide, particularly on elderly population. This study aims to explore how risks of infection interact across age groups and compared risk patterns between South Korea and US Florida state.

Methods: Daily new COVID-19 cases were scraped from online sources. A multivariate vector autoregressive model for time series 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 both South Korea and Florida, the risk of coronavirus infection among elderly people was significantly affected by other age groups. An increase of virus infection among people aged 20 -39 could double the risk of infection among elderly people. Meanwhile, an increase in virus infection among elderly people also significantly increased 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 long term.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.17.20105049; this version posted May 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Introduction

The coronavirus disease (COVID-19) is caused by the infection of a novel Severe Acute Respiratory Syndrome associated coronavirus (SARS-CoV2) [1]. Since December 2019, over 4 million people have been infected with SARS-CoV2 and over 291,000 people died (https://coronavirus.jhu.edu/map.html, accessed on May 12, 2020). Of them, elderly people and people with underlying chronic conditions suffered the heaviest disease burden [2, 3] [4]. For example, about 80% 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 aged 65 or above [5]. In the state of Florida,US, people aged 65 or above accounted for 54% of hospitalizations, and the mortality rate was 11% if infected with virus [6].

The reasons for the disproportional burden among elderly people were unclear [7]. Elderly people generally have weaker immune system than younger people due to aging, and they are also more likely to have multiple chronic conditions[8, 9]. Thus, elderly people may have severe symptoms if infected with coronavirus [10, 11]. On the other hand, elderly people may have exposed myriads of infections over their lifetime which may provide immunity against new virus infection. But it was unknown whether elderly people might have any effective immunity against SARS-CoV2.

Meanwhile, the COVID-19 pandemic is waning down in many countries since May 1, 2020, and society is gradually returning to normalcy [12, 13]. A potential rebound of new cases has been warned by many public health experts [14]. 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]. In addition, a recent study predicted a lasting post-

medRxiv preprint doi: https://doi.org/10.1101/2020.05.17.20105049; this version posted May 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

pandemic circulation of new virus worldwide[16]. Therefore, it is imperative to examine risk patterns of coronavirus infection among elderly people after the peak of epidemic.

Unfortunately, due to lack of testing kits and heterogeneous diagnosis criteria, epidemiological data on COVID-19 among different countries (and even within a country) were often noncomparable [17]. One notable exception is data from 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 [15]. Thus, depicting a complete picture of the epidemic process was possible.

In this study, we will examine how risks of coronavirus infection interact across age groups using time series analysis. We will focus on the tail period of the epidemic process to evaluate the risk of infection among elderly people during the period of society re-opening. The risk patterns will be compared between South Korea and US Florida.

Methods

Daily new COVID-19 case counts from South Korea were obtained from the website (https://github.com/jihoo-kim/Data-Science-for-COVID-19) which were scraped from the Korea Center for Disease Control website. Daily COVID-19 case counts from Florida were downloaded from the website (https://github.com/gadenbuie/covid19-florida/) which were scraped from the Florida Department of Health website. Since the peak of epidemic in South Korea ended around March 10, we limited the time series of new cases between March 10 to April 30, 2020 for South Korea. Similarly, since the epidemic in Florida passed its peak after April 1, 2020 [18], the case series were limited between April 1, 2020 to May 12, 2020 for Florida. In addition, the test kits became more available in Florida after April 1. All daily cases

medRxiv preprint doi: https://doi.org/10.1101/2020.05.17.20105049; this version posted May 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

were stratified by age groups (0-19, 20-39, 40-59, and 60 or above). To be consistent between South Korea and Florida data, those aged 60 or above were referred as elderly people.

We developed a vector autoregressive (VAR) model to examine the associations of the infection risks across age groups simultaneously [19]. 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) [20]. The model also included case counts from previous days (lags) across age groups as predictors to form a dynamic model (see Appendix 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.

$$y_{j,t} \mid \alpha_j, \beta_{j,k}, b_{j,t} \sim generalized\ Poisson(\mu_{j,t}, \xi_j)$$

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

$$b_{j,t} \mid \Sigma \sim MultiNormal(0, \Sigma)$$

Where $j = 1,\ldots,J$ represented age groups, $t=1,\ldots,T$ represented days, and $k = 1,\ldots,K$ represented the number of time lags. We reported results from five-lag models but three-lag models were also performed. Results from both models were consistent. The scale parameter ξ 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.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.17.20105049; this version posted May 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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 [21]. The coefficients βs 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 Bayesian software stan through Rstan interface (http://mc-stan.org) [22]. To keep the model simple, we assumed weakly informative priors of student t distributions for all αs and βs, and an LKJ prior with modal density around diagonals for correlations between case series (see Appendix). 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 fit and results were similar to those reported here except for wider confidence intervals (appendix Tables). The data and replicable codes were available online (*github address after blind review*).

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.

Results

In South Korea, there were 3,383 COVID-19 cases between March 10 and April 30, 2020. Of 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 or above (28.7%). On the other hand, in Florida, there were 35,279 cases during the

medRxiv preprint doi: https://doi.org/10.1101/2020.05.17.20105049; this version posted May 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

period from April 1 to May 12, 2020. Of them, 1,235 COVID-19 cases aged 0-19 (3.5%), 9,849 aged 20-39 (27.9%), 11,957 aged 40-59 (33.9%), and 12,004 aged 60 or above (34.0%).

Figure 1a and 1b presented the epidemic curves by age groups for South Korea and Florida. Although it is difficult to discern patterns between lines, all age groups generally followed similar trends over time for both regions. There was a solo spike among elderly people in South Korea around March 18.

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 doubled 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 also increased the 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.

The risk patterns were slightly different in Florida (Table 2). First, the current risk of infection among elderly people was affected by their previous infection and also by five-day prior infections among people aged 40-59, with a risk ratio of 3. Second, risks of infection among all other age groups were significantly affected by infections among elderly, with risk ratios ranging from 3 - 4. Third, risks of infection among those aged 40 – 59 appeared similar to that of elderly

medRxiv preprint doi: https://doi.org/10.1101/2020.05.17.20105049; this version posted May 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

people, while risks of infection among those aged 20-39 or 0-19 were largely unaffected by infections among other groups (except from the infection in elderly people). Finally, except for those aged 0 – 19, there was significant overdispersion in daily new case counts across age groups with observed variances was about 12-14 times of means, suggesting large variations in daily case counts in Florida.

Discussion

This was the first study to quantify risk interactions of SARS-CoV2 infection across age groups based on vector autoregressive models and compare the risk patterns between South Korea and the state of Florida in the US. We found that in both South Korea and Florida, the risk of infection among elderly people was significantly affected by other age groups. An increase in virus infection among elderly people also significantly 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 elderly people born disproportional disease burden and suffered highest mortality [4, 23]. Although virus transmission might differ between age groups [7, 24], our results highlighted the importance of implementing and enforcing effective interventions in the whole society [25-27], and the highest priority of protecting elderly people[24]. Furthermore, we showed that an increase of coronavirus infection among elderly people could increase 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[28].

medRxiv preprint doi: https://doi.org/10.1101/2020.05.17.20105049; this version posted May 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Respiratory infectious diseases often spread through personal contacts [29]. 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[30]. During the period with stay-at-home rule in the US, contacts between age groups were reduced significantly. Thus, as shown in our study, there were 2-5 lagging days in the risk interactions across age groups, especial 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. Passive community interactions such as grocery shopping might play an important role in sustaining the epidemic.

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. On the other hand, in Florida data, those cases with mild or no symptoms were more likely missed, as suggested by the smaller percent of cases among younger age groups in Florida compared with that of South Korea. We chose our study period for Florida from April 1 to May 12 to mitigate this bias because testing became more available nationally in the US after April 1. Furthermore, there were other factors such as gender, socio-economic status and neighborhood environment might also affect the risk of infection.

In addition, although we interpreted the results with action terms, they had no causative meanings. For example, younger people tended to have milder or no symptoms (i.e., subclinical cases) if infected with virus [31, 32]. Thus, it was possible that an increased number of detected cases among young people implied the existence of an increase of subclinical cases in the community who might unknowingly infected other people, including elderly people. Subclinical

medRxiv preprint doi: https://doi.org/10.1101/2020.05.17.20105049; this version posted May 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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, we compared risk interactions across age groups and between South Korea and Florida. Data from South Korea were more likely complete which would provide information about underlying epidemic mechanisms. Although different social norms and health care systems might explain some differences in risk patterns between South Korea and Florida, 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[20]. Unlike many other studies that used mechanistic epidemic models which was useful to describe the epidemic process[33], our statistical models extended traditional relative risk models to time series of count data. The principle of our methods was similar to that of IHME[34] and UT-Austin models[35], all of which relied on time series analysis of count data. However, we did not attempt to predict future cases. Rather, we focused on untangling risk interactions of infection across age groups, which was more important and relevant in disease preventions.

Finally, during the process of re-opening the economy and society, the number of new cases may rebound, and a second wave of epidemic is possible. A contentious issue was whether and how to protect high risk populations such as elderly people. 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 like elderly people should still take serious precautions during the post-epidemic period.

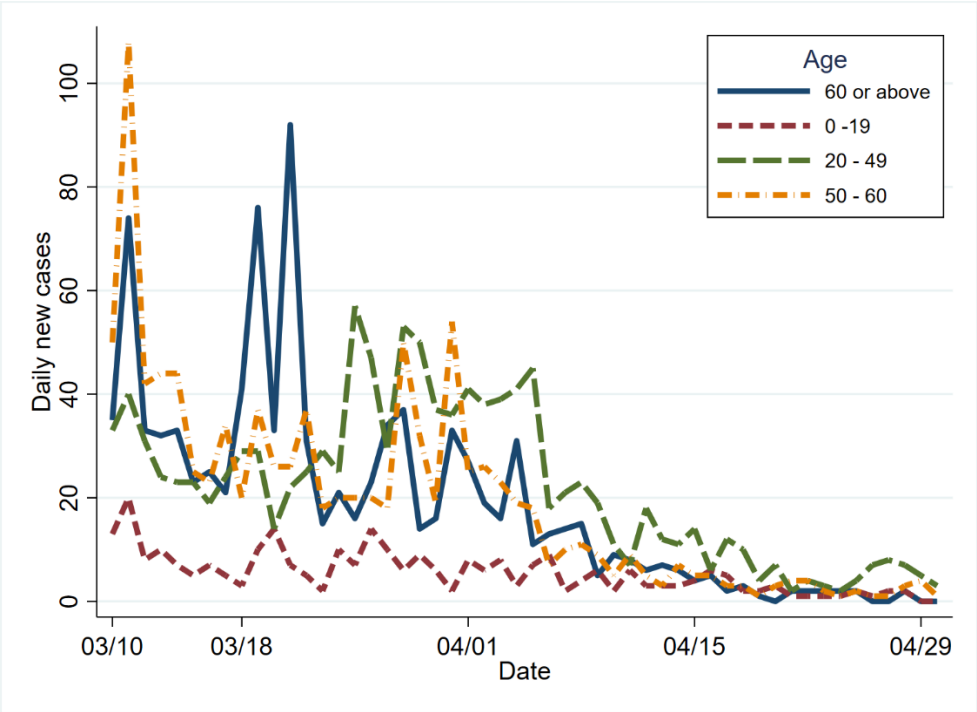
medRxiv preprint doi: https://doi.org/10.1101/2020.05.17.20105049; this version posted May 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Conclusions

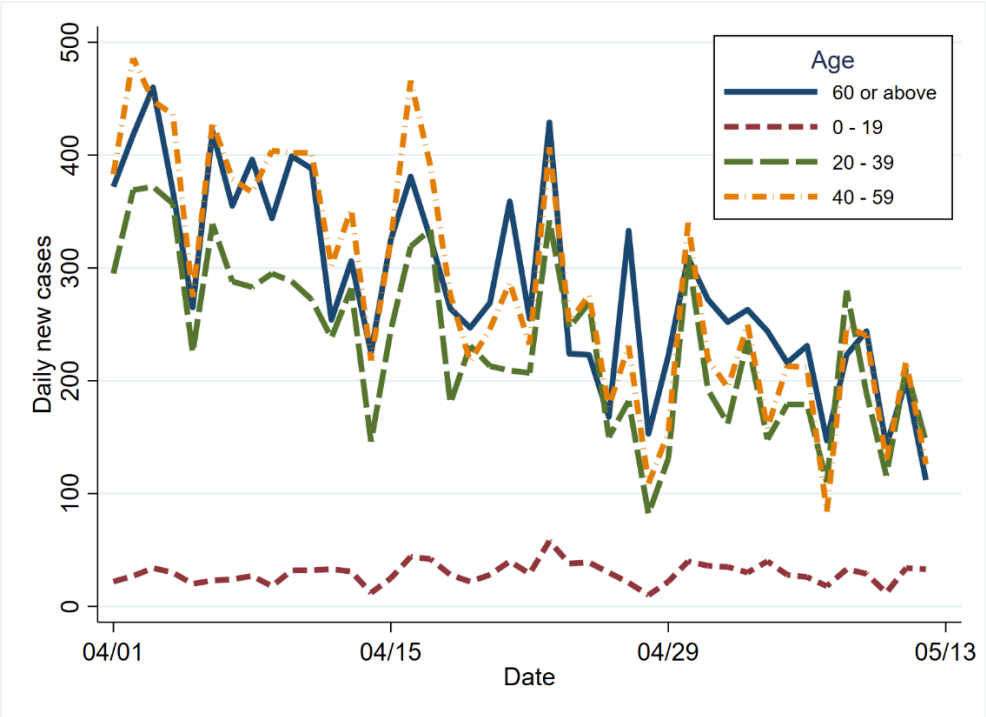
In summary, protecting elderly people from coronavirus infection could not only reduce the risk of infection among themselves but also ameliorate risks of virus infection among other age groups. Therefore, elderly people should keep on practicing social distancing and maintaining effective personal protections until the epidemic is completely over.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.17.20105049; this version posted May 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Figure 1: Epidemic curves of COVID-19 after the peak of Epidemic, South Korea (a) and US Florida (b)



(a)



(b)

Table 1: Risk interactions in coronavirus infection across age groups, COVID-19, South Korea

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 - 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 $p < 0.05$

Table 2: Risk interactions in coronavirus infection across age groups, COVID-19, US Florida

Model		Lags of predictors				
Outcomes	Predictors	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5
Aged 60 or above						
	60 or above	1.13 (0.48 - 2.64)	2.81 (1.32 - 6.19) *	1.23 (0.50 - 2.79)	1.26 (0.53 - 2.95)	0.91 (0.39 - 2.08)
	40 - 59	0.64 (0.24 - 1.68)	0.51 (0.20 - 1.30)	0.90 (0.40 - 2.04)	1.09 (0.47 - 2.58)	2.99 (1.21 - 7.19) *
	20 - 39	1.33 (0.60 - 2.90)	1.30 (0.62 - 2.68)	1.22 (0.53 - 2.64)	0.97 (0.40 - 2.19)	0.37 (0.15 - 0.91)
	0 - 19	1.01 (0.63 - 1.57)	0.64 (0.39 - 1.00)	0.87 (0.52 - 1.45)	0.89 (0.53 - 1.53)	1.02 (0.56 - 1.85)
Aged 40 - 59						
	60 or above	0.67 (0.28 - 1.60)	3.32 (1.38 - 7.61) *	1.35 (0.56 - 3.14)	1.33 (0.52 - 3.30)	0.78 (0.32 - 1.82)
	40 - 59	0.80 (0.30 - 2.17)	0.61 (0.24 - 1.66)	0.99 (0.40 - 2.51)	0.90 (0.38 - 2.09)	2.94 (1.10 - 7.49) *
	20 - 39	1.56 (0.69 - 3.54)	1.23 (0.57 - 2.67)	1.47 (0.63 - 3.49)	0.75 (0.30 - 1.92)	0.46 (0.17 - 1.18)
	0 - 19	1.15 (0.71 - 1.90)	0.48 (0.30 - 0.79)	0.72 (0.42 - 1.24)	1.15 (0.66 - 2.04)	0.99 (0.53 - 1.88)
Aged 20 - 39						
	60 or above	0.83 (0.35 - 1.93)	2.68 (1.15 - 6.10) *	1.75 (0.74 - 3.99)	1.47 (0.60 - 3.60)	0.79 (0.34 - 1.84)
	40 - 59	0.64 (0.25 - 1.71)	0.89 (0.35 - 2.37)	0.91 (0.37 - 2.26)	0.72 (0.32 - 1.76)	2.12 (0.78 - 5.44)
	20 - 39	1.47 (0.66 - 3.35)	1.10 (0.50 - 2.39)	1.42 (0.58 - 3.37)	0.87 (0.34 - 2.06)	0.45 (0.16 - 1.19)
	0 - 19	1.18 (0.71 - 1.88)	0.51 (0.32 - 0.81) #	0.67 (0.38 - 1.17)	1.35 (0.78 - 2.41)	1.11 (0.61 - 2.06)
Aged 0 - 19						
	60 or above	0.69 (0.27 - 1.74)	4.03 (1.67 - 9.17) *	1.23 (0.45 - 3.15)	1.01 (0.38 - 2.76)	1.27 (0.48 - 3.30)
	40 - 59	0.74 (0.25 - 2.21)	1.15 (0.41 - 3.29)	1.42 (0.53 - 4.00)	0.46 (0.18 - 1.27)	1.84 (0.66 - 4.80)
	20 - 39	1.67 (0.64 - 4.37)	0.56 (0.23 - 1.36)	0.64 (0.23 - 1.74)	1.38 (0.55 - 3.46)	0.36 (0.13 - 0.95) #
	0 - 19	1.56 (0.88 - 2.69)	0.56 (0.33 - 0.95) #	0.96 (0.52 - 1.74)	1.47 (0.81 - 2.63)	1.14 (0.57 - 2.32)

Note: * and # for $p < 0.05$

medRxiv preprint doi: https://doi.org/10.1101/2020.05.17.20105049; this version posted May 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Appendix A: Model setups and comparisons

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

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

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

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

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.

The generalized Poisson distribution introduces an additional scale parameter ξ [36] as quoted in Hilbe JM. 2014 [21]. The PMF is:

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

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

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

Thus, $\varphi \stackrel{\text{def}}{=} \frac{1}{(1-\xi)^2}$ is the dispersion factor indicating how variance changes with the mean.

Therefore, if ξ = 0, then φ = 1, corresponds to a standard Poisson (mean = variance); 0 < ξ < 1, then φ > 1, models overdispersion (mean < variance); and ξ < 0, then φ < 1, models under-dispersion (mean > variance).

Reparametrize the PMF of generalized Poisson distribution with μ and ξ[20]:

medRxiv preprint doi: https://doi.org/10.1101/2020.05.17.20105049; this version posted May 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

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

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*:

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

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

$$Var(Y) = \frac{rp}{(1-p)^2} = \frac{\mu}{1-p} = \mu + \frac{\mu^2}{r}$$

Thus, the overdispersion of Y is controlled by the shape parameter *r*.

Reparametrize the PMF with μ and *r*,

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

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

Where gamma function Γ(*x*+1) = *x*! (*x* factorial) 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 ~ 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,

medRxiv preprint doi: https://doi.org/10.1101/2020.05.17.20105049; this version posted May 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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:

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

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*.

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 were included in the appendix. 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[21].

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

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

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

$$b_{j,t} \mid \Sigma \sim MultiNormal(0, \Sigma)$$

medRxiv preprint doi: https://doi.org/10.1101/2020.05.17.20105049; this version posted May 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

The above VAR model included new case counts from previous days (lags) across age groups as predictors[19], 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 μ is equivalent to common relative risk models in epidemiological studies, i.e., log-linear models for count data. Under this multiplicative scale framework, the interpretation of βs are relative risks given one unit increase of predictors.

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

$$\alpha_j \sim t(df = 5, location = 0, scale = 2.5)$$

$$\beta_{j,k} \sim t(df = 5, location = 0, scale = 2.5)$$

$$\xi_j \sim normal(0,0.3), and\ -0.9 \leq \xi_j \leq 0.9$$

$$\frac{1}{\sqrt{r}} \stackrel{\text{def}}{=} \varphi \sim \text{positive half} - \text{normal}(0,1)$$

$$\Sigma = D'RD,$$

$$where\ standard\ deviation\ D = diag(sd_j), sd_j \sim cauchy(0,5)$$

$$and\ R = correlation\ matrix, R \sim LKJ(2)$$

medRxiv preprint doi: https://doi.org/10.1101/2020.05.17.20105049; this version posted May 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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 Rstan interface [22]. 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 (*github address after blind review*).

Appendix B: Output from additional models

Table 1: Risk interactions across age groups in coronavirus infection assuming case counts as negative binomial distribution, South Korea

Model		**Lags of predictors**				
Outcomes	**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)

Table 2: Risk interactions across age groups in coronavirus infection assuming case counts as negative binomial distribution, Florida

Model		Lags of predictors				
Outcomes	Predictors	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5
Aged 60 or above						
	60 or above	1.00 (0.39 - 2.53)	2.52 (1.05 - 6.04) *	1.08 (0.43 - 2.81)	1.48 (0.55 - 4.10)	1.00 (0.38 - 2.51)
	40 - 59	0.73 (0.24 - 2.26)	0.54 (0.19 - 1.54)	0.97 (0.37 - 2.62)	1.03 (0.40 - 2.68)	2.60 (0.97 - 6.79)
	20 - 39	1.01 (0.40 - 2.64)	1.21 (0.50 - 2.98)	1.31 (0.53 - 3.16)	1.00 (0.40 - 2.53)	0.43 (0.17 - 1.18)
	0 - 19	1.17 (0.70 - 1.95)	0.68 (0.39 - 1.18)	0.79 (0.43 - 1.46)	0.87 (0.45 - 1.69)	1.00 (0.50 - 2.06)
Aged 40 - 59						
	60 or above	0.64 (0.24 - 1.67)	3.11 (1.28 - 7.67) *	1.16 (0.46 - 3.16)	1.48 (0.53 - 4.11)	0.84 (0.32 - 2.27)
	40 - 59	0.91 (0.29 - 2.82)	0.59 (0.20 - 1.75)	1.20 (0.46 - 3.19)	0.83 (0.31 - 2.24)	2.58 (0.94 - 7.03)
	20 - 39	1.30 (0.51 - 3.39)	1.23 (0.49 - 2.95)	1.39 (0.55 - 3.42)	0.76 (0.31 - 1.91)	0.55 (0.20 - 1.54)
	0 - 19	1.23 (0.74 - 2.04)	0.52 (0.30 - 0.93)	0.68 (0.36 - 1.27)	1.16 (0.61 - 2.21)	0.94 (0.44 - 1.96)
Aged 20 - 39						
	60 or above	0.80 (0.30 - 2.08)	2.78 (1.16 - 6.62) *	1.63 (0.65 - 4.18)	1.63 (0.60 - 4.46)	0.82 (0.32 - 2.22)
	40 - 59	0.77 (0.25 - 2.34)	0.84 (0.29 - 2.37)	1.08 (0.42 - 2.77)	0.66 (0.26 - 1.71)	1.87 (0.69 - 5.10)
	20 - 39	1.27 (0.51 - 3.23)	1.09 (0.44 - 2.68)	1.23 (0.52 - 2.97)	0.83 (0.34 - 2.05)	0.53 (0.20 - 1.46)
	0 - 19	1.19 (0.72 - 2.01)	0.52 (0.30 - 0.92) #	0.69 (0.37 - 1.28)	1.36 (0.71 - 2.58)	1.05 (0.50 - 2.14)
Aged 0 - 19						
	60 or above	0.70 (0.24 - 2.06)	4.08 (1.56 - 10.60) *	1.16 (0.41 - 3.28)	1.14 (0.39 - 3.51)	1.26 (0.46 - 3.50)
	40 - 59	0.78 (0.23 - 2.64)	1.15 (0.37 - 3.63)	1.52 (0.53 - 4.52)	0.50 (0.19 - 1.41)	1.66 (0.55 - 4.72)
	20 - 39	1.52 (0.54 - 4.18)	0.57 (0.21 - 1.47)	0.63 (0.24 - 1.69)	1.19 (0.43 - 3.08)	0.37 (0.12 - 1.11)
	0 - 19	1.53 (0.89 - 2.69)	0.53 (0.29 - 0.97) #	1.00 (0.51 - 1.93)	1.42 (0.72 - 2.78)	1.22 (0.56 - 2.75)

medRxiv preprint doi: https://doi.org/10.1101/2020.05.17.20105049; this version posted May 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

References:

1. Zhu, N., et al., A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med*, 2020. **382**(8): p. 727-733.
2. Wu, Z. and J.M. McGoogan, 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.
3. Guan, W.J., et al., Clinical Characteristics of Coronavirus Disease 2019 in China. *N Engl J Med*, 2020.
4. Richardson, S., et al., Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. *JAMA*, 2020.
5. Garg, S., 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**(15): p. 458-464.
6. Yu, X., Did elderly people living in small towns or rural areas suffer heavier disease burden during the COVID-19 epidemic? *medRxiv*, 2020: p. 2020.05.01.20087791.
7. Hay, J.A., et al., Implications of the Age Profile of the Novel Coronavirus *PNAS*, 2020.
8. Ward, B.W., J.S. Schiller, and R.A. Goodman, Multiple chronic conditions among US adults: a 2012 update. *Prev Chronic Dis*, 2014. **11**: p. E62.
9. Chavan, P., S. Kedia, and X. Yu, Physical and Functional Limitations in US Older Cancer Survivors, J of Palliative Care & Medicine, 2017. 2017.
10. Imai, N., et al., *Report 3: Transmissibility of 2019-nCoV* 2020, Imperial College of London: London, OK.
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(69): p. 343-346.
12. Ferguson, N., et al., *Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand* 2020, Imperial College: UK.
13. Anderson, R.M., et al., How will country-based mitigation measures influence the course of the COVID-19 epidemic? *Lancet*, 2020. **395**(10228): p. 931-934.
14. Chowell, G. and K. Mizumoto, The COVID-19 pandemic in the USA: what might we expect? *Lancet*, 2020. **395**(10230): p. 1093-1094.
15. Shim, E., et al., Transmission potential and severity of COVID-19 in South Korea. *Int J Infect Dis*, 2020. **93**: p. 339-344.
16. Kissler, S.M., et al., Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science*, 2020.
17. Lipsitch, M., D.L. Swerdlow, and L. Finelli, Defining the Epidemiology of Covid-19 - Studies Needed. *N Engl J Med*, 2020. **382**(13): p. 1194-1196.
18. Yu, X., Impact of mitigating interventions and temperature on the instantaneous reproduction number in the COVID-19 epidemic among 30 US metropolitan areas. *MEDRXIV/2020/081083*, 2020.
19. Brandt, P.T. and T. Sandler, A bayesian Poisson vector autoregressive model. *Political Analysis*, 2012. **20**: p. 23.
20. Yang, Z., et al., Testing approaches for overdispersion in poisson regression versus the generalized poisson model. *Biom J*, 2007. **49**(4): p. 565-84.
21. Hilbe, J.M., Modeling Count Data. 2014*: Cambridge University Press*.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.17.20105049; this version posted May 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

22. Carpenter, B., et al., Stan: A probabilistic programming language. *Journal of Statistical Software* 2017. **76(1)**.
23. Li, Q., et al., Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *N Engl J Med*, 2020. **382**(13): p. 1199-1207.
24. Davies, N.G., et al., Age-dependent effects in the transmission and control of COVID-19 epidemics. *medRxiv*, 2020: p. 2020.03.24.20043018.
25. Pan, A., et al., Association of Public Health Interventions With the Epidemiology of the COVID-19 Outbreak in Wuhan, China. *JAMA*, 2020.
26. Gostic, K., et al., Estimated effectiveness of symptom and risk screening to prevent the spread of COVID-19. *Elife*, 2020. **9**.
27. Cowling, B.J., et al., Impact assessment of non-pharmaceutical interventions against coronavirus disease 2019 and influenza in Hong Kong: an observational study. *Lancet Public Health*, 2020.
28. Yu, X., Modeling Return of the Epidemic: Impact of Population Structure, Asymptomatic Infection, Case Importation and Personal Contacts. *medRxiv*, 2020: p. 2020.04.26.20081109.
29. Zhang, J., et al., Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. *Science*, 2020.
30. Mossong, J., et al., Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med*, 2008. **5**(3): p. e74.
31. Li, C., et al., Asymptomatic and Human-to-Human Transmission of SARS-CoV-2 in a 2-Family Cluster, Xuzhou, China. *Emerg Infect Dis*, 2020. **26**(7).
32. Li, G., et al., Asymptomatic and Presymptomatic Infectors: Hidden Sources of COVID-19 Disease. *Clin Infect Dis*, 2020.
33. Kucharski, A.J., et al., Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect Dis*, 2020.
34. IHME, *Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilatordays and deaths by US state in the next 4 months. IHME COVID-19 health service utilization forecasting team*. 2020.
35. Woody, S., et al., Projections for first-wave COVID-19 deaths across the US using social-distancing measures derived from mobile phones. *medRxiv*, 2020: p. 2020.04.16.20068163.
36. Consul, P., generalized poisson distribution: properties and applications. 1989, *New York: Marcel Decker*.