A Data Driven Study on the Variant of Covid-19 in Hong Kong

Yongmei Ding[*] and Lingxiao Xiang

College of Science, Wuhan University of Science and Technology, Huangjiahu West Road, 430065 Wuhan, China

Abstract: The new wave of COVID-19 in Hong Kong, China was overwhelming again by "dynamic zero" strategy and non-pharmaceutical interventions (DZ-NPIs), which makes a time challenge to control the variant of this epidemic. We describe the variant of Covid-19 in Kong Hong to the infected proportion of the population, cumulative confirmed cases, cumulative deaths and current hospitalizations by age group via statistical measure firstly, then establish time series model for fitting the accumulative confirmed cases, further to predict the trend for searching out possible turning time-points. Non-linear regression model is created to feature the deaths series, then we figure out the parameters and educe the controlling condition for this epidemic. We expect our data-driven modeling process providing some insights to the controlling strategy for the new wave of the Covid-19 variant in Hong Kong, even in the mainland of China.

INTRODUCTION

COVID-19, which broke out in late 2019, has swept the world and seriously threatened human health. It has become a major challenge of health and security in global [1]. By the end of 2021, as most of the world was moving towards various "living with covid" strategies, Hong Kong has always had a good reputation for being dynamic and flexible. But by the time the omicron variant hit, the strict contact tracing and isolation measures adopted up to that point were no longer able to control the spread of such a contagious variant [2].

As of July 10, 2022, there were 2773 newly confirmed cases daily (includes the number of asymptomatic infections and re-positive patients) and 9,410 cumulative deaths in Hong Kong, with the fatality rate of 0.74%, much higher than the 0.1% fatality rate of influenza. At the peak of the rapid spread of the epidemic in Hong Kong, 19 new cases of BA.2.12.1 Omicron variant infection were identified with 22 new cases of BA.4 or BA.5. The proportion of BA.4 cases in the local daily new cases increased gradually, from 0 percent on June 21 to 1 percent of the population recently.

Multiple waves of overall death in COVID-19 are increasing in Hong Kong, while the total unemployment rate has increased to 5.4%(from February to April 2022) and the underemployment rate increased merely by 0.7 percentage points to 3.8% [3]. When Arisina Ma and Jane Parry assert Hong Kong is in the dark times [2], and Taylor L thinks Hong Kong becoming the place of the worst covid-19 death rate in the world [4], researchers surge their efforts to figure out the feature of these epidemic waves. Two multiple linear regression models were compared to reflect the infection risk [5], while a time series modeling for the

association between unemployment and suicide provides the estimations of suicide risk among unemployed people during different economic situations under the background of this epidemic [6]. The exported cases in Hong Kong has been tested and found that the infection risk was happened on travers' trips, rather than locally when the pandemic in the departure city is well controlled; The impact of multiple non-pharmaceutical interventions without lockdown was also examined through incorporated variation in efficiency of contact tracing [7], provide a better understanding of what caused rapid growth of the outbreak; Health belief model and theory of planned behavior has been established to find out the reason of the low vaccine rat for children and its associated factors, guide the way of how it was suppressed [8].

Evidently, an exact modeling from the real data for epidemic can offer valuable insights to evaluate and intervene the possible effect [9-11]. Here we chose the observational data to describe the feature of the proportion of the population, cumulative confirmed cases, cumulative deaths and current hospitalizations by age group through statistical measure，also the hospitalization, severe cases, cumulative death among different vaccinations firstly, then model and predict the confirmed cases for search out the exact point of non-increasing spot in the near future via time series model. A new non-linear regression model is created to feature the death data, and figure out the turning point of decreasing the death. We deduce the reproductive number out for this new variant of COVID-19, and the rate of death will drop to 0.01 after October 9, 2022.

MATERIALS

Data are retrieved from public sources, primarily the Hong Kong Department of Health with the link https://www.coronavirus.gov.hk/sim/index.html. According to the availability, the time span of data is from Feb.1 to July 10, 2022, and there are two tips:

*Address correspondence to this author at the College of Science, Wuhan University of Science and Technology, Huangjiahu West Road, 430065 Wuhan, China; E-mail: dingyongmei@wust.edu.cn

Firstly, the vaccination count is based on the number of people who have received at least one dose of vaccine in Hong Kong or elsewhere for 14 days as of the current date, and the new count is classified as recovered after 28 days from the date of COVID-19 infection; Secondly, Number of first dose vaccinations records the total number of persons vaccinated with the first dose in Hong Kong, which includes non-resident cases.

METHODS

ARIMA Model

The difference operation has a powerful deterministic information extraction capability, and many non-stationary series will show the properties of a stationary series after differencing, at which point we call this non-stationary series a differential stationary series. The ARIMA model can be used to fit the differential smooth series. A model with the following structure is called a autoregressive integrated moving average model, which is simply called $ARIMA(p, d, q)$:

$$\begin{cases} \Phi(B)\nabla^d x_t = \Theta(B)\varepsilon_t \\ E(\varepsilon_t) = 0, Var(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t\varepsilon_s) = 0, t \neq s \\ E(x_s\varepsilon_t) = 0, \forall s < t \end{cases}$$

where

$\nabla^d = (1 - B)^d$, where B is a delay operator, and $(1 - B)x_t = x_t - x_{t-1}$;

$\Phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ is the Autoregressive coefficient polynomial of a Stationary Reversible ARMA(p, q) model;

$\Theta(B) = 1 - \phi\theta_1 B - \cdots - \phi\theta_p B^q$ is the moving average coefficient polynomial of a Stationary Reversible ARMA(p, q) model;

$\{\varepsilon_t\}$ is a zero mean white noise sequence.

Non-Linear Model

The following is a common fitted regression model for drug response:

$$y_t = \alpha - \frac{\alpha}{1 + (\frac{x_t}{\beta})^\gamma} + \varepsilon_t$$

where

the independent variable $x_t = t, t \in \mathbb{N}^*$, is the amount of the agent.

the dependent variable y_t is the degree of drug response.

parameter α is the maximum degree of response.

parameter γ is a measure of the intensity of the overall drug reaction process

Parameter β is the value corresponding to the point x_t in the drug reaction process with the fastest reaction rate.

The procedure for calculating the rate of change is given below:

Taking the derivative of the above equation gives:

$$\text{Growth Rate} = \frac{dy}{dx} = \alpha \cdot \frac{1}{\left(1 + \left(\frac{x}{\beta}\right)^\gamma\right)^2} \cdot \frac{\gamma - 1}{\beta^\gamma} \cdot x^{\gamma-1}$$
$$= \frac{\alpha(\gamma - 1)\beta^\gamma}{(\beta^\gamma + x^\gamma)^2} x^{\gamma-1}$$

which is:

$$\text{Growth Rate} = \frac{\alpha(\gamma - 1)\beta^\gamma}{(\beta^\gamma + x^\gamma)^2} x^{\gamma-1}$$

RESULTS

Baseline Scenario

Age Structure Affected by the Epidemic

Here we plot stacked bar chart of population, cumulative diagnoses, cumulative deaths, and current hospitalizations by age group, as shown in Figure **1**. The age structure of Hong Kong's population is similar to that of confirmed cases, which implies there is no significant correlation between age and infectious cases. The elderly with age over 60 years, who are the smaller proportion of confirmed cases, account for more than half of cumulative deaths and current hospitalizations. In contrast, young people, such as those under 40 years old, account for an extremely smaller proportion of deaths.
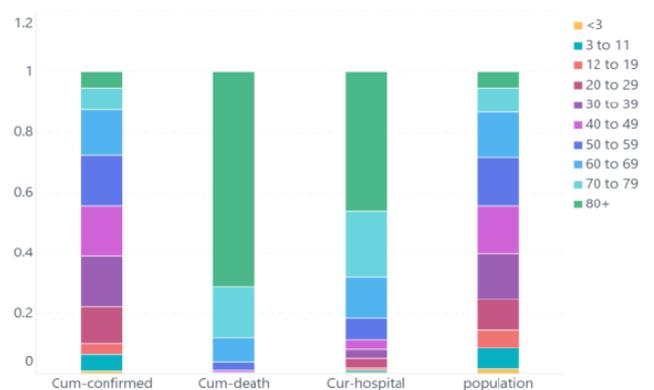

Figure 1: Chart of age affected by COVID-19.

We then figure out the severe cases in different age structure (Figure **2**), which shows no more than five severe cases in any age group. it was found that severe patients were mostly in the elderly, while no severe patients under 30 years old.
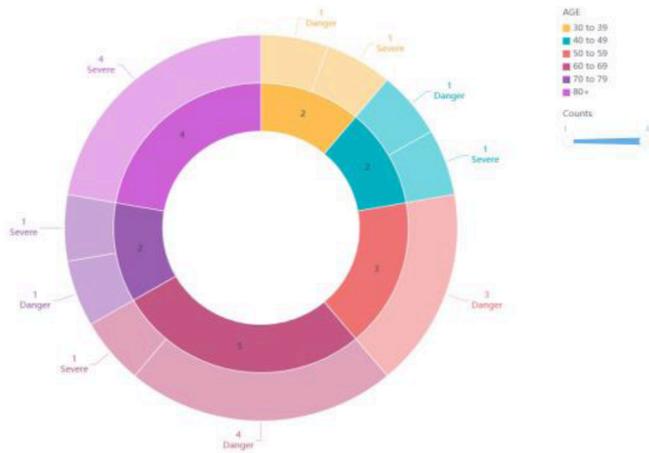
Figure 2: Diagram of age and disease degree.

Vaccination

The bar chart of composition of different vaccinated populations in current hospitalizations, current severe cases, cumulative deaths, and cumulative diagnoses shows the high rate of one dose vaccine (Figure **3**), The unvaccinated population accounts for the largest proportion of current hospitalizations, severe cases, and cumulative deaths. More than 40% of the population have received three doses of vaccine, and less than 1% of cumulative deaths have occurred in this group. It is evident that lower vaccine, higher risk of infection, even higher mortality rate.
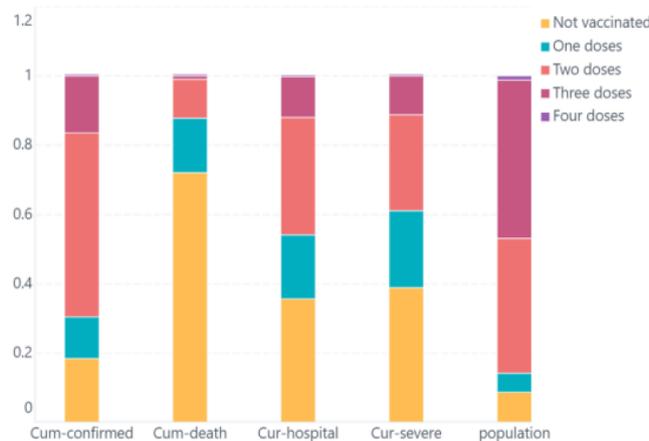

Figure 3: Vaccine.

Death

Almost 70% of the people who died were unvaccinated. The more times of inoculation, the lower the case fatality rate. The elderly endures the highest risk in the population with different times of inoculation, while the fatality rate of the young is lower (Figure **4**).

Hospitalization

The number of hospitalizations has shown a gradual upward trend since 1 June, 2022, and the majority is the elderly. The immediate cause was the resurgence of the epidemic in Hong Kong on May 30,

when 275 new cases were confirmed. As for the time series changes of the current severe patients (Figure **5**), severe patients are no more than 10, and most of them are the elderly.

Figure 4: The correlation of vaccination, age and death.

Impact DZ-NPIs Strategy on Accumulative Confirmed Cases

In order to figure out the trend of confirmed cases and cease point of stopping deaths increasing, then search out the point of flattening the curve, we chose ARIMA model to fit the accumulative confirmed cases and predict the possible trend.

Time series graph of accumulative confirmed cases is provided in Figure **6**. The fifth wave peaked on March 4, 2022, with 8,764 COVID-19 cases per million population [3], then declined to 2,716.0 by March 21, 2022. On April 21, the Hong Kong government announced that it would relax most of its control measures, opening public places including cultural, educational, sports and entertainment institutions for the first time in four months, while restaurants in the city resumed serving dinner. A small rebound happened at the end of this month, then keep smoothly until July 7, the reported news cases are up to 3028 including asymptomatic infections and re-positive patients.

We test the stationarity for the series of accumulative confirmed cases (noted it as variable "y"), p-value shows accepting the null hypothesis under the significance level 0.05, which indicates non-stationary to the original series, but accepting the alternative hypothesis after one-order difference calculation. The statistical value of AIC shows less information lost in the differential calculation. Here we use "y.difference" to note the differential series.

The coefficients of autocorrelation and partial autocorrelation are provided to determine the optimal order for classical time series model of ARMA (p, q).
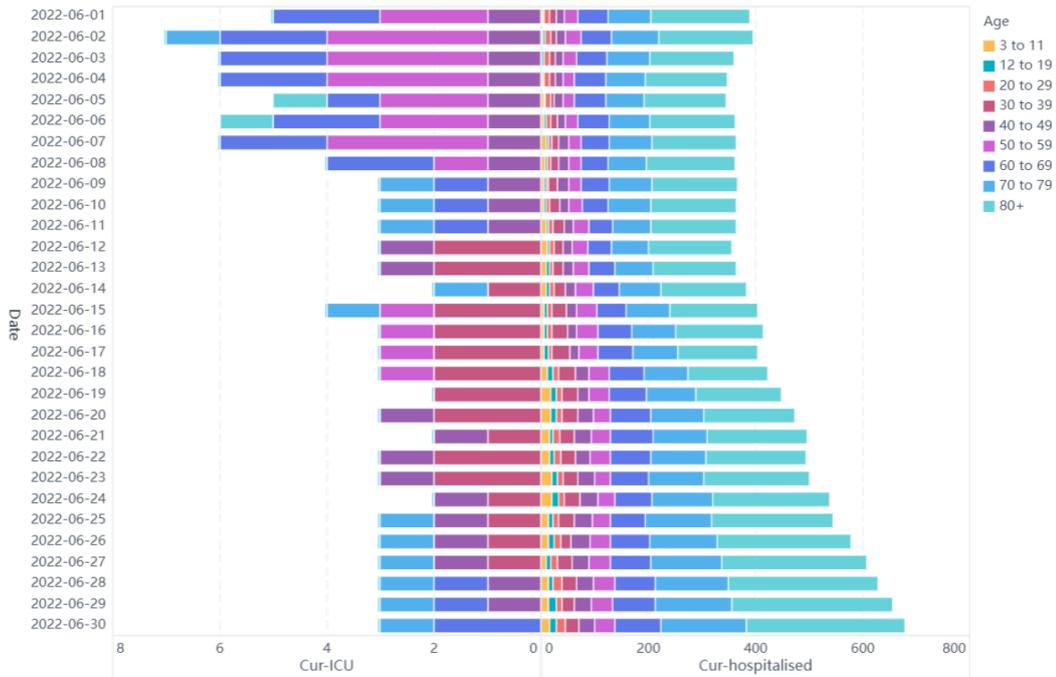
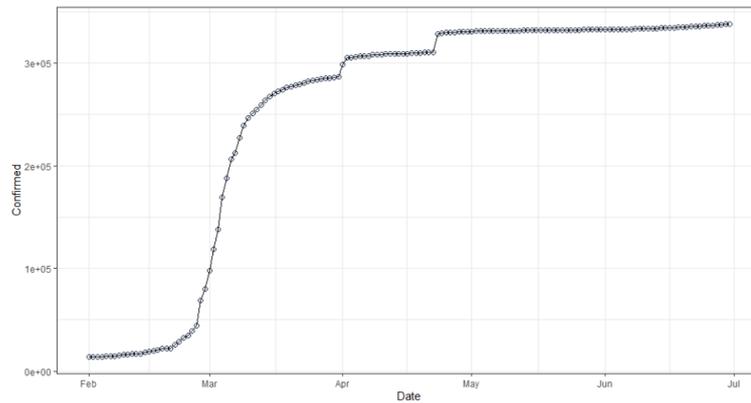**Figure 5:** Hospitalization at different age structure.

Figure 6: Time series for accumulative confirmed cases.

Table 1:   ADF Test for the Accumulative Confirmed Cases

Data series	Order of difference	T statistic	P value	AIC
Accumulative confirmed cases	0	-2.604	0.092*	2567.676
	1	-3.261	0.017**	2558.922

Here p is the order of auto-regression (AR) part and q is the order of moving-average part.

It can be shown from Figure **7** that the coefficients of auto-correlation function (ACF) are trailing (7a), while the coefficients of partial auto-correlation function (PACF) are truncated two steps (7b), which provide the optimal p and q for the auto-regression integrated moving-average model(ARIMA(p,d,q) with p and q respectively. Here d is the order of differential calculation. The optimal model is ARIMA (2,1,0) with 0.999 of R².

According to the Table **2**, the fitted model is shown in (1). Here y_t is the series of accumulative confirmed cases, and t is the time point.

$$\nabla y_t = 2176.081 + 0.463 \nabla y_{t-1} + 0.374 \nabla y_{t-2} \tag{1}$$

where $\nabla y_t = (1 - B)y_t = y_t - y_{t-1}$, and B is the lag operator.

We then test the residual through Q statistic, which shows no correlation on the residual of lag sixth(Q6) and twelfth(Q12) order at the significant level of 0.05,
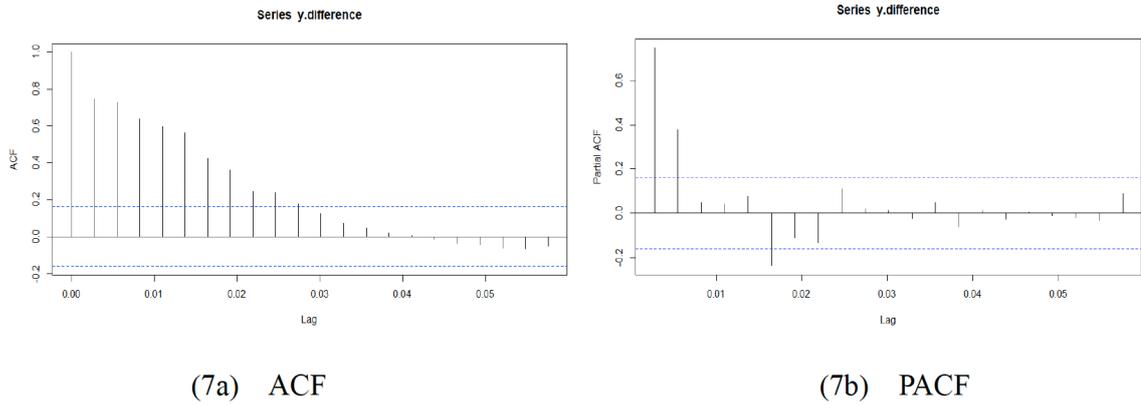
| (7a) ACF | (7b) PACF |

Figure 7: correlation coefficients for 1-order differential series.

Table 2: Parameters Estimation for ARIMA(2,1,O)

parameter	estimation	S.E	t	p>|t|
constant	2176.081	1520.692	1.431	0.152
φ_1	0.463	0.075	6.153	0
φ_2	0.374	0.075	4.972	0

as both P value are bigger than 0.05 with 0.839 and 0.097 respectively (Table **3**). The accumulative confirmed cases for future 30 days is predicted and shown in Figure **8**. The forecast shows the cumulative number of confirmed cases will continue to grow steadily in July, reaching 39,324 as of July 30, 2022.

Impact DZ-NPIs Strategy on Accumulative Number of Deaths

As far as we're concerned, the deaths number is the most series and reliable data, which shows the S-like shape (Figure **9**). Considering the death caused by the variant of the COVID-19, the number of deaths would accelerate at the initial stage and then slow down. Suppose the maximum number of cumulative deaths is a definite number, denoted it as α; the inflection point of the epidemic is β, and the reproductive number is γ, we create a non-linear model to feature this character with the expression as shown in (2):

$$y_t = \alpha - \frac{\alpha}{1+\left(\frac{x_t(t)}{\beta}\right)^\gamma} + \varepsilon_t \tag{2}$$

Table 3: Residual Test

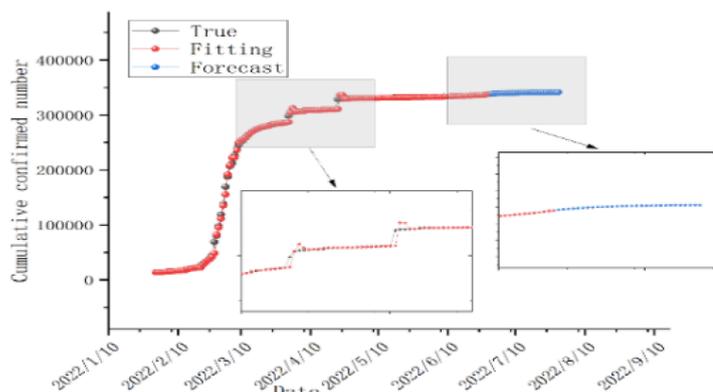| item	index	estimation	item	index	estimation
Total population	N	150		Df Residuals	146
Q statistc for residual serires	Q6(p-value)	0.042(0.839)	Information Criteria	AIC	2833.93
	Q12	10.725(0.097*)		BIC	2845.946
	Q18	17.382(0.136)	goodness of fit	R²	0.999

Figure 8: The prediction for cumulative number of confirmed cases.
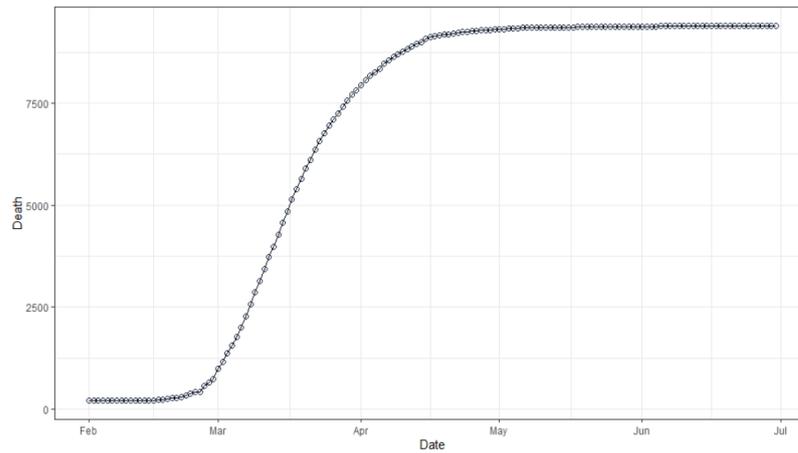
Figure 9: Cumulative death.

Table 4:   Parameter Estimation for Accumulative Death Series

Formula: y ~ α - α/(1 + (x/β)^γ)				
Parameters:				
	Estimate	**Std. Error**	**t value**	**Pr(>|t|)**
α	9460	12.00	791.1	<2e-16 ***
β	43.52	0.07	588	<2e-16 ***
γ	5.30	0.04	122	<2e-16 ***

Residual standard error: 90.44 on 147 degrees of freedom.
Number of iterations to convergence: 4.
Achieved convergence tolerance: 5.499e-06.

where $x_t = t, t \in \mathbb{N}^*$ is the time, which means the *t* day from Feb.1,2022, and y_t is the accumulative death on day *t*. Here we solve it through iteration. Suppose the initial values for the three parameters are 10000, 100, and 100. The solution is convergent after iterating 4 times, as shown in Table **4**.

All three parameters pass the significance test, and the goodness of fit R^2=0.999, indicating that the nonlinear regression fitting effect is great. The fitted model is shown in (3). We deduce out the upper limit of accumulative death is 9460, and the reflection point happened 43 days later from the first day of measure time, which happened in March 11,2022 (as it's 28 days in Feb.) with matching the reality exactly. The reproductive number for this new variant of COVID-19 is about 5.3, much higher than the corresponding number of the original virus [10]. Then we predict according the fitted model, which shows the cumulative death will increase at a slower rate in July and will reach 9,454 as of 30 July 2022 (Figure **10**).

$$\hat{y}(t) = 9460 - \frac{9460}{1+\left(\frac{x(t)}{43.52}\right)^{5.304}} \tag{3}$$

In order to search out the exact time point of normal level for death, we derive the derivation through the non-linear model, the rate of increase in the number of deaths is described in (4).

$$\frac{dy}{dx} = \frac{\alpha(\gamma-1)\beta^{\gamma}}{(\beta^{\gamma}+x^{\gamma})^2}x^{\gamma-1} \tag{4}$$
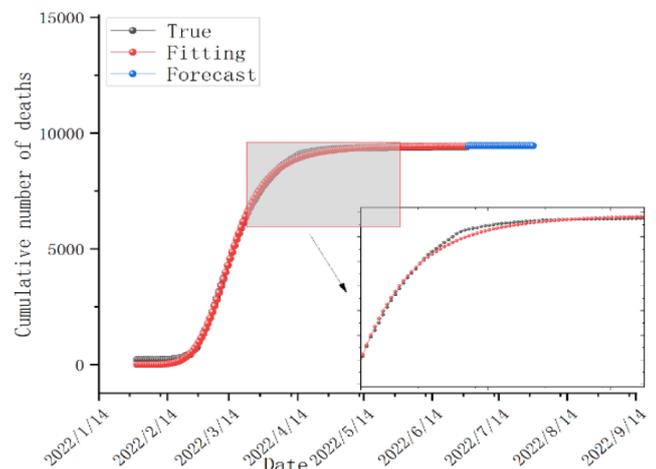

Figure 10: The prediction of accumulative deaths.

The curve of speed rate from our model shows the inflection point at the rate of increasing for cumulative deaths occurred on March 13, 2022 (Figure **11**); Actually, it's March 11 as February has only 28 days. The rate of death will drop to 0.01 after October 9, 2022(the predict cumulative death will reached 9,459, as the upper number of deaths would be 9460), which means there will be no new deaths if no outside import cases rash into Hong Kong.

DISCUSSION

We did a descriptive analysis of different age groups with different vaccination status, and then modelled and predicted the cumulative number of confirmed diagnoses through time series model, and built a new non-linear regression model to characterize the mortality data.
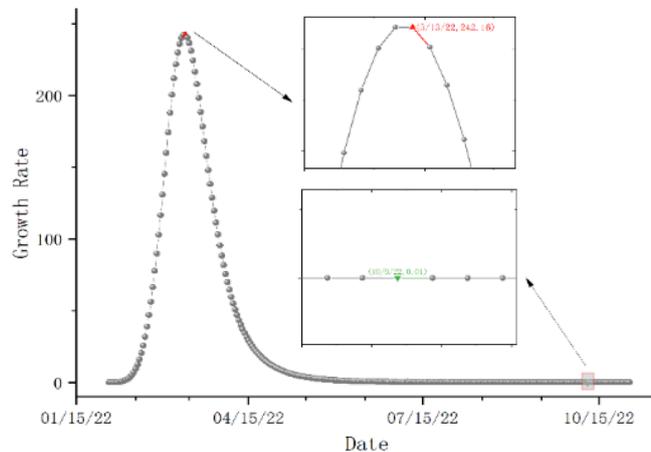

Figure 11: Speed rate of accumulative death.

A statistical analysis of the characteristics of the different populations showed that there was no significant relationship between infection with the new coronavirus and age. On the contrary, there was a strong association with age, with older infected individuals tending to be sicker. Secondly, those who had been vaccinated more often generally had lower infection and mortality rates, suggesting that the vaccine was effective against the new coronavirus.

In the time-series and non-linear regression models, we considered historical data on the number of confirmed diagnoses and deaths, and were able to describe broad trend information on the changing dynamics of the epidemic. However, the absence of external factors such as government interventions in prevention and control policies and advances in medical care may make the models less than perfect.

CONCLUSION

Based on the time series ARIMA model and nonlinear regression model established in this paper, our real data modeling shows the cumulative number of confirmed cases in Hong Kong would continue to rise in July, reaching 39,324 as of July 30, 2022. The cumulative death in Hong Kong will gradually slow, rising to 9,454 as of July 30, with 53 new deaths in July. No death caused by this epidemic will happen after 9 October, 2022 in the view of our model's prediction.

Researchers from the WHO Coordinating Centre for Epidemiology and Control of Infectious Diseases commented in The Lancet [12] that, while it is true that omicron is less virulent than previous strains in terms of the number of people infected and the proportion of serious illnesses, the extreme infectiousness and immune escape of omicron can lead to a sudden increase in the number of infections.

In Hong Kong, increasing vaccination rates among the older people may be the key to flatten the curve and overcome the challenge of the Omicron finally. In wave 5, there were 7207 deaths in Hong Kong, representing nearly 1 in 1000 of the total population. 86.96% of the deaths were in people aged 60 years or older, 88% of whom did not complete two doses of vaccination. It's also an effective way in avoiding putting unbearable pressure on the healthcare system through increasing vaccination rates.

This paper discussed and analyzed the relationship between different population groups affected by the new crown epidemic and the predicted future cumulative number of confirmed and cumulative deaths in Hong Kong. The prediction of the epidemic involving the whole community can help the government to formulate appropriate policies for reasonable and effective preventive and control measures in a timely manner, as well as to reduce the pressure felt by people about the epidemic.

FUNDING

AUTHOR CONTRIBUTIONS

Yongmei Ding and Lingxiao Xiang wrote the first draft of the manuscript, and Yongmei Ding worked on subsequent versions. All authors contributed to writing and interpretation of results. All authors read and approved the final manuscript.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing or conflicting financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

[1]   Fu G-E, Li J, Geng P-R, *et al*. Comparative study on COVID-19 prevention and control forces in typical cities of Guangdong-Hong Kong-Macao Greater Bay Area from the

perspective of medical treatment [J]. Chinese Journal of Bioengineering 2021; 41(12): 16.

[2]   Ma A, Parry J. When Hong Kong's "dynamic zero" covid-19 strategy met omicron, low vaccination rates sent deaths soaring. BMJ 2022; 377.
https://doi.org/10.1136/bmj.o980

[3]   The Government of the Hong Kong Special Administrative Region. Together, we fight the virus! Hong Kong, China: The Government of the Hong Kong Special Administrative Region; 2022.
https://www.coronavirus.gov.hk/eng/index.html.

[4]   Taylor L. Covid-19: Hong Kong reports world's highest death rate as zero covid strategy fails. BMJ 2022; 376: o707.
https://doi.org/10.1136/bmj.o707

[5]   Xie M, Dong N, Zhang X, He D. Exported cases were infected on the way: A conjecture derived from analysis on Hong Kong monthly exported COVID-19 cases. International Journal of Infectious Diseases 2022; 118: 62-4.
https://doi.org/10.1016/j.ijid.2022.02.027

[6]   Men YV, Yeung CY, Yip PSF. The association between unemployment and suicide among employed and unemployed people in Hong Kong: A time-series analysis. Journal of Affective Disorders 2022; 305: 240-243.
https://doi.org/10.1016/j.jad.2022.03.013

[7]   Yuan HY, Blakemore C. The impact of multiple non-pharmaceutical interventions on controlling COVID-19 outbreak without lockdown in Hong Kong: A modelling study.

The Lancet Regional Health-Western Pacific 2022; 20: 100343.
https://doi.org/10.1016/j.lanwpc.2021.100343

[8]   Li J-B, Lau EYH, Chan DKC. Why do Hong Kong parents have low intention to vaccinate their children against COVID-19? testing health belief model and theory of planned behavior in a large-scale survey. Vaccine 2022; 40(19): 2772-2780.
https://doi.org/10.1016/j.vaccine.2022.03.040

[9]   Koh K, Tang KC, Axhausen K. Loo BP. A metropolitan-scale, three-dimensional agent-based model to assess the effectiveness of the COVID-19 Omicron wave interventions in a hyperdense city: a case study of Hong Kong. International Journal of Infectious Diseases 2022.
https://doi.org/10.1016/j.ijid.2022.06.042

[10] Huang R, Liu M, Ding Y. Spatial-temporal distribution of COVID-19 in China and its prediction: A data-driven modeling analysis. The Journal of Infection in Developing Countries 2020; 14(03): 246-53.
https://doi.org/10.3855/jidc.12585

[11] Ding Y, Gao L. An evaluation of COVID-19 in Italy: A data-driven modeling analysis. Infectious Disease Modelling. 2020; 5: 495-501.
https://doi.org/10.1016/j.idm.2020.06.007

[12] Nealon J, Cowling BJ. Omicron severity: milder but not mild. The Lancet 2022; 399(10323): 412-413.
https://doi.org/10.1016/S0140-6736(22)00056-3