Understanding unreported cases in the COVID-19 epidemic outbreak and the importance of major public health interventions

Pierre Magal

University of Bordeaux, France

GdT Maths4covid19

https://www.ljll.math.upmc.fr/fr/evenements/workshops-et-conferences
Co-authors

- **Quentin Griette**, *University of Bordeaux, France*

![Quentin Griette](Image)

- **Zhihua Liu**, *Beijing Normal University, China*

![Zhihua Liu](Image)

- **Ousmane Seydi**, *Ecole Polethnique de Thies, Senegal*

![Ousmane Seydi](Image)

- **Glenn Webb**, *Vanderbilt University, USA*

![Glenn Webb](Image)
Abstract

We develop a mathematical model to provide epidemic predictions for the COVID-19 epidemic in China. We use reported case data from the Chinese Center for Disease Control and Prevention and the Wuhan Municipal Health Commission to parameterize the model. From the parameterized model we identify the number of unreported cases. We then use the model to project the epidemic forward with varying level of public health interventions. The model predictions emphasize the importance of major public health interventions in controlling COVID-19 epidemics.
1 Introduction

2 Results

3 Numerical Simulations
What are unreported cases?

**Unreported cases** are missed because authorities aren’t doing enough testing, or ‘preclinical cases’ in which people are incubating the virus but not yet showing symptoms.

Research published\(^1\) traced COVID-19 infections which resulted from a business meeting in Germany attended by someone infected but **who showed no symptoms at the time**. Four people were ultimately infected from that single contact.

---

Why unreported cases are important?

A team in Japan\textsuperscript{2} reports that 13 evacuees from the \textit{Diamond Princess} were infected, of whom 4, or 31\%, \textbf{never developed symptoms}.

A team in China \textsuperscript{3} suggests that by 18 February, there were 37,400 people with the virus in Wuhan whom authorities didn’t know about.


\textsuperscript{3}Wang et al. Evolving Epidemiology and Impact of Non-pharmaceutical Interventions on the Outbreak of Coronavirus Disease 2019 in Wuhan, China, \textit{medRxiv} (2020)
Early models designed for the COVID-19

- Wu et al. 4 used a susceptible-exposed-infectious-recovered metapopulation model to simulate the epidemics across all major cities in China.

- Tang et al. 5 proposed an SEIR compartmental model based on the clinical progression based on the clinical progression of the disease, epidemiological status of the individuals, and the intervention measures which did not consider unreported cases.

---


Early results on identification the number of unreported cases

Identifying the number of unreported cases was considered recently in

- Magal and Webb\(^6\)
- Ducrot et al.\(^7\)

In these works we consider an SIR model and we consider the Hong-Kong seasonal influenza epidemic in New York City in 1968-1969.

---


The model

Our model consists of the following system of ordinary differential equations

\[
\begin{align*}
S'(t) &= -\tau S(t)[I(t) + U(t)], \\
I'(t) &= \tau S(t)[I(t) + U(t)] - \nu I(t), \\
R'(t) &= \nu_1 I(t) - \eta R(t), \\
U'(t) &= \nu_2 I(t) - \eta U(t).
\end{align*}
\]  

(2.1)

Here \( t \geq t_0 \) is time in days, \( t_0 \) is the beginning date 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 (i.e. symptomatic infectious with severe symptoms) at time \( t \), and \( U(t) \) is the number of unreported symptomatic infectious individuals (i.e. symptomatic infectious with mild symptoms) at time \( t \). This system is supplemented by initial data

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

(2.2)
Exposed individuals are infected but not yet capable to transmit the pathogen.

A team in China detected high viral loads in 17 people with COVID-19 soon after they became ill. Moreover, another infected individual never developed symptoms but shed a similar amount of virus to those who did.

In Liu et al. we compare the model (2.1) with exposure and the best fit is obtained for an average exposed period of 6-12 hours.

---

Figure: Compartments and flow chart of the model.
Parameters of the model

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

Table: Parameters of the model.
Comparison of Model (2.1) with the Data

For influenza disease outbreaks, the parameters $\tau$, $\nu$, $\nu_1$, $\nu_2$, $\eta$, as well as the initial conditions $S(t_0)$, $I(t_0)$, and $U(t_0)$, are usually unknown. Our goal is to identify them from specific time data of reported symptomatic infectious cases. To identify the unreported asymptomatic infectious cases, we assume that the cumulative reported symptomatic infectious cases at time $t$ consist of a constant fraction along time of the total number of symptomatic infectious cases up to time $t$. In other words, we assume that the removal rate $\nu$ takes the following form: $\nu = \nu_1 + \nu_2$, where $\nu_1$ is the removal rate of reported symptomatic infectious individuals, and $\nu_2$ is the removal rate of unreported symptomatic infectious individuals due to all other causes, such as mild symptom, or other reasons.
Cumulative number of reported cases

The cumulative number of reported symptomatic infectious cases at time $t$, denoted by $CR(t)$, is

$$CR(t) = \nu_1 \int_{t_0}^{t} I(s) ds.$$  \hfill (2.3)
Phenomenological model

We assume that $CR(t)$ has the following special form:

$$CR(t) = \chi_1 \exp (\chi_2 t) - \chi_3.$$  

(2.4)

We evaluate $\chi_1, \chi_2, \chi_3$ using the reported cases data.

We obtain the model starting time of the epidemic $t_0$ from 2.6

$$CR(t_0) = 0 \iff \chi_1 \exp (\chi_2 t_0) - \chi_3 = 0 \iff t_0 = \frac{1}{\chi_2} (\ln (\chi_3) - \ln (\chi_1)).$$
Estimation of the parameters

<table>
<thead>
<tr>
<th>Name of the Parameter</th>
<th>$\chi_1$</th>
<th>$\chi_2$</th>
<th>$\chi_3$</th>
<th>$t_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>From China</td>
<td>0.16</td>
<td>0.38</td>
<td>1.1</td>
<td>5.12</td>
</tr>
<tr>
<td>From Hubei</td>
<td>0.23</td>
<td>0.34</td>
<td>0.1</td>
<td>$-2.45$</td>
</tr>
<tr>
<td>From Wuhan</td>
<td>0.36</td>
<td>0.28</td>
<td>0.1</td>
<td>$-4.52$</td>
</tr>
</tbody>
</table>

Table: Estimation of the parameters $\chi_1, \chi_2, \chi_3$ and $t_0$ by using the cumulative reported cases Chinese CDC and Wuhan Municipal Health Commission.

Remark 2.1

The time $t = 0$ will correspond to 31 December. Thus, in Table 17, the value $t_0 = 5.12$ means that the starting time of the epidemic is 5 January, the value $t_0 = -2.45$ means that the starting time of the epidemic is 28 December, and $t_0 = -4.52$ means that the starting time of the epidemic is 26 December.
Data

We use three sets of reported data to model the epidemic in Wuhan: First, data from the Chinese CDC for **mainland China**, second, data from the Wuhan Municipal Health Commission for **Hubei Province**, and third, data from the Wuhan Municipal Health Commission for **Wuhan Municipality**. These data vary but represent the epidemic transmission in Wuhan, from which almost all the cases originated from Hubei province.
Fit of the exponential model (2.6) to the data
Remark 2.2

As long as the number of reported cases follows (2.1), we can predict the future values of $CR(t)$. For $\chi_1 = 0.16$, $\chi_2 = 0.38$, and $\chi_3 = 1.1$, we obtain

<table>
<thead>
<tr>
<th></th>
<th>Jan. 30</th>
<th>Jan. 31</th>
<th>Feb. 1</th>
<th>Feb. 2</th>
<th>Feb. 3</th>
<th>Feb. 4</th>
<th>Feb. 5</th>
<th>Feb. 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>8510</td>
<td>12390</td>
<td>18050</td>
<td>26290</td>
<td>38290</td>
<td>55770</td>
<td>81240</td>
<td>118320</td>
</tr>
</tbody>
</table>

The actual number of reported cases for China are 8163 confirmed for 30 January, 11791 confirmed for Jan. 31, and 14380 confirmed for Feb. 1. Thus, the exponential formula (2.6) overestimates the number reported after Jan. 30.
Estimation of the parameters for the model (2.1)

Since the COVID-19 is a new virus for the population in Wuhan, we can assume that all the population of Wuhan is susceptible. So we fix

\[ S_0 = 11.081 \times 10^6 \]

which corresponds to the total population of Wuhan (11 millions people).

From now on, we fix the fraction \( f \) of symptomatic infectious cases that are reported. We assume that between 80\% and 100\% of infectious cases are reported. Thus, \( f \) varies between 0.8 and 1.

We assume \( 1/\nu \), and the average time during which the patients are asymptomatic infectious varies between 1 day and 7 days. We assume that \( 1/\eta \), the average time during which a patient is symptomatic infectious, varies between 1 day and 7 days. We can compute

\[ \nu_1 = f\nu \quad \text{and} \quad \nu_2 = (1 - f)\nu. \]  \hspace{1cm} (2.5)
Estimation of the parameters for the model (2.1)

\[ CR(t) = \chi_1 \exp(\chi_2 t) - \chi_3 = \nu_1 \int_{t_0}^{t} I(s) ds. \]  

(2.6)

and by fixing \( S(t) = S_0 \) in the \( I \)-equation of system (2.1), we obtain

\[ I_0 = \frac{\chi_1 \chi_2 \exp(\chi_2 t_0)}{f \nu} = \frac{\chi_3 \chi_2}{f \nu}, \]  

(2.7)

\[ \tau = \frac{\chi_2 + \nu}{S_0} \frac{\eta + \chi_2}{\nu_2 + \eta + \chi_2}, \]  

(2.8)

and

\[ U_0 = \frac{(1 - f) \nu}{\eta + \chi_2} I_0 \text{ and } R_0 = \frac{f \nu}{\eta + \chi_2} I_0. \]  

(2.9)
Computation of the basic reproductive number $R_0$

By using the approach described in Diekmann et al.\textsuperscript{10} and by Van den Driessche and Watmough\textsuperscript{11} the basic reproductive number for model (2.1) is given by

$$R_0 = \frac{\tau S_0}{\nu} \left( 1 + \frac{\nu_2}{\eta} \right).$$

By using $\nu_2 = (1 - f) \nu$ and (2.8), we obtain

$$R_0 = \frac{\chi_2 + \nu}{\nu} \frac{\eta + \chi_2}{(1 - f) \nu + \eta + \chi_2} \left( 1 + \frac{(1 - f) \nu}{\eta} \right). \quad (2.10)$$


Outline

1. Introduction
2. Results
3. Numerical Simulations
Numerical Simulations

We can find multiple values of $\eta$, $\nu$ and $f$ which provide a good fit for the data. For application of our model, $\eta$, $\nu$ and $f$ must vary in a reasonable range. For the corona virus COVID-19 epidemic in Wuhan at its current stage, the values of $\eta$, $\nu$ and $f$ are not known. From preliminary information, we use the values

$$f = 0.8, \quad \eta = 1/7, \quad \nu = 1/7.$$ 

By using the formula (2.10) for the basic reproduction number, we obtain from the data for mainland China, that $R_0 = 4.13$ (respectively from the data for Hubei province $R_0 = 3.82$ and the data for Wuhan municipality $R_0 = 3.35$).
Turning point

Definition 3.1
We define the **turning point** $t_p$ as the time at which the function $t \to R(t)$ (red solid curve) (i.e., the curve of the non-cumulative reported infectious cases) reaches its maximum value.
Absence of major public health interventions

For example, in the figure below, the turning point $t_p$ is day 54, which corresponds to 23 February for Wuhan.

Figure: In this figure, we plot the graphs of $t \rightarrow CR(t)$ (black solid line), $t \rightarrow U(t)$ (blue solid line) and $t \rightarrow R(t)$ (red solid line).
Strong confinement measures

In the following, we take into account the fact that very strong isolation measures have been imposed for all China since 23 January. Specifically, since 23 January, families in China are required to stay at home. In order to take into account such a public intervention, we assume that the transmission of COVID-19 from infectious to susceptible individuals stopped after 25 January.

Remark 3.2

The private transportation was stopped in Wuhan after February 26. Furthermore, around 5 millions people left Wuhan just before the February 23 2020 to escape the epidemic. So the rate transmission should be divided by 2 after February 23 2020.
Therefore, we consider the following model: for $t \geq t_0$,

\[
\begin{align*}
S'(t) &= -\tau(t)S(t)[I(t) + U(t)], \\
I'(t) &= \tau(t)S(t)[I(t) + U(t)] - \nu I(t) \\
R'(t) &= \nu_1 I(t) - \eta R(t) \\
U'(t) &= \nu_2 I(t) - \eta U(t)
\end{align*}
\]  

(3.1)

where

\[
\tau(t) = \begin{cases} 
4.44 \times 10^{-08}, & \text{for } t \in [t_0, 25], \\
0, & \text{for } t > 25.
\end{cases}
\]  

(3.2)
Using the parameters $\chi_1, \chi_2, \chi_3$ obtained from the data for mainland China.
Using the parameters $\chi_1, \chi_2, \chi_3$ obtain from the data for Hubei province.
Using the parameters $\chi_1, \chi_2, \chi_3$ obtained from the data for Wuhan city.
Basic reproductive number
Another model for $\tau(t)$

The formula for $\tau(t)$ during the exponential decreasing phase was derived by a fitting procedure. The formula for $\tau(t)$ is

$$
\begin{align*}
\tau(t) &= \tau_0, \quad 0 \leq t \leq N, \\
\tau(t) &= \tau_0 \exp(-\mu(t - N)), \quad N < t.
\end{align*}
$$

(3.3)

The date $N$ is the **first day of the confinement** and the value of $\mu$ is the **intensity of the confinement**. The parameters $N$ and $\mu$ are chosen so that the cumulative reported cases in the numerical simulation of the epidemic aligns with the cumulative reported case data during a period of time after January 19. We choose $N = 25$ (January 25) for our simulations.
Figure: Graph of $\tau(t)$ with $N = 25$ (January 25) and $\mu = 0.16$. The transmission rate is effectively 0.0 after day 53 (February 22).
Predicting the epidemic in China

A: Data January 19 - January 31

μ = 0.16

B: Data January 19 - February 7

μ = 0.14

C: Data January 19 - February 14

μ = 0.14

D: Data January 19 - February 21

μ = 0.139

E: Data January 19 - February 28

μ = 0.139

F: Data January 19 - March 6

μ = 0.139
Predicting the Cumulative data for China with $f = 0.8$
Predicting the weekly data for China
Daily number of cases

The daily number of reported cases from the model can be obtained by computing the solution of the following equation:

\[
DR'(t) = \nu f I(t) - DR(t), \text{ for } t \geq t_0 \text{ and } DR(t_0) = DR_0. \tag{3.4}
\]
Predicting the weekly data in China
Multiple good fit simulations

We vary the time interval \([d_1, d_2]\) during which we use the data to obtain \(\chi_1\) and \(\chi_2\) by using an exponential fit. In the simulations below we vary the first day \(d_1\), the last day \(d_2\), \(N\) (date at which public intervention measures became effective) such that all possible sets of parameters \((d_1, d_2, N)\) will be considered. For each \((d_1, d_2, N)\) we evaluate \(\mu\) to obtain the best fit of the model to the data. We use the mean absolute deviation as the distance to data to evaluate the best fit to the data. We obtain a large number of best fit depending on \((d_1, d_2, N, f)\) and we plot the smallest mean absolute deviation \(\text{MAD}_{\text{min}}\). Then we plot all the best fit with mean absolute deviation between \(\text{MAD}_{\text{min}}\) and \(\text{MAD}_{\text{min}} + 5\).

Remark 3.3

The number 5 chosen in \(\text{MAD}_{\text{min}} + 5\) is questionable. We use this value for all the simulations since it gives sufficiently many runs that are fitting very well the data and which gives later a sufficiently large deviation.
Cumulative data for China until February 6 with $f = 0.6$
Cumulative data for China until February 20 with $f = 0.6$
Cumulative data for China until March 12 with $f = 0.6$
Daily data for China until February 6 with $f = 0.6$
Daily data for China until February 20 with $f = 0.6$
Daily data for China until March 12 with $f = 0.6$
Estimating the last day for COVID-19 outbreak in China

<table>
<thead>
<tr>
<th>Level of risk</th>
<th>10%</th>
<th>5%</th>
<th>1%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extinction date ((f = 0.8))</td>
<td>May 19</td>
<td>May 24</td>
<td>June 5</td>
</tr>
<tr>
<td>Extinction date ((f = 0.6))</td>
<td>May 25</td>
<td>May 31</td>
<td>June 12</td>
</tr>
<tr>
<td>Extinction date ((f = 0.4))</td>
<td>May 31</td>
<td>June 5</td>
<td>June 17</td>
</tr>
<tr>
<td>Extinction date ((f = 0.2))</td>
<td>June 7</td>
<td>June 12</td>
<td>June 24</td>
</tr>
</tbody>
</table>

Table: *In this table we record the last day of epidemic.*
Cumulative data for France until Mars 30 with $f = 0.4$
Cumulative data for France until April 20 with $f = 0.4$
Daily data for France until Mars 30 with $f = 0.4$
Daily data for France until April 20 with $f = 0.4$
References


- Z. Liu, P. Magal and G. Webb, Predicting the number of reported and unreported cases for the COVID-19 epidemic in China, South Korea, Italy, France, Germany and United Kingdom, *medRxiv*