Article

Predicting the cumulative number of cases for the COVID-19 epidemic in China from early data

Zhihua Liu [1,‡], **Pierre Magal** [2,‡], **Ousmane Seydi** [3,‡] **and Glenn Webb** [4,‡*]

1. School of Mathematical Sciences, Beijing Normal University, Beijing 100875, People's Republic of China; zhihualiu@bnu.edu.cn
2. Univ. Bordeaux, IMB, UMR 5251, F-33400 Talence, France. CNRS, IMB, UMR 5251, F-33400 Talence, France; pierre.magal@u-bordeaux.fr
3. Département Tronc Commun, École Polytechnique de Thiés, Sénégal; oseydi@ept.sn
4. Mathematics Department, Vanderbilt University, Nashville, TN, USA; glenn.f.webb@vanderbilt.edu
*   Correspondence: glenn.f.webb@Vanderbilt.Edu
‡   These authors contributed equally to this work.

Abstract: We model the COVID-19 coronavirus epidemic in China. We use early reported case data to predict the cumulative number of reported cases to a final size. The key features of our model are the timing of implementation of major public policies restricting social movement, the identification and isolation of unreported cases, and the impact of asymptomatic infectious cases.

Keywords: corona virus, reported and unreported cases, isolation, quarantine, public closings, epidemic mathematical model.

1. Introduction

Many mathematical models of the COVID-19 coronavirus epidemic in China have been developed, and some of these are listed in our references [4,7,9–15]. We develop here a model describing this epidemic, focused on the effects of the Chinese government imposed public policies designed to contain this epidemic, and the number of reported and unreported cases that have occurred. Our model here is based on our model of this epidemic in [5], which was focused on the early phase of this epidemic (January 20 through January 29) in the city of Wuhan, the epicenter of the early outbreak. During this early phase, the cumulative number of daily reported cases grew exponentially. In [5], we identified a constant transmission rate corresponding to this exponential growth rate of the cumulative reported cases, during this early phase in Wuhan.

On January 23, 2020, the Chinese government imposed major public restrictions on the population of Wuhan. Soon after, the epidemic in Wuhan passed beyond the early exponential growth phase, to a phase with slowing growth. In this work, we assume that these major government measures caused the transmission rate to change from a constant rate to a time dependent exponentially decreasing rate. We identify this exponentially decreasing transmission rate based on reported case data after January 29. We then extend our model of the epidemic to the central region of China, where most cases occurred. Within just a few days after January 29, our model can be used to project the time-line of the model forward in time, with increasing accuracy, to a final size.

2. Model

The model consists of the following system of ordinary differential equations:

$$\begin{cases} S'(t) = -\tau(t)S(t)[I(t) + U(t)], \\ I'(t) = \tau(t)S(t)[I(t) + U(t)] - \nu I(t), \\ R'(t) = \nu_1 I(t) - \eta R(t), \\ U'(t) = \nu_2 I(t) - \eta U(t). \end{cases} \tag{1}$$

This system is supplemented by initial data

$$S(t_0) = S_0 > 0, \ I(t_0) = I_0 > 0, \ R(t_0) = 0 \text{ and } U(t_0) = U_0 \geq 0. \tag{2}$$

Here $t \geq t_0$ is time in days, t_0 is the beginning date of the model of the epidemic, $S(t)$ is the number of individuals susceptible to infection at time t, $I(t)$ is the number of asymptomatic infectious individuals at time t, $R(t)$ is the number of reported symptomatic infectious individuals at time t, and $U(t)$ is the number of unreported symptomatic infectious individuals at time t.

Asymptomatic infectious individuals $I(t)$ are infectious for an average period of $1/\nu$ days. Unreported symptomatic infectious individuals $U(t)$ are infectious for an average period of $1/\eta$ days. We assume that reported symptomatic infectious individuals $R(t)$ are reported and isolated immediately upon showing symptoms, and cause no further infections. The asymptomatic infectious individuals $I(t)$ can also be viewed as having a low-level symptomatic state, which does not result in their being reported and isolated. All infections are acquired from either $I(t)$ or $U(t)$ individuals.

The parameters of the model are listed in Table 1.

Symbol	Interpretation	Method
t_0	Time at which the epidemic started	fitted
S_0	Number of susceptible at time t_0	fixed
I_0	Number of asymptomatic infectious at time t_0	fitted
U_0	Number of unreported symptomatic infectious at time t_0	fitted
$\tau(t)$	Transmission rate at time t	fitted
$1/\nu$	Average time during which asymptomatic infectious are asymptomatic	fixed
f	Fraction of asymptomatic infectious that become reported symptomatic infectious	fixed
$\nu_1 = f\nu$	Rate at which asymptomatic infectious become reported symptomatic	fitted
$\nu_2 = (1-f)\nu$	Rate at which asymptomatic infectious become unreported symptomatic	fitted
$1/\eta$	Average time symptomatic infectious have symptoms	fixed

Table 1. *Parameters of the model.*

3. Data

We use data from the National Health Commission of the People's Republic of China and the Chinese CDC for mainland China as of February 15, 2020:

January	20	21	22	23	24	25	26	27	28	29	3 0	31
	291	440	571	830	1287	1975	2744	4515	5974	7711	9692	11791

February	1	2	3	4	5	6	7	8	9	10	11	12
	14380	17205	20438	24324	28018	31161	34546	37198	40171	42638	44653	46472

February	13	14	15
	48467	49970	51091

Table 2. *Cumulative daily reported case data from January 20, 2020 to February 15, 2020 for mainland China from the National Health Commission of the People's Republic of China and the Chinese CDC.*

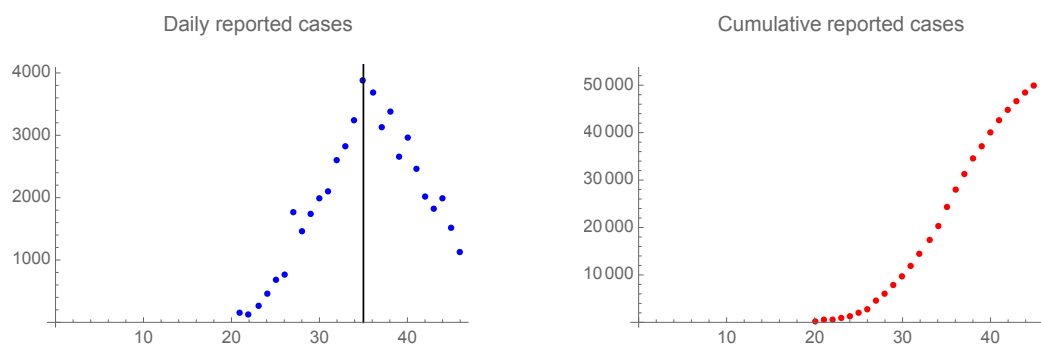We plot the data for daily reported cases and the cumulative reported cases in Figure 1.

Figure 1. *Daily reported cases data (left) and cumulative reported cases data (right). The epidemic turning point of the reported case data is approximately February 4, 2020 (day 35, day 1 = January 1, 2020).*

4. Model parameters

We assume $f = 0.8$, which means that 20% of symptomatic infectious cases go unreported. We assume $\eta = 1/7$, which means that the average period of infectiousness of both unreported symptomatic infectious individuals and reported symptomatic infectious individuals is 7 days. We assume $\nu = 1/7$, which means that the average period of infectiousness of asymptomatic infectious individuals is 7 days. These values can be modified as further epidemiological information becomes known.

In our previous work, we assumed that in the early phase of the epidemic (January 20 through January 29), the cumulative number of reported cases grew approximately exponentially, according to the formula:

$$CR(t) = \chi_1 \exp(\chi_2 t) - \chi_3, \ t \geq t_0$$

with values $\chi_1 = 0.16$, $\chi_2 = 0.38$, $\chi_3 = 1.1$. These values of χ_1, χ_2, and χ_3 were fitted to reported case data from January 20 to January 29. We assumed the initial value $S_0 = 11,000,000$, the population of the city Wuhan, which was the epicenter of the epidemic outbreak. The other initial conditions are

$$I_0 = \frac{\chi_2 \chi_3}{f(\nu_1 + \nu_2)} = 3.7, \ \ U_0 = \left(\frac{(1-f)(\nu_1 + \nu_2)}{\eta + \chi_2} \right) I_0 = 0.2, \ \ R_0 = 0.0.$$

The value of the transmission rate $\tau(t)$, during the early phase of the epidemic, when the cumulative number of reported cases was approximately exponential, is the constant value

$$\tau_0 = \left(\frac{\chi_2 + \nu_1 + \nu_2}{S_0} \right) \left(\frac{\eta + \chi_2}{\nu_2 + \eta + \chi_2} \right) = 4.47 \times 10^{-8}.$$

The initial time is

$$t_0 = 5 = \frac{1}{\chi_2} \left(\log(\chi_3) - \log(\chi_1) \right).$$

The value of the basic reproductive number is

$$\mathcal{R}_0 = \left(\frac{\tau_0 S_0}{\nu_1 + \nu_2} \right) \left(1 + \frac{\nu_2}{\eta} \right) = 4.16.$$

These parameter formulas were derived in [5].

After January 23, strong government measures in all of China, such as isolation, quarantine, and public closings, strongly impacted the transmission of new cases. The actual effects of these measures were complex, and we use an exponential decrease for the transmission rate $\tau(t)$ to incorporate these effects after the early exponential increase phase. The formula for $\tau(t)$ during the exponential decreasing phase was derived by a fitting procedure. The formula for $\tau(t)$ is

$$\begin{cases} \tau(t) = \tau_0, \ 0 \le t \le 24, \\ \tau(t) = \tau_0 \exp\left(-\mu\left(t - 24\right)\right), \ 24 < t, \end{cases} \tag{3}$$

where January 24 and $\mu = 0.12$ are fitted from on-going reported case data after January 24. In Figure 2, we plot the graph of $\tau(t)$.
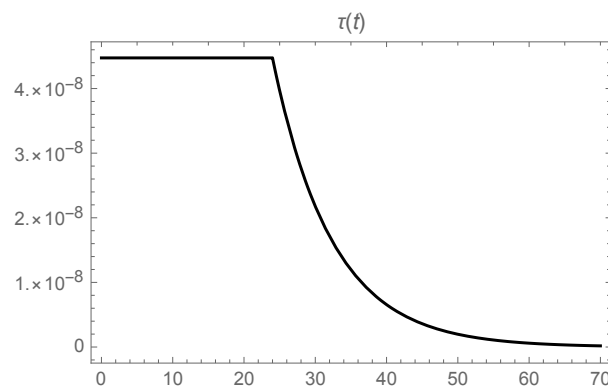

Figure 2. *Graph of the transmission rate $\tau(t)$.*

5. Model simulation

We numerically simulated the model (2.1) to project forward in time the time-line of the epidemic after the government imposed interventions. We set $\tau(t)$ in (4.1) to $S_0 \tau(t)/1,400,050,000$, where $S_0 = 11,000,000$ and $1,400,050,000$ is the population of mainland China, excluding Hong Kong, Macao and Taiwan. We set $S(t_0)$ in (2.2) to $1,400,050,000$. We set $t_0 = 5.0$, $I(t_0) = 3.7$, $U(t_0) = 0.2$ and $R(t_0) = 0$. In Figure 3, we plot the graphs of $CR(t)$ (cumulative reported cases), $CU(t)$ (cumulative unreported cases), $R(t)$, and $U(t)$ from the numerical simulation.
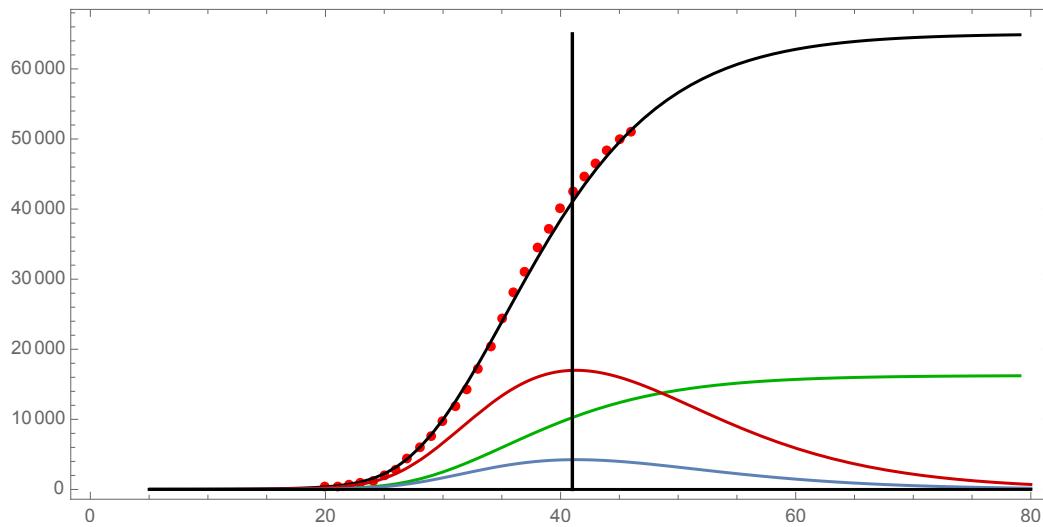
Figure 3. *Graphs of $CR(t)$ (black), $CU(t)$ (green), $U(t)$ (blue) and $R(t)$ (red). The red dots are the reported case data. The final size of the epidemic is approximately $65,000$ reported cases, approximately $16,000$ unreported cases, and approximately $81,000$ total cases. The basic reproductive number is $\mathcal{R}_0 = 4.16$. The turning point of the epidemic is approximately day $41 = February 10, 2020$.*

In Figure 4 we plot the graphs of the reported cases $R(t)$ and the infectious pre-symptomatic cases $I(t)$. The blue dots are obtained from the reported cases data (Table 2) for each day beginning on January 26, by subtracting from each day, the value of the reported cases one week earlier.
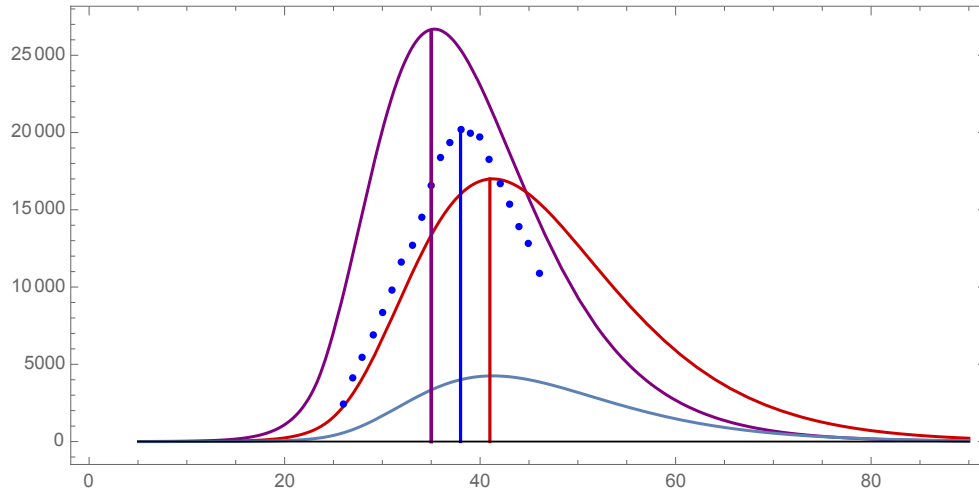

Figure 4. *Graphs of $R(t)$ (red), $U(t)$ (blue), and $I(t)$ (purple). The blue dots are the day by day weekly reported data. The turning point of the asymptomatic infectious cases $I(t)$ is approximately day 35. The turning point of the reported cases $R(t)$ and the unreported cases $U(t)$ is approximately day 41. The turning point of the day by day weekly reported data is approximately day 38.*

Our model transmission rate $\tau(t)$ can be modified to illustrate the effects of an earlier or later implementation of the major public policy interventions that occurred in this epidemic. The implementation one week earlier (24 is replaced by 17 in (4.1)) is graphed in Figure 5 (top). All other parameters and the initial conditions remain the same. The total reported cases is approximately $4,500$ and the total unreported cases is approximately $1,100$. The implementation one week later (24 is replaced by 31 in (4.1)) is graphed in Figure 5 (bottom). The total reported cases is approximately

$820,000$ and the total unreported cases is approximately $200,000$. The timing of the institution of major social restrictions is critically important in mitigating the epidemic.
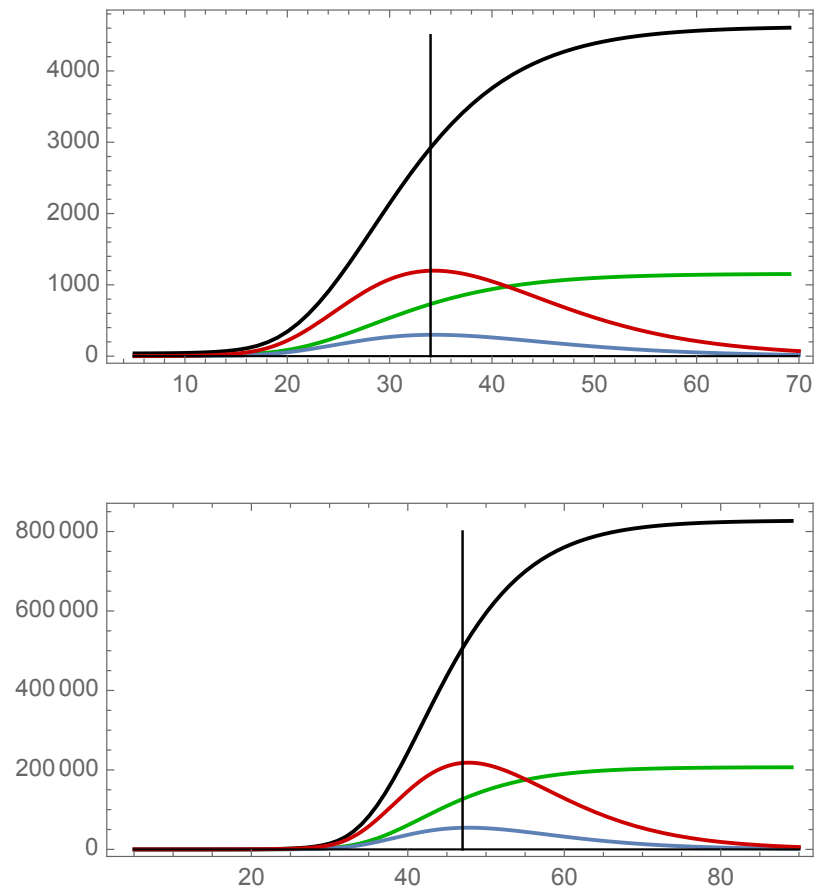

Figure 5. *Graphs of $CR(t)$ (black), $CU(t)$ (green), $U(t)$ (blue), and $R(t)$ (red). Top: The major public policy interventions were implemented one week earlier (January 17). Bottom: The major public policy interventions were implemented one week later (February 1). The one week earlier implementation resulted in a final size of approximately $5,750$ total cases, with turning point day 34 .The one week later implementation resulted in a final size of approximately $1,234,000$ total cases, with turning point day 47.*

The number of unreported cases is of major importance in understanding the evolution of an epidemic, and involves great difficulty in their estimation. The data from January 20 to February 15 for reported cases in Table 2, was only for tested cases. Between February 11 and February 15, additional clinically diagnosed case data, based on medical imaging showing signs of pneumonia, was also reported by the Chinese CDC. Since February 16, only tested case data has been reported by the Chinese CDC, because new NHC guidelines removed the clinically diagnosed category. Thus, after February 15, there is a gap in the reported case data that we used up to February 15. The uncertainty of the number of unreported cases for this epidemic includes this gap, but goes even further to include additional unreported cases.

We assumed previously that the fraction f of reported cases was $f = 0.8$ and the fraction of unreported cases was $1 - f = 0.2$. Our model formulation can be applied with varying values for the fraction f. In Figure 6 below, we provide illustrations with the fraction $f = 0.4$ (top) and $f = 0.6$ (bottom). The formula for the time dependent transmission rate $\tau(t)$ in (4.1) involves a new value for τ_0 and μ for each case. The final size of the epidemic when $f = 0.4$ is approximately 164,700 cases, and

the final size of the epidemic when $f = 0.6$ is approximately 110,700 cases. From these simulations, we see that estimation of the number of unreported cases has major importance in understanding the severity of this epidemic.
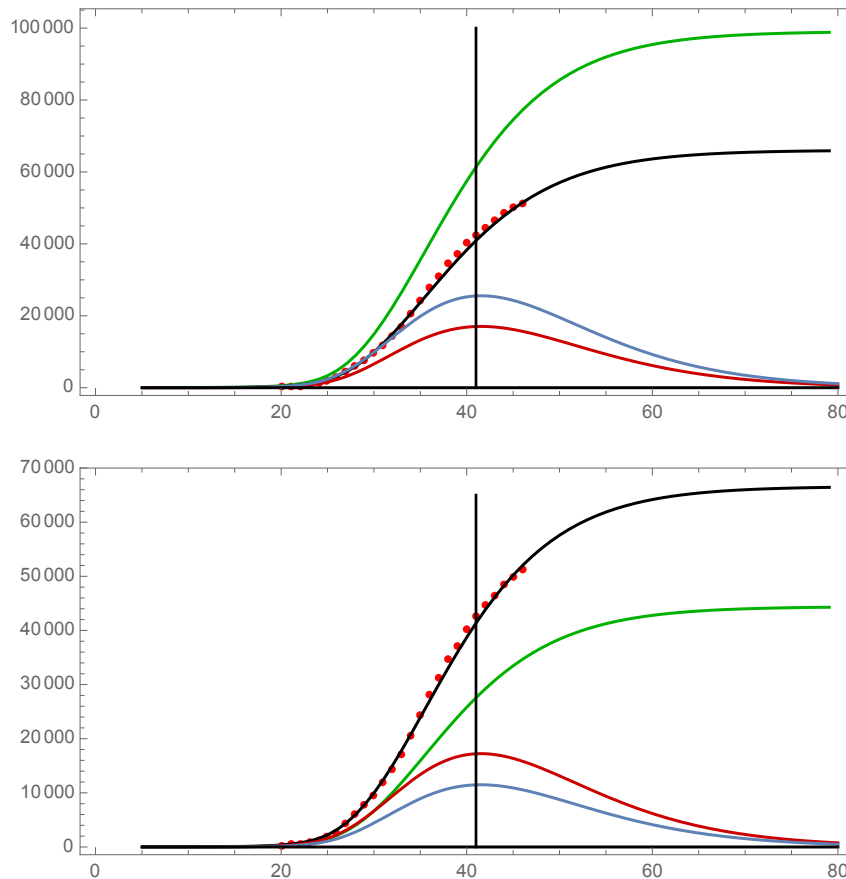

Figure 6. *Graphs of $CR(t)$ (black), $CU(t)$ (green), $U(t)$ (blue), and $R(t)$ (red). The red dots are the cumulative reported case data from Table 2. Top: $f = 0.4$, $t_0 = 5.0$, $\tau_0 = 4.054 \times 10^{-8}$, $\mu = 0.127$, $I(t_0) = 7.32$, $U(t_0) = 1.20$, $R(t_0) = 0.0$, and the basic reproductive number is $\mathcal{R}_0 = 5.03$. The final size of the epidemic is approximately $65,900$ reported cases, approximately $98,800$ unreported cases, and approximately $164,700$ total cases. The turning point of the epidemic is approximately day 41 = February 10. Bottom: $f = 0.6$, $t_0 = 5.0$, $\tau_0 = 4.254 \times 10^{-8}$, $\mu = 0.123$, $I(t_0) = 4.88$, $U(t_0) = 0.53$, $R(t_0) = 0.0$, and the basic reproductive number is $\mathcal{R}_0 = 4.25$. The final size of the epidemic is approximately $66,500$ reported cases, approximately $44,300$ unreported cases, and approximately $110,800$ total cases. The turning point of the epidemic is approximately day 41 = February 10.*

6. Discussion

We have developed a model of the COVID-19 epidemic in China that incorporates key features of this epidemic: (1) the importance of the timing of the implementation of major government public restrictions designed to mitigate the severity of the epidemic; (2) the importance of both reported and unreported cases in interpreting the number of reported cases; and (3) the importance of asymptotic infectious cases in the disease transmission. In our model formulation, we divide infectious individuals into asymptomatic and symptomatic infectious individuals. The symptomatic infectious phase is also divided into reported and unreported cases. Our model formulation is based on our work [5], in which we developed a method to estimate epidemic parameters at an early stage of an epidemic, when the number of cumulative cases grows exponentially. The general method in [5], was applied to

the COVID-19 epidemic in Wuhan, China, to identify the constant transmission rate corresponding to the early exponential growth phase. We then model the effects of the major government imposed public restrictions in China, beginning on January 23, as a time-dependent exponentially decaying transmission rate after January 24. With this time dependent exponentially decreasing transmission rate, we are able to fit with increasing accuracy, our model simulations to the Chinese CDC reported case data for all of China, forward in time through February 15, 2020.

Our model demonstrates the effects of implementing major government public policy measures. By varying the date of the implementation of these measures in our model, we show that had implementation occurred one week earlier, then a significant reduction in the total number of cases would have resulted. We also showed that if these measures had occurred one week later, then a significant increase in the total number of cases would have occurred. It is evident, that control of a COVID-19 epidemic is very dependent on the early implementation of major restrictions on public functions.

We also varied the fraction of unreported cases involved in the transmission dynamics. We showed that if this fraction is higher, then a significant increase in the number of total cases results. If it is lower, then a significant reduction occurs. It is evident, that control of a COVID-19 epidemic is very dependent on identifying and isolating symptomatic infectious cases. It is also possible to vary the parameter ν (the reciprocal of the average period of asymptomatic infectiousness) and η (the reciprocal of the average period of symptomatic infectiousness), to obtain similar dynamics of the epidemic in agreement with cumulative reported case data.

Our model was specified to the COVID-19 outbreak in China, but it is applicable to any outbreak location for a COVID-19 epidemic.

Author Contributions: Z.L, O.S, P.M and G.W conceived and designed the study. P.M and O.S analyzed the data, carried out the analysis and performed numerical simulations, Z.L and G.W conducted the literature review. All authors participated in writing and reviewing of the manuscript.

Conflicts of Interest: Declare conflicts of interest or state "The authors declare no conflict of interest."

Abbreviations

The following abbreviations are used in this manuscript:

MDPI	Multidisciplinary Digital Publishing Institute
DOAJ	Directory of open access journals
TLA	Three letter acronym
LD	linear dichroism

References

1. A. Ducrot, P. Magal, T. Nguyen and G. F. Webb, Identifying the number of unreported cases in SIR epidemic models, *Math. Med. Biol. (to appear).*
2. K.P. Hadeler, Parameter identification in epidemic models, *Math. Biosci.*, **229** (2011), 185-189.
3. K.P. Hadeler, Parameter estimation in epidemic models: simplified formulas, *Can. Appl. Math. Q.*, **19** (2011), 343-356.
4. D.S. Hui, *et al.* The continuing 2019-nCoV epidemic threat of novel corona viruses to global health - The latest 2019 novel corona virus outbreak in Wuhan, China, *Int. J. Infect. Dis.* **91** (2020), 264-266.
5. Z. Liu, P. Magal, O. Seydi, and G. Webb, Understanding unreported cases in the 2019-nCov epidemic outbreak in Wuhan, China, and the importance of major public health interventions, *submitted* (2020).
6. P. Magal and G. Webb, The parameter identification problem for SIR epidemic models: Identifying Unreported Cases, *J. Math. Biol.* **77(6-7)** (2018), 1629-1648.
7. H. Nishiura, N. M. Linton, and A. R. Akhmetzhanov, Initial cluster of novel coronavirus (2019-nCoV) infections in Wuhan, China is consistent with substantial human-to-human transmission, *J. Clin. Med.* (2020).

8.   H. Nishiura et al., The Rate of Under ascertainment of Novel Coronavirus (2019-nCoV) Infection: Estimation Using Japanese Passengers Data on Evacuation Flights, *J. Clin. Med.* (2020).

9.   K. Roosa, *et al.*, Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, *Infect. Dis. Model.* (2020).

10. Y. Shao and J. Wu, IDM editorial statement on the 2019-nCoV. *Infect. Dis. Model.* (2020).

11. B. Tang, N. Bragazzi, Q. Li, S. Tang, Y. Xiao, and J. Wu, An updated estimation of the risk of transmission of the novel coronavirus (2019-nCov), *Infect. Dis. Model.* **5** (2020), 248-255.

12. B. Tang, X. Wang, Q. Li, N. Bragazzi, S. Tang, Y. Xiao, and J. Wu, Estimation of the transmission risk of the 2019-nCoV and its implication for public health interventions. *J. Clin. Med.*, **9(2)** (2020).

13. R. N. Thompson, Novel coronavirus outbreak in Wuhan, China, 2020: Intense surveillance Is vital for preventing sustained transmission in new locations. *J. Clin. Med.* **9(2)**, (2020).

14. J. T. Wu, K. Leung,and G. M. Leung, Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study, *The Lancet* (2020).

15. S. Zhao, *et al.*, Estimating the unreported number of novel Coronavirus (2019-nCoV) cases in China in the first half of January 2020: A data-driven modelling analysis of the early outbreak, *J. Clin. Med.* **9(2)** (2020), 388.