The reproductive number R_0 of COVID-19 based on estimate of a statistical time delay dynamical system

Nian Shao*, Jin Cheng†, Wenbin Chen‡

Abstract. In this paper, we estimate the reproductive number R_0 of COVID-19 based on Walling and Lipsitch framework [10] and a novel statistical time delay dynamic system.We use the observed data reported in CCDC's paper to estimate distribution of the generation interval of the infection and apply the simulation results from the time delay dynamic system as well as released data from CCDC to fit the growth rate. The conclusion is: Based our Fudan-CCDC model, the growth rate r of COVID-19 is almost in $[0.27, 0.29]$ which is larger than the growth rate 0.1 estimated by CCDC [8], and the reproductive number R_0 of COVID-19 is estimated by $3.04 \leq R_0 \leq 3.13$, which is similar to that of SARS. Some evolutions and predictions are listed.

1 One formula for the distribution of generation interval

The serial interval distribution presented in the paper "Early transmission dynamics in Wuhan, China, of novel coronavirus-infected peneumonia" [8] by the group of Chinese Center for Disease Control and Prevention(CCDC) et al, which is defined as the duration between symptom onset insuccessive cases in the transmission chain in Fig.2B of [8]. We assume that the distribution is Gamma distribution, i.e., the corresponding probability density function is

$$P(t) = \frac{b^n}{\Gamma(n)} t^{n-1} e^{-bt} = \frac{\left(\frac{2}{3}\right)^5}{\Gamma(5)} t^4 e^{-\frac{2t}{3}}. \tag{1.1}$$

the parameters are $n = 5$ and $b = \frac{2}{3}$. So the mean of the distribution is $\frac{n}{b} = 7.5$ which is exact equal to the number announced in [8], where they claimed that the mean serial interval of COVID-19 is 7.5 days (95% CI, 5.3 to 19). The formula is so beautiful, that **we believe it must be true**. The other distributions are also given in our recent report [9].

We think this distribution just as the one of generation interval mentioned in [10]. Now we can determine the reproductive numbers R_0 of COVID-19 if we know the growth rate r of COVID-19 by

$$R_c = 1 + rT_c, \tag{1.2}$$

where $T_c = 7.5$ can be obtained form (1.1). Before we discuss how to get the estimate of the growth rate r of COVID-19, let us review the elegant framework by Wallinga and Lipsitch [10].

*School of Mathematical Sciences, Fudan University, 220 Handan Road, Yangpu District, Shanghai, 200433, China, P. R. China. (16307130024@fudan.edu.cn).

†School of Mathematical Sciences and SKLCAM, Fudan University, 220 Handan Road, Yangpu District, Shanghai, 200433, China, P. R. China. (jcheng@fudan.edu.cn).

‡School of Mathematical Sciences and SKLCAM, Fudan University, 220 Handan Road, Yangpu District, Shanghai, 200433, China, P. R. China. (Corresponding author. wbchen@fudan.edu.cn).

2 Wallinga and Lipsitch's framework

According to the framework of Wallinga and Lipsitch [10], in order to obtain the reproductive numbers R_0, some variables and distributions should be known based on the data before we use their uniform framework:

- the growth rate of the epidemic r;
- the mean generation interval of the infection T_c;
- or the generation interval distribution of the infection $g(t)$.

We just mention that the mean generation interval T_c is the first moment of the generation interval distribution $g(t)$:

$$T_c = \int_0^\infty t g(t) dt. \tag{2.1}$$

If r and T_c are computed, the value of the reproductive number R_0 can be simply estimated through a linear equation([10, 1, 7, 4, 5]):

$$R_c = 1 + rT_c. \tag{2.2}$$

And if r and $g(t)$ are estimated, R_0 can be computed as

$$R_g = \frac{1}{M(-r)}, \tag{2.3}$$

where $M(z)$ is the moment generating function of the distribution $g(t)$:

$$M(z) = \int_0^\infty e^{zt} g(t) dt. \tag{2.4}$$

Wallinga and Lipstch's processes are simple but powerful, where clearly demonstrate how generation intervals shape the relationship between growth rate and reproductive numbers. Specially, if $g(t)$ is Gamma distribution with parameters (b, n) as in [10], then we can have two estimates of R_0, one is by (2.2)

$$R_{\Gamma,c} = 1 + \frac{nr}{b}, \tag{2.5}$$

and the another is by (2.3)

$$R_{\Gamma,g} = \left(1 + \frac{r}{b}\right)^n. \tag{2.6}$$

Based on the formula of the generation interval, we know that $b = \frac{2}{3}$ and $n = 5$, $R_{\Gamma,g}$ will be too sensitive on r, and we will choose (1.2) to obtain R_0, and our main focus is to get an accurate and stable estimate for the growth rate r.

In [10], Wallinga and Lipsitch point out that if the epidemic model is assumed to be susceptible-exposed-infectious-recovered(SEIR) model, then the generation interval distribution is Gamma distribution.

So we can obtain our conjecture:

Conjecture 2.1. *The epidemic model for COVID-19 is SEIR if CCDC's data are correct.*

Now we introduce how we can estimate the growth rate r.

3 Fudan-CCDC model

Our group has developed some models for the growth rate of COVID-19 (TDD-NCP models [2, 11] and Fudan-CCDC models[9]) after Jin suggested to use the time delay model to fit the real data [2]. Because of the outstanding agreement with data from CCDC and the estimations on the growth rate are very stable, we decide to use the following Fudan-CCDC model. Other models also work very well to fit the CCDC data.

Firstly, Let us introduce some variables:

- $I(t)$: (the distribution of) the cumulative number of cases in the infected stage at time t;
- $J(t)$: (the distribution of) the cumulative ones in hospitalization stage at time t;
- $G(t)$: (the distribution of) the cumulative number of infected and isolated but yet diagnosed cases at time t. They are infected in fact, but are not confirmed by the hospital, then not appeared in the infected list of CCDC;
- $I_0(t)$: $I_0(t) = I(t) - J(t) - G(t)$.

We further make some assumptions on the transition probability functions $f_2(t)$, $f_3(t)$ and $f_4(t)$:

- $f_2(t)$: the transition probability from infection to illness onset;
- $f_3(t)$: the transition probability from illness onset to hospitalization;
- $f_4(t)$: the transition probability from infection to hospitalization, which can be calculated via the convolution of $f_2(t)$ and $f_3(t)$.

We assume the log-normal distribution for $f_2(t)$ and the Weibull distrubution for $f_3(t)$[9], and the distribution parameters can be estimated form CCDC by fitting the figures in [8]. In particular, two important parameters are used:

- β: denotes the growth rate, which is exactly the same as r: $\beta = r$.
- ℓ: denotes the quarantine strategy, for example, specially, we consider this parameter as the isolation rate. Large value of ℓ suggests that the government more likes to control the spreading of COVID-19.

Now we introduce the Fudan-CCDC our Fudan-CCDC model: $I(t)$, $J(t)$ and $G(t)$ can be approximated by the following (statistical) dynamical system:

$$\frac{\mathrm{d}I}{\mathrm{d}t} = rI_0(t), \tag{3.1}$$

$$\frac{\mathrm{d}J}{\mathrm{d}t} = r\int_{-\infty}^{t} f_4(t-s)I_0(s)\mathrm{d}s, \tag{3.2}$$

$$\frac{\mathrm{d}G}{\mathrm{d}t} = \ell(t)\int_{-\infty}^{t} f_2(t-s)I_0(s)ds - \ell(t)\int_{-\infty}^{t} f_4(t-s)I_0(s)\mathrm{d}s. \tag{3.3}$$

One can also use the discrete system with each step representing one day just as we have implemented in the code:

$$I(t+1) = I(t) + rI_0(t), \tag{3.4}$$

$$J(t+1) = J(t) + r\sum_{s\le t} f_4(t-s)I_0(s), \tag{3.5}$$

$$G(t+1) = G(t) + \ell(t)\sum_{s\le t} f_2(t-s)I_0(s) - \ell(t)\sum_{s\le t} f_4(t-s)I_0(s). \tag{3.6}$$

We will fit and estimate appropriate values of β(or r) and ℓ based on the cumulative infection data from CCDC. Other versions of our Fudan-CCDC model are developing, specially, we are considering the stochastic version: that is to take random inputs into account, which is important to account for the iinfluences from other regions. In math,we can easily generalize the model by random terms such as Brownian motion in the system, meanwhile the parameters have to be obtained from the real data. The another important version is to consider the interactions of two or more regions, Jin's group have developed some models based on TTD-NCP model[3, 6]. These models will also be developed in need of disease Control and prevention.

The simulation is implemented via Matlab with some additional constraints imposed. Here the results are summarized in Table 1 with the real data employed up to Feb 16, 2020.

Table 1: The growth rate r and the reproductive numbers R_0

	the growth rate	the reproductive number
Wuhan	0.283	3.1225
Hubei(without Wuhan)	0.278	3.085
China(without Hubei)	0.273	3.0474

Moreover, the growth rate in other region are almost same, for example, both Bejing and Shanghai are 0.28. Even for the Diamond Princess Cruise ship docked in Yakushima, Japan with limited data, the growth rate is 0.275. We are surprised that our code can get uniform estimators on the growth rate r and the reproductive numbers R_0.

Quarantine strategy: We can observe that ℓ is distinct for different time stages at the different regions, we use the following assumption:

$$\ell = \begin{cases} 0, & \text{if } t < t_\ell \\ \ell, & \text{otherwise,} \end{cases}$$

which means that the quarantine strategy was observably changed at the day t_ℓ, we report the results in Table 2. We also find that ℓ for the Diamond Princes Cruise Ship is 0.55.

Starting date t_0: Our model can also track the initial date of the event t_0 if $I(t_0)$ is fixed. It is REALLY important to find the initial starting date statistically correct in different regions.

4 Evolution and prediction

Now we present some figures of the evolution and predictions of the event computed from the model, The data used in the simulation is ONLY the cumulative infections of different region

Table 2: The isolation rate ℓ and the changed point t_ℓ

	ℓ	t_ℓ
Wuhan	0.43	Jan. 20
Hubei(without Wuhan)	0.56	Jan. 20
China(without Hubei)	0.84	Jan. 20
Beijing	0.67	Jan. 15
Shanghai	0.63	Jan. 14

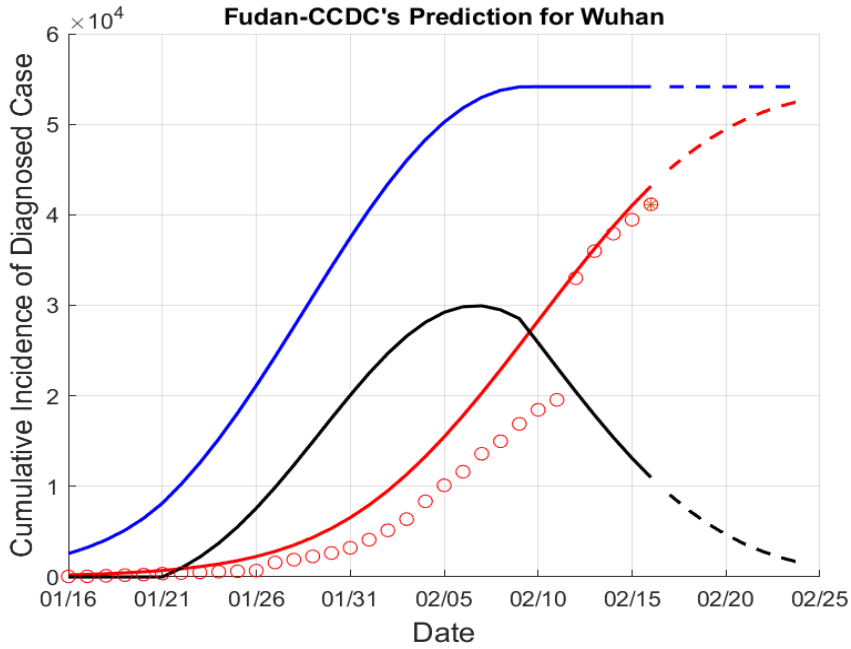
Table 3: The initial infections $I(t_0)$ and the estimated initial date t_0

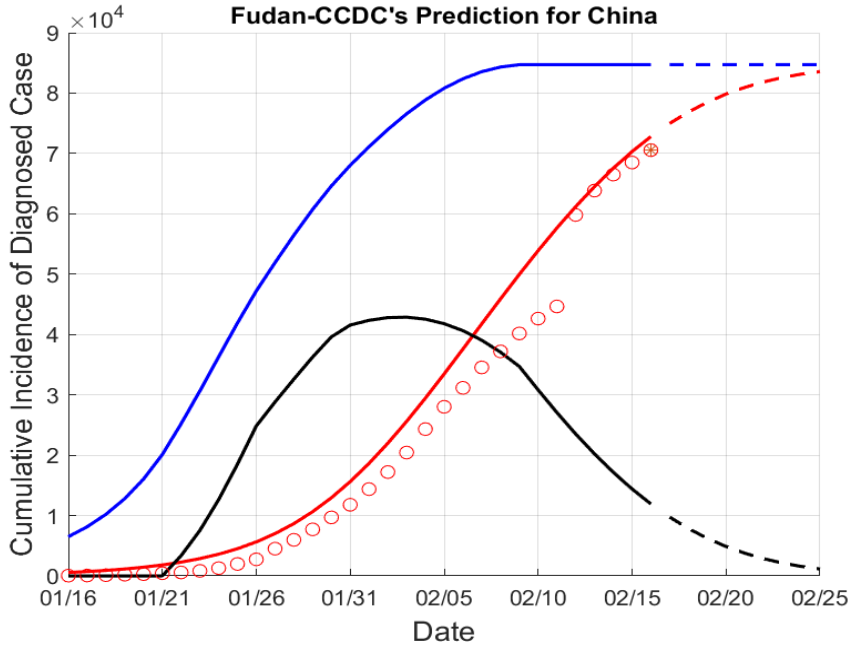
	parameter $I(t_0)$	t_0
Wuhan	4	Dec. 19
Hubei(without Wuhan)	4	Dec. 20
China(without Hubei)	4	Dec. 19
Beijing	1	Dec. 31
Shanghai	1	Dec. 31

from CCDC until Feb. 16, everyone can download it from many website or APP, we find the data from WIND(like Bloomberg).

Evolution: Here is the evolution of the cumulative number of the infected cases, cumulative number of cases in hospitalization $J(t)$, and the cumulative number of infected and isolated but yet diagnosed cases $G(t)$. According to our model, total number of infected cases will be **54140**(Wuhan), and **84701**(China). Based on this data, it is easy to estimate various data, specially for the fatality rate that can be simply estimated. We would like to point out that the rate calculated here will be varying once the quarantine strategy is changed as the rate depend on ℓ.

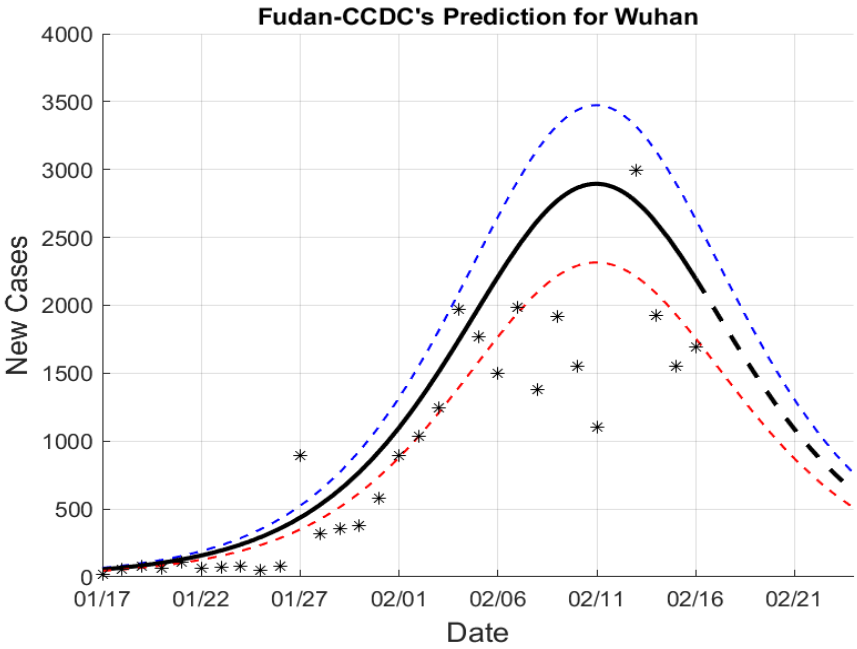
Figure 1: Evolution of COVID-19





We also can obtain the cumulative incidence of diagnosed case and new cases that occurred in a give period which is displayed in Figure 2.

Figure 2: New cases



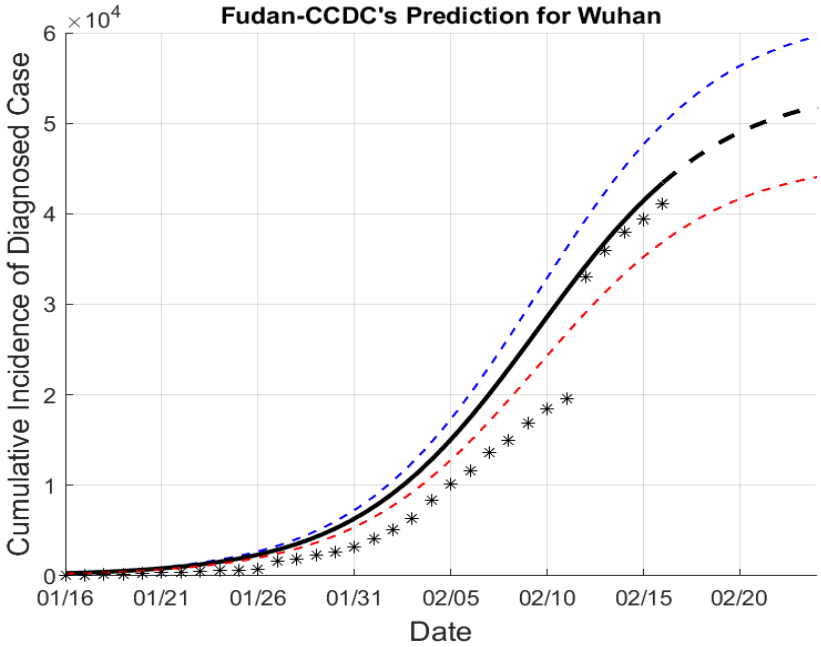
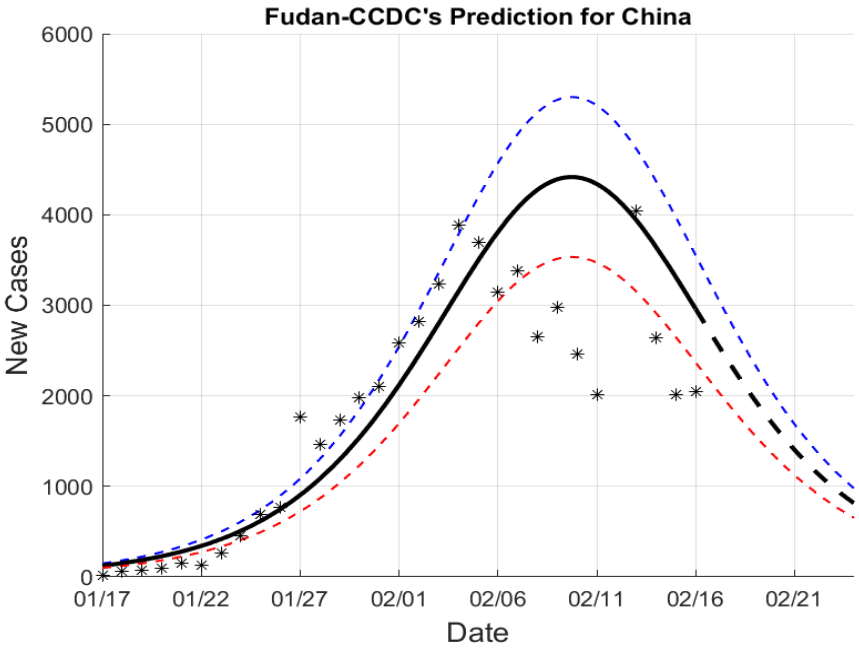
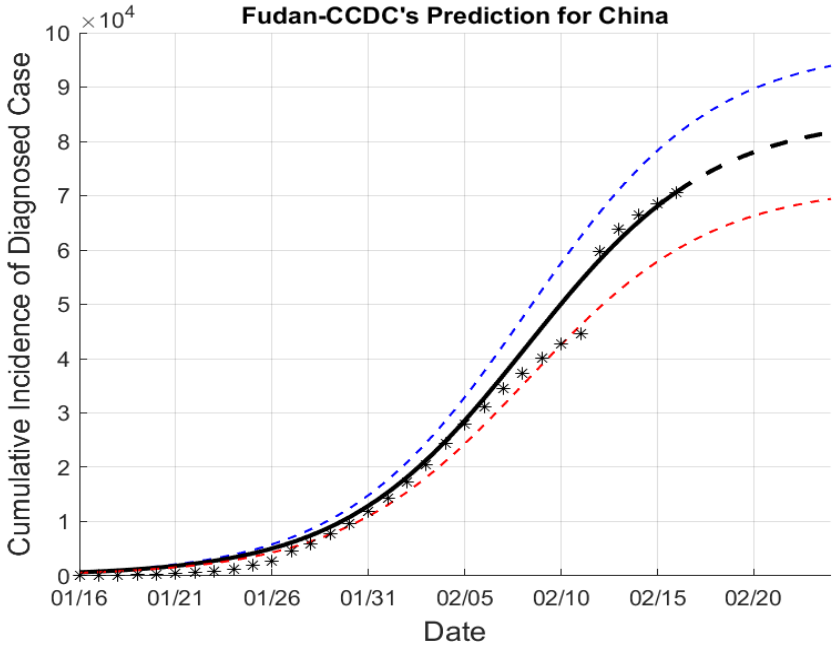


Figure 3: Cumulative incidence of diagnosed case





Some figures and other results have been submitted to Chinese Center for Disease Control and Prevention(CCDC), Wuhan.

5 Acknowledgements

We should truly thank Prof. Daqian Li and Prof. Guanhong Ding for their leading work on the SARS models, we are also sincerely grateful to every member in our group, special to Dr. Yu Jiang, Dr. Yue Ye, Dr. Boxi Xu, Dr. Yu Chen, Dr. Keji Liu, Miss Xinyue Luo, Dr. Ming Zhong,

Dr. Xiang Xu and Prof. Shuai Lu and other group members. We also thank Yongzhen Wang and Prof. Shanjian Tang(Fudan), Gang Liang(China CDC in Wuhai, Hubei), our old friend Dr. Zhihua Shen(Fusion Fin Trade, fusionfintrade.com), Prof. Xingjie Li(University of North Carolina at Charlotte), Ling Ye(China CDC in Daishan, Zhejiang), Long Chen(University of California, Irvine), Prof. Jiongming Yong(University of Central Florida) and Xinkang Chao(Mathworks). Wenbin Chen specially thank our colleague Prof. Rongmin Li who mentioned us to use the Gamma distribution. Wenbin Chen was supported in part by the National Science Foundation of China (11671098, 91630309) and 111 project(B08018), Jin Cheng was supported in part by the National Science Foundation of China (11971121). Wenbin Chen thanks Dr. Xu Chen's group in the Winning Health Technology Group Company(winning.com.cn), who always believe and support us, he also thanks his wife Dr. Jie Xu, hers hometown is Wuhan.

We should thanks everyone who are fighting for Wuhan, specially we present our deep respects to Doctor Wenliang Li and Prof. Nanshan Zhong.[1].

References

[1] R. M. Anderson and R. M. May. Infectious diseases of humans: dyanmics and control. Oxford, Uk: Oxford University Press.

[2] Y. Chen, J. Cheng, Y.Jiang and K. Liu. A time delay dynamical model for outbreak of 2019-nCov and the pramameter identification[EB/OL]. To appear in Journal in Inverse and Ill-posed Problems, 2020. arXiv:2002.00418.

[3] Y. Chen, J. Cheng, Y. Jiang and K. Liu. A time delay dynamic system with external source for the local outbreak of 2019-nCoV. arXiv:2002.02590, 2020.

[4] N. Ferguson et al. Strategies for containing an emerging influenza pandemic in Southeast Asia. Nature 437, 209-214.

[5] N. Ferguson et al. Strategies for mitigating and influenza pandemic. Nature 442, 448-452.

[6] K. Liu, Y. Jiang, Y. Yan, and W. Chen. A time delay dynamic model with external source and the basic reproductive number estimation for the outbreak of Novel Coronavirus Pneumonia. Submitted to Contral Theory and Applications (in Chinese), 2020.

[7] O. G. Pybus et al. The epidemic behaviour of hepatitis C virus. Science 22, 2323-2325.

[8] L. Qun, G. Xuhua, W. Peng, W. Xiaoye, Z. Lei, T. Yeqing, R. Ruiqi, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. N Engl J Med, 2020. DOI: 10.1056/NEJMoa2001316.

[9] N. Shao, Y. Chen,J. Cheng and W. Chen. Some novel statistical time delay dynamic model by statistics data from CCDC on Novel Coronavirus Pneumonia, submitted to Journal of Control Theory and Applications(in Chinese), 2020.

[1] There must exists lots of typos and many references are note cited, since the paper is finished in a hurry and the first version is submitted at Feb. 17, 5:25am, 2020, after Xingjie Li helps me to modify the notes. Please let us know if you have any question and suggestions, please email us at wbchen@fudan.edu.cn. Moreover, we don't know where to publish our results without any proof in mathematics. We both are math people, we believe the proof. But now I do not care about the proof. Now we believe the model, and hope it can be useful for the people in Wuhan.

[10] J. Wallinga and M. Lipstch. How generation intervals shape the relatoinship between growth rates and reproductive numbers. Proc. R. Soc. B, 274, 599-604, 2007.

[11] Y. Yan, Y. Chen, K. Liu, X. Luo, B. Xu, Y. Jiang and J. Cheng. Modeling and prediction for the trend of outbreak of 2019-nCoV based on a time-delay dynamic system (in Chinese). Sci. Sin. Math., 50(3), 1-8, 2020. https://doi.org/10.1360/SSM-2020-0026.