Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Modeling the underestimation of COVID-19 infection

Ichiro Nakamoto \(^a\), Jilin Zhang \(^b\)

\(^a\) School of Internet Economics and Business, Fujian University of Technology, Fuzhou City, Fujian Province, China
\(^b\) Department of Computer Science and Mathematics, Fujian University of Technology, Fuzhou City, Fujian Province, China

**ABSTRACT**

Estimation of the undocumented cases of COVID-19 is critical for understanding the epidemic potential of the disease and informing pandemic response. The COVID-19 pandemic originated from a severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a virus similar to severe acute respiratory syndrome (SARS) that was formerly identified in 2003. The contagiousness, dynamics of the pathogen, and mobility of the general population incurred the occurrence of underestimation of infection (i.e., the unidentified cases and the gap with the identified cases) that was potentially substantial in magnitude, which was supposed to connect with subsequent cyclical outbreaks in practice. We employed a Susceptible-Infected-Removed-Contained (SIR-C) mathematical model to infer critical epidemiological characteristics associated with COVID-19, then asymptotically simulated the peak sizes and peak dates of the identified and unidentified cases, the underestimation, and the dynamics of the gap. The simulation outcomes indicated that unidentified peak dates in practice could predate the reported peak dates for a variable period of weeks or months. In comparison, the saturation sizes of infection remained at commensurate levels. The curve of the initial exponential-like outbreak for the undocumented cases would flatten when the gap between concurrent identified cases and unidentified cases decreased. The rate of non-pharmaceutical containment could impact the trend of disease transmission ceteris paribus, and the greater the rate the larger reduction of infections. When the rate reached a certain level of threshold, the undocumented infection (DOI) manifested that greater values of DOI were associated with greater peak sizes and greater peak dates for both documented and undocumented cases. Conditional on assumptions, calibration of DOI from 8 days to 18 days would increase the unidentified peak size by nearly 56% and the peak date by almost 18 days.

**Introduction**

The COVID-19 epidemic incurred viral respiratory diseases and pneumonia outbreaks worldwide. It was projected that before effective vaccines were available, the cyclical outbreak of the COVID-19 epidemic might persist for years, which would substantially impact the progress of returning to normal social activities \([1]\). By the end of November 2020, officially confirmed cases reached 60,534,526 globally, of which 1,426,101 deaths were identified, in more than 200 countries and areas based on the report by WHO \([2]\). This has potentially changed the past concept of local and sporadic outbreaks of epidemics to an extensive and compounding cycle of response and recovery \([3]\). A variety of unrivaled pharmaceutical quarantine and non-pharmaceutical containment measures has been employed in response to the COVID-19 pandemic \([4]\). While comprehensive adverse effects on the infected individuals are yet to be uncovered, countries and communities with early isolating of cases, curtailing person-to-person contacts, physical distancing, tracing close contact in combination with hygiene practices (e.g., utilization of masks or disinfectant) have obtained meaningful control. Consequently, the practice may be of interest to other areas where the cyclical outbreak of COVID-19 was observed and where the number of infected cases was underestimated \([4,5]\).

The dynamics of human-to-human transmission risk are correlated with multiple factors including attributes of the pathogen, measures of interventions, and other time-contingent factors that might impact the trajectory \([6]\). In the occurrence where complete information on the fluctuation of transmission is difficult to obtain, correct decision-making would be a challenge due to underestimation. Substantial asymptomatic and pre-symptomatic cases were reported in many countries, hence asymptotic inference might still provide insight for controlling the

\(\text{https://doi.org/10.1016/j.rinp.2021.104271}\)

Received 3 December 2020; Received in revised form 29 April 2021; Accepted 30 April 2021

Available online 13 May 2021

2211-3797/© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license
spread of the pandemic. Pharmaceutical quarantine and non-pharmaceutical measures such as physical distancing and contact tracing were used in many affected countries, which effectively influenced the trajectory of progression in shrinking the outbreak of COVID-19. For instance, patients were generally quarantined once identified COVID-19 positive, and residents were typically called upon to stay home when able and to perform physical distancing and hygiene practice when not. Control measures aimed at reducing close contact in the general population altered the trajectory of the epidemic [7,8].

Discrete transmission events and sporadic cluster infections in connection to a later explosion of spread had been identified during past SARS outbreaks and were not exceptional for the COVID-19 epidemic as well [9,10]. One concern was that sub-clinical transmission typically caused by asymptomatic or pre-symptomatic individuals might complicate the pandemic, causing correct and prompt response hard to be followed.

If infectious individuals were identified timely and effectively then the efficacy of measures could be increased significantly and thus concentrating effort on the quarantining of the most infectious cases would be at work than the case of mere random control [9]. However, such kind of mechanism was compromised during a pandemic in which transmission was established before the onset of symptoms or visible symptoms were not present [10]. The rate of spread hinged partially on the dynamics of the pathogen such as reproduction number and the duration of infection (DOI), which were defined as the average secondary infactions for each index case and the inverse of removal rate respectively [11]. The greater these values, the more difficult for the outbreak to settle down. Hence, modeling and simulation of the scale and timing of the underestimated outbreak were meaningful in gaining updated knowledge of pandemic evolution and providing insight for decision-making [12].

Past outbreaks of infectious diseases were characterized by remarkable differences in total cases of infection for the affected regions [13]. Preventing further spread and controlling the subsequent recurrent occurrence of pandemic remained to be a priority [14]. As of now, the pandemic is still not completely contained, partially due to reasons such as underestimated asymptomatic and pre-symptomatic cases [15]. The human-to-human transmission was confirmed in many countries since the first identified case [16]. Countries could migrate to be worldwide epicenters of outbreak unless substantial health interventions at a variety of levels were implemented. A large scale of asymptomatic and pre-symptomatic transmission in the absence of public health interventions would induce international seeding and subsequent local establishment of epidemics would be inevitable [17,18].

The purpose of our research included multiple facets. First, this study aimed to contribute to the literature by quantifying the gap of peak sizes and peak dates between the identified and unidentified cases, the magnitude of underestimation, and how the trajectory of the underestimation would be impacted by the measures of pharmaceutical quarantine, non-pharmaceutical containment and the dynamics of the pathogen. Second, verify how parameters of transmission impacted the spread of COVID-19 and sketch the potential outline of underestimation. Third, propose the concept of latent “plus alpha” effect relative to the disclosed data, and thereby provide updated understanding for COVID-19 related decision-making. Fourth, investigate the potential effects of non-pharmaceutical containment and pharmaceutical quarantine when uncertainties and risks of being infected remained the concerns. Finally, explore the asymptotic magnitude of impact by asymptomatic and pre-symptomatic patients. We employed a mathematical model to estimate the characteristics of underestimation of COVID-19, including the gap of peak sizes and gap of peak dates between the confirmed and unconfirmed cases, and simulation was performed to asymptotically quantify the inherent relationship.

**Methods**

**SIR-C model**

The dynamics of the transmission of COVID-19 over time are governed by the set of differential equations as follows:

\[
\frac{dS(t)}{dt} = - \alpha S(t)I(t) - \theta_0 S(t)
\]

(1)

\[
\frac{dI(t)}{dt} = \alpha S(t)I(t) - \beta I(t) - \theta_0 I(t) - \theta_1 I(t)
\]

(2)

\[
\frac{dR(t)}{dt} = \beta I(t) + \theta_0 S(t)
\]

(3)

\[
\frac{dC(t)}{dt} = (\delta_0 + \theta)I(t)
\]

(4)

\[
S(t) + I(t) + R(t) = 1
\]

(5)

This SIR-C model is derived from the standard SIR model. \(S(t)\), \(I(t)\) and \(R(t)\) denote the time-contingent fractions of the susceptible, the infected, and the removed cases respectively. To account for the potential effects on the population resulting from containment, one new time-contingent compartment \(C(t)\) is introduced, which is used to quantify the observable dynamic fraction. In contrast, \(I(t)\) is the fraction of infection incapable of being exactly identified in practice due to multiple reasons such as substantial pre-symptomatic and asymptomatic infections. \(\theta_0\) denotes the rate of non-pharmaceutical containment such as staying at home, physical distancing, and hence captures the uniform effect on the general population; in contrast, \(\theta\) typically captures the effect of rate of pharmaceutical quarantine on the infected individuals.

The assumptions are (1) \(I(t)\) maintains a positive and close relationship with the empirically identified naive cases \(C(t)\); (2) non-pharmaceutical measures influence the susceptible, the infected, and the removed individuals identically. Parameters \(\alpha\) and \(\beta\) quantify the rate of transmission and removal respectively. For this set of first-order differential systems, a complete symbolic solution for Eq. (1) is given by:

\[
S(t) = \exp(-\theta_0 t)\exp\left(-\alpha \int_0^t I(\tau) d\tau\right)
\]

(6)

For a high contagious disease like COVID-19, we assume that non-pharmaceutical containment measures impact the progression of the outbreak in a more dominant way such that the influence resulting from the transmission process is negligible. Consequently approximated solution for the susceptible simplifies to the following [19,20]:

\[
S(t) = \exp(-\theta_0 t)
\]

(7)

Substitute (7) into Eq. (2) and derive the general solution for the underestimated fraction of the infected individuals as:

\[
I(t) = \delta_0 \exp\left(\frac{\alpha}{\delta_0} \left(1 - e^{-\delta_0 t}\right) - (\theta_0 + \theta + \beta) t\right) = \delta_0 \exp[I(0)] \cdot \frac{e^{-\delta_0 t}}{\delta_0} \text{ if } \delta_0 \neq 0
\]

(8)

and:

\[
I(t) = \delta_0 \exp\left((\alpha - \beta - \theta) t\right) \text{ if } \delta_0 = 0
\]

(9)

(8) captures the effect in which non-pharmaceutical containment exerts non-negligible effect, in contrast, (9) catches the effect in which non-pharmaceutical containment exerts trivial or no effect at all. In Eq. (8), the trend of the underestimated cases \(I(t)\) hinges on multiple factors, including the rate of transmission, the rate of removal, the rate of non-pharmaceutical containment, the rate of pharmaceutical containment, and the time. In eq. (8) assume \(I(t) = \frac{\delta_0}{\alpha} (1 - e^{\theta_0 t}) - (\theta_0 + \theta + \beta) t\), hence if F.O.C \(\frac{\delta_0}{\alpha} > 0\), then the underestimated cases \(I(t)\) increases over time; comparatively, if F.O.C \(\frac{\delta_0}{\alpha} < 0\) the opposite trend of \(I(t)\) is observed. At
the locus where $\Delta \omega = 0$, $I(t)$ reaches its saturation (i.e., peak size), from which we derive the relationship depicted in Eq. (10):

$$t' = \text{argmax}(I) = \frac{1}{\theta_0} \log \left( \frac{a}{\beta + \theta_0 + \theta} \right)$$

(10)

where $R_c = e^{\frac{a}{\theta_0}}$ denotes the effective reproduction number when pharmaceutical or non-pharmaceutical control measures are enacted and $R_0 = \theta_0$ the basic reproduction number when no containment measures are implemented. Consequently, when the rate of non-pharmaceutical containment increases ceteris paribus, $R_c$ decreases accordingly. A similar trend applies to the case when the rate of pharmaceutical quarantine increases ceteris paribus. Both types of reproduction numbers capture the subsequent cases of infection for each index case on average before removal from the transmission. Since both rates are no less than zero, the effective reproduction number is generally smaller than the basic reproduction number. Reduced reproduction number is associated with slowing down or decaying spread of the pathogen, and when the numeric value is less than one the disease will cease transmission and the pandemic will be contained ultimately. When the rate of non-pharmaceutical containment $\theta_0$ keeps extremely low and approximates infinitely to zero, $t'$ potentially approximates to a considerable value, causing the spread of disease hard to diminish. When the rate of non-pharmaceutical containment is exceptionally prompt, the value of $\theta_0$ would increase and thus the ratio between effective and basic reproduction number would decrease, in such a case, $t'$ can be smaller and earlier arrival of peak dates is expected. After $t'$ is passed, the pandemic would diminish over time. Similar analysis can apply to the scenario concerning the change in $\theta$. The maximum size of the underestimated cases is governed by:

$$I_{\text{est}}(t') = I_0 \exp \left( \frac{a}{\theta_0} (t') \right) = I_0 \exp \left( \frac{a}{\theta_0} \left( 1 - e^{-\theta_0 t'} \right) - (\theta_0 + \theta + \beta) t' \right)$$

(11)

On the other hand, in Eq. (9), when no-pharmaceutical containment is implemented, the pandemic size would increase over time. In summary, non-pharmaceutical containment plays a more critical role in determining the timing of peak dates than pharmaceutical quarantine. To establish the parameters, we employed the non-linear least-squares of Levenberg–Marquardt [20] and a Taylor-series expansion approaching the halfway point $t_0 = \frac{t'}{2}$:

$$I(t) = I(t_0) + \frac{1}{2} I'(t_0)(t-t_0)^2 + \frac{1}{6} I''(t_0)(t-t_0)^3 + O(\Delta^4)$$

(12)

where $(t-t_0)$ is set as the step size for the asymptotic computation. There are two types of intrinsic errors involved in this asymptotic approach: First, the truncation of the Taylor series incurs an error that limits the ultimate accuracy of the model. Second, utilization of the approximation of $I(t_0)$ given by the previous iteration when computing $I(t)$ generates an additional disturbance that may accumulate over successive iterations, and eventually affects the quantitative sensitivity of the method. The approximate number of unconfirmed infections at time $t$ is determined by:

$$I(t) = \tilde{I}(t) = I_0 \exp \left( \frac{a R_c^{-0.5}(1 + 0.5 log R_c) - (\theta_0 + \theta + \beta) t - 0.5 \theta_0 R_c^{-0.5}}{5} \right)$$

(13)

And the inherent association between time-adjacent unidentified infections is given by:

$$\tilde{I}(t+1) = \tilde{I}(t) \exp \left( \frac{a R_c^{-0.5}(1 + 0.5 log R_c) - (\theta_0 + \theta + \beta)}{5 \theta_0 R_c^{-0.5}} \right) - 0.5 \theta_0 R_c^{-0.5} (1 + 2 t)$$

(14)

In contrast, confirmed cases can asymptotically be decided by the relation of below [19]:

$$\tilde{C}(t) = \tilde{I}_0 \frac{a}{\theta_0} \left( \theta + \beta_0 \right) \left( \frac{R_c}{\theta_0} \right)^{0.5} \exp \left( \frac{a^2 \theta_0 R_c^{-0.5}}{8 \theta_0^2 R_c^{-1}} \right) \left( \left( 1 - e^{-\theta_0 t} \right) - 2a(\theta_0 t - 1) \right)$$

$$- \theta_0 \theta_0 R_c^{-0.5} \left( 1 - e^{-2 \theta_0 t} \right) - 2 \theta_0 \theta_0 R_c^{-0.5} \left( 1 - e^{-\theta_0 t} \right)$$

(15)

Schema of model scenarios

The typical scenarios in which containment measures employed in response to an epidemic can be summarized in their simplest arrangement (Fig. 1). This applies to settings in which the healthcare capacity is of relative tininess or the equilibrium of healthcare is not achieved. Treatment of infected cases is differentiated depending on the infectiousness of the pathogen, the symptoms of the patients, and the accommodation capacity of the healthcare system. Privilege is given to severely symptomatic cases and thus immediate pharmaceutical quarantine is needed for these patients. In contrast, asymptomatic, pre-symptomatic, or mild-symptom patients are typically called upon to comply with non-pharmaceutical containment strategy by staying home or at locally assigned places. Pre-symptomatic, mild, or asymptomatic patients can either deteriorate to severe infections or recover from the disease attributable to the dynamics of COVID-19. Severe symptomatic patients might recover or die from the pharmaceutical quarantine by chance.

Simulation

Simulation scenario

To simulate, we assumed that the first case was identified on Jan 22, 2020, and the data were continuously collected until May 3, 2020. The general population exploited measures such as social distancing and hygiene practice since April 7 and the outbreak experienced slow growth (daily increment not exceeded one thousand) in terms of infected cases up to the points where data were observable. The accumulated number of cases was set to 15,000 on May 3, 2020, 100 days since the first case was officially reported [19]. The number of the total population was equal to 125.96 million at the time of analysis. In subsequent scenarios, we assumed response patterns altered in pharmaceutical quarantine and non-pharmaceutical containment respectively starting April 7, and simulated the trajectory of underestimation beyond this point contingent on the least-squares fitted parameters ceteris paribus. To scrutinize the sensitivity, we explored how heterogeneity of DOI affected the trajectory of the outbreak by increasing its value from 8 to 18 days with a step of 2-day increment.

Simulation parameters

To estimate and simulate, part of the parameters attained from
The inherent containment effect was reflected in the fitting outcome, and analysis was performed to evaluate the fitting between discretely observed cases and the fitted curves. Past research had identified that the mean basic reproduction number of SARS-CoV-2 was distributed in a range between 2 and 7 \([11,21]\). The practice in China had also shown that a value of 6.2 for the basic reproduction number qualitatively worked well \([19]\). Conditional on the given parameters, the effective reproduction number was estimated to be a value of 2.48. The virus was highly infectious and had a long but still uncertain transmission interval as partial infections were ascertained to establish before the onset of symptoms. According to the reports by WHO, the incubation period of COVID-19 ranged from 1 to 14 days with a median appraisal of 5 days. The removal rate consisted of the sum of recovery and mortality rate, it was known that variation existed in the removal procedure for the infected individuals contingent on the status of sickness and the potential health problems resulting from COVID-19. Simultaneously, heterogeneity was present in the distribution of the duration of infection, which was defined as the inverse of the removal rate. Prior study found that the variation of DOI was estimated to be a fuzzy range of \([3 \text{ days}, 20 \text{ days}]\) \([19]\). However, this duration could potentially also vary in a narrower interval of \([5 \text{ days}, 20 \text{ days}]\) \([20]\). In the simulation scenario presented in this study, we utilized an interval of 16 days. How the trajectory of underestimation was impacted by the change of DOI was investigated in the sensitivity section. In the scenario illustrated, the rate of pharmaceutical quarantine was estimated to be a small value close to the boundary, and the rate of non-pharmaceutical containment was estimated to be \(0.016 \text{ day}^{-1}\).

### Simulation results

In the first scenario, we simulated the least-squares fitted outcomes for the confirmed and unconfirmed cases with presumed effective reproduction number as well as the duration of infection (presented in Table 1) by using the aforementioned outbreak data from late January to early May (Fig. 2). In the subsequent scenarios, we then inherited all the fitted parameters except the rate of containment by utilizing an increment of one-day step. We calibrated the changes concerning each case of containment starting from April 7 (almost 75 days after the identification of the first case) to forecast the underestimation trajectory of the infectious process.

---

**Table 1**

Simulation Parameters of the SIR-C model.

| Parameters | Value | References |
|------------|-------|------------|
| Base reproduction number without any containment: \(R_0 = \frac{\alpha}{\beta}\) | 6.2 | Maier et al \([19]\) |
| Transmission rate: \([\text{day}^{-1}]\) | 0.39 | Calculated from \(R_0\) and \(\beta\) |
| Duration of infection(DOI): \(\frac{1}{\beta}\) [\text{day}] | 16 | Assumed |
| Pharmaceutical quarantine rate: \(\theta\) | 7-6e-10 | Least-squares Fitted |
| Non-pharmaceutical containment rate: \(\theta_0\) | 0.016 | Least-squares Fitted |
| Effective reproduction number with containment: \(R_c = \frac{\alpha}{\beta + \theta + \theta_0}\) | 2.48 | Least-squares Fitted |

---

Published literature was presented in Table 1, for the remaining we least-squares fitted in calibration to previous assumptions.
outbreak beyond that point (Fig. 3; Fig. 4).

The estimates implied that the peak dates of the unconfirmed and the confirmed curves did not coincide. The dates of peak for the former case occurred around May 6, in contrast, it was delayed almost two-and-half months for the latter, which reached its peak around 21 July. Hence, real peak dates in practice might arrive earlier than the peak dates that were reported. Comparatively, the ultimate saturation size in magnitude for both was at a commensurate level, indicating that the approximate number of infected individuals would potentially reach 34 thousand cases. In between the two peaks, identified cases continued to increase but unidentified cases followed the trend of decrease. Prior to the saturation of the unidentified curve, the size of underestimation was increasing over time but the opposite was observed after the saturation was exceeded and thereafter the pandemic faded out gradually or

Fig. 2. Estimates of confirmed $C(t)$ and unconfirmed $I(t)$ cases of COVID-19 from late January to early May 2020 Parameters estimates were performed in calibration to assumed cases predating May 3 based on least-squares fitting. Basic reproduction number without containment and duration of infection were set to be 6.2 and 16 days respectively. The effective reproduction number with containment was estimated to be 2.48. The model predicted the inherent underestimation in terms of size and timing between the confirmed and the unconfirmed cases. The former peaked around July 21. In contrast, the latter saturated around May 6, almost two-and-half months earlier. Peak sizes in magnitude for both were at the commensurate level, approximated to 34,000 cases.

Fig. 3. Simulation of underestimation of COVID-19 with changes in the rate of non-pharmaceutical containment All parameters in the simulations were time invariant and inherited from the baseline least-squared fitted values except the rate of non-pharmaceutical containment. We assumed rate changes took place two-and-half months later on April 7th, 2020. The data inside the arrows denoted the dates of saturation and the corresponding magnitude.
became the next starting point of recurrent outbreaks, as would be corrected by the dynamics of COVID-19. The simulation indicated that at a certain point, the curve of the confirmed cases shifted to a sub-exponential or algebraic growth until saturation was reached. It was projected that the trajectory of the initial exponential-like outbreak would flatten after the underestimation size started to decline (Fig. 2).

Simulation of underestimation when the change in non-pharmaceutical containment occurred

To simulate the association between the trajectory of undocumented cases and the change of rate of non-pharmaceutical containment ceteris paribus, we suppose an occurrence of rate change due to reasons such as enhanced risk awareness of the public or implementation of containment strategies by the administration starting April 7, almost two and half months after the first case was identified. In comparison with the previous baseline simulation outcome, in the case in which the rate of non-pharmaceutical containment increased by 1%, the date of saturation was estimated to antedate nearly 18 days compared to no concurrent change of containment measures, shifting to an earlier date of April 18, which was accompanied by a reduced peak size of 22,590 cases (almost one-third decrease in size). Hence, for highly infectious diseases like COVID-19, implementation of non-pharmaceutical control measures could achieve meaningful effects. A greater downsizing trend was observed if the rate of non-pharmaceutical containment increased by 5%, generating almost three weeks earlier advent of the peak date and nearly 37⋅4% of reduction in peak size to 20,993 cases. The slowing-down effect would shift to a decaying trend when the rate of non-pharmaceutical containment increased approximately by 20% or more, causing a faster pace of dying out of the outbreak. To achieve a significant reduction in the number of unidentified cases, more stringent containment had to be enacted (Fig. 3).

Simulation of underestimation when the change in pharmaceutical containment occurred

In the scenario where the rate of pharmaceutical quarantine changed, the reduction in effective reproduction number would potentially cause the spread of pandemics to slow down or extinguish. This hinged on multiple factors including the capacity, quality, and promptness of healthcare. Simulation conditional on this type of scenario illustrated the potential change in peak sizes and peak dates, assuming the rate of change commenced on April 7 (two-and-half months later). If the rate of pharmaceutical quarantine grew by 10-folds compared with the baseline, it was predicted that the peak date would arrive roughly 17 days earlier with a shrinking peak size of 23,144 cases. This remained at a commensurate level of curtailment compared with the non-pharmaceutical case when one percentage of containment change was observed. The more increment of the rate change, the greater curtailing effect on the peak sizes and the earlier arrival of peak dates. When the rate of pharmaceutical quarantine reached 0.03 per day or larger, a decaying trend of the unidentified cases would take effect and thus a speedier diminishing of the pandemic could be expected (Fig. 4). Note that the initial least-squares estimate fluctuated around the boundary value, therefore this suggested a remarkable improvement in the capacity of healthcare. It was relatively easier for non-pharmaceutical quarantine than for pharmaceutical containment to achieve an effective reproduction number less than one in the case where the capacity of healthcare was supposed to be relatively stable within a certain period.

Sensitivity analysis

To evaluate how heterogeneity of DOI impacted the trajectories of underestimated and observed cases, we calibrated DOI size from 8 to 18 days following a step of two-day increment (Fig. 5). Generally, the greater values of DOI, the longer period of exposure to the infection, and the larger values of peak sizes and peak dates. This applied to the trajectories of the confirmed and the unconfirmed identically. In comparison with the baseline DOI of 16 days, the increase of the unidentified peak sizes was 13,399 (or almost 56% growth) and the peak date relocated from April 22 to May 10 (nearly 18 days of delay) when DOI changed from 8 days to 18 days. Similar trends were observed for the curves of confirmed cases. By late July, all the curves of confirmed cases approximated to their saturation in magnitude asymptotically. By asymptotic computation, every two-day increase of DOI linearly postponed the peak dates by three to four days; however, the growth in peak sizes did not present a similar linear trend. Change of DOI from 8 days to
10 days was correlated with a growth of 1529 cases, then rose to 2103 on the subsequent step, thereafter 2656, 3232, and 3879 respectively. Thus, when the DOI was trending in a similar vein, the saturation size of the outbreak was forecast to be much larger. In contrast, when the DOI was relatively short, the spread of disease would be contained more rapidly as the overall size of the outbreak was relatively small. At low levels of DOI, the peak sizes of the unconfirmed cases potentially surpassed the sizes of the confirmed sizes, and this trend shrank with the increase in DOI. Uncertainty in the dynamics of the pathogen might partially contribute to this, however, the peak sizes of both are generally maintained at comparable levels.

Discussions

The contribution of our study was manifolds. First, we both quantitatively and qualitatively identified the asymptotic underestimation characteristics of COVID-19 in terms of peak dates and peak sizes, their relationship, and how it was influenced by the size of DOI. Second, we investigated the effect where the change of either non-pharmaceutical or pharmaceutical measures ceteris paribus impacted the trend of the pandemic and explored the effort needed to achieve for the sake of meaningful containment of the risk transmission. Third, simulation in compliance with calibrated DOI parameters sketched the linear dynamic outline for peak dates and non-linear dynamic profile for peak sizes respectively. Finally, we asymptotically outlined the magnitude of impact by non-captured factors including asymptomatic and pre-symptomatic patients that might be difficult to follow in practice.

We found the quantitative correlation between the identified and unidentified cases regarding peak dates, peak sizes, and trends. Conditional on the assumed parameters and scenarios, the dates of saturation for the unidentified curve did not coincide with the identified curve, implying a fundamental episode of underestimation and gap. In the baseline simulation, the undocumented peak date predated by nearly two-and-half months. In contrast, the peak sizes of both curves were at a commensurate level. The magnitude of underestimation hinged on multiple factors such as the rate of transmission, the rate of removal, the rate of non-pharmaceutical containment, the rate of pharmaceutical containment, and the time. Sensitivity analysis was performed through the channel of DOI and the results suggested that DOI could substantially impact the dynamics of underestimation.

Prior to the saturation of the unidentified cases, the gap in size increased over time; in contrast, the opposite was observed when the unidentified peak was reached. A lengthier duration postponed the peak dates and enlarged the peak sizes accordingly, causing the outbreak more difficult and more costly to settle down. The changes in the rate of pharmaceutical quarantine and non-pharmaceutical containment could impact the trend of unidentified curves and the spread of the pathogen. In settings where the capacity of healthcare was presumed with relative stability, non-pharmaceutical containment such as staying home or at local places, physical distancing, and tracing close contacts was expected to be enacted with priority. Our simulations implied that non-pharmaceutical containment could potentially achieve meaningful effects when effectively followed. We projected that under a duration of infection interval [8d,18d], the peak sizes of COVID-19 outbreak would be of a range from 24 to 38 thousand and the peak dates of which would potentially fluctuate between late April and mid of May (between two-and-half months and three months after the initial case). A comparable level of peak sizes could be expected, and the dates of saturation in practice presented a delaying effect and converged until late May to late July depending on the strength of the infectiousness.

The trajectory of the epidemic reflected the interactions of containment strategies and the transmission attributes of pathogen coupled with changed behaviors in response to the outbreak. The heterogeneity in confirmed and unconfirmed cases, rate of pharmaceutical quarantine, and rate of non-pharmaceutical containment did influence the trajectory of the outbreak. And this heterogeneity prospectively played critical roles during the spread of COVID-19. We extrapolated that confirmed cases would saturate in late July or around that time and unidentified cases would reach the peak before that time, almost two-and-half months earlier. More stringent containment was of necessity to acquire a greater diminishing effect on the outbreak.

The forecasts could potentially be impacted by other omitted factors
such as capacity (e.g., pathogen testing capacity) and quality of healthcare, changing biological effects, social (e.g., the degree of a shared sense of crisis in population) and spatial heterogeneity [22,23]. For instance, spatial variations such as the structure of population mixing were found to exist in countries like Japan, as the prefectures close to Tokyo were highly populated. And for these communities, more stringent quarantine adherence was needed to achieve a more flattened curve [24]. In the case where the resistance to drastic disease-control measures was at work, the curve of infections would fluctuate with considerable uncertainty. We estimated the scenarios where pharmaceutical quarantine and non-pharmaceutical containment impacted the underestimation trends of the COVID-19 pandemic. Preventing further transmission by decreasing the potential channels of infections and effective reproduction number. However, the feasibility of these strategies would be compromised if the number of infected patients reached a certain threshold of the total population. One critical concern was how asymptomatic, pre-symptomatic and mild-symptomatic individuals responded to COVID-19 to fulfill the feasibility of prevention of cyclical outbreaks. Other factors that might impact the trajectory of the COVID-19 pandemic were not accounted for in our model as well. In the case where presumed conditions were relaxed, the likelihood in connection with differentiated scenarios would be present. The simulation contingent on the fitted parameters indicated that the unidentified cases were generally greater than the reported cases before the undocumented saturation. We investigated a range of scenarios where the heterogeneity of containment changed the trajectory of the underestimation of COVID-19. As uncertainty existed in factors such as the interval of infectiousness for the asymptomatic individuals, non-pharmaceutical containment would be of importance to enhance the effect of control in combination with pharmaceutical quarantine and other measures. The model could be modified to incorporate other unaccounted factors impacting the dynamics of the transmission, which might implicitly alter the trend of the COVID-19 outbreak [7,8]. As clinical knowledge of this novel pathogen and its dynamics accrues, it is feasible that the outcomes will improve. It therefore will be of concern to revise these estimates as the epidemics continue to unfold [25].

The inherent relationship between the confirmed and unconfirmed cases was asymptotically identified and hence its precision was up to the point where the inherent approach applied. The analysis could not differentiate the efficacy of specific containment, nor could it differentiate the effect for the asymptomatic, pre-symptomatic, symptomatic, and infection-route-unknown compartment respectively. However, it reinforced other findings by showing that pharmaceutical and non-pharmaceutical control, when timely and successfully implemented, were feasible in decelerating or even diminishing the spread of the pandemic with scaled-down peak dates and peak sizes. This could be informative when it was too lengthy for the transmission to converge and when the motivation was to asymptotically interpret the “plus a” effect of the outbreak. The contribution of this study partially lied in the point that we asymptotically identified the quantitative and qualitative nature of the underestimation. Containment measures were preferred to be performed at an earlier stage and the effectiveness has been identified in other countries and past outbreaks. Changes in behaviors were observed in response to the epidemic where asymptomatic, mild- and pre-symptomatic patients were not readily ascertained. This might be of great importance for developing control measures for recurrent outbreaks of COVID-19. Countries with a lower capacity of healthcare are more likely to understate the outbreak sizes, and transmission chains connecting with extensive spread can not be correctly established [2]. Multiple important lessons emerged in that integration of healthcare services across disparate sectors disproportionately amplified the resilience to respond to the pandemic and misinformation remained to be unresolved [26]. As the capacity of healthcare became overwhelmed, the coordination between local healthcare providers and administration could be remarkably challenged. To avoid recurrent outbreaks of SARS-CoV-2 after the initial wave, it was critical that the capacities of healthcare did not exceed its saturation absent other interventions. Discussions thus far hovered over the comparatively low capacity of COVID-19 tests in some countries albeit the effectiveness of pathogen testing has been confirmed in other countries [27]. The dynamics of COVID-19 caused the intrinsic existence of undetected individuals being infected by SARS-CoV-2. It is of priority to deliver effort according to WHO’s recommendations of a combination of measures: rapid and adequate diagnosis, immediate isolation of confirmed cases, rigorous tracking, and precautionary self-isolation of close contacts. Pharmaceutical quarantine and non-pharmaceutical control have been implemented by a variety of countries to prevent further spread and helped reduce the transmission [18].

While it was critical to balance the control of spread and economic impact from COVID-19, when this was not feasible then priority was needed [28,29]. The transmission of COVID-19 was supposed to be more infectious than past SARS outbreaks and the likelihood of recurrence was high [30]. Infectiousness was estimated to peak on or before symptom onset, therefore, many infections potentially took place in an unnoticeable mechanism. Disease control measures should be adjusted to account for substantial subclinical transmission and underestimation [31]. COVID-19 had more severe pre-symptomatic and asymptomatic infections than influenza A and SARS, clinical studies delivered effort to evaluate the viremia and the dynamics. Heterogeneity might exist in the between-communities measures as well as in the responses to containment, growth in sporadic events could gradually overwhelm the contact tracing system, leading to the necessity for broader-scale social interventions. Ongoing data collection, epidemiological analysis alongside clinical research on COVID-19 are therefore essential parts of assessing the impacts of measures [28]. Prolonged or intermittent non-pharmaceutical measures such as social distancing are of necessity before effective vaccinations are readily available for the general population [30]. This might potentially redefine the daily routine that we are experiencing by this time [22].

It was shown that the crude size of the epidemic could be roughly estimated based on its distinguished dynamics. The possible trajectories of an outbreak hinged on the levels of public health interventions such as quarantine and precautionary measures [32,32]. The uncertainty of the peak sizes and dates was attributable to multiple factors, including stochasticity of early dynamics, heterogeneity of contact patterns, spatial variation, and dynamics of the epidemiological parameters [28]. SARS was eventually contained through prompt isolation, strict quarantine of contacts, and top-down enforcement of community containment. The first waves of COVID-19 outbreak have been controlled in some countries to date. Striking similarities between SARS and COVID-19 were identified, but more difference was of necessity to be ascertained. Even if traditional public health measures are incapable to completely contain the outbreak of COVID-19, they will still be operative in reducing the peak incidence and mortality [33]. While rigorous control policies were associated with detained growth in cases, in the extremity where stay-at-home restrictions were unlikely to be the one-shot deal, a gradual approach to restrictive measures might be of concern [22]. It will be particularly meaningful to design measures for long-term pharmaceutical and non-pharmaceutical control of COVID-19, along with large-scale testing, contact tracing, and isolation. Research should concentrate on refining specific estimates of susceptibility to infection, which is instrumental to appraise the impact of these strategies [28]. In the absence of effective measures, the pandemic spread widely, and thus considerable efforts at a variety of levels were essential for the outbreak to settle down. Such efforts will be critical to quench local outbreaks and reduce the risk of further global dissemination [34,35]. Protective measures could compromise the effectiveness as cases accrue, the impact of underestimation tended to be easily misevaluated. The dynamics of COVID-19 could alter significantly attributable to the recent identification of multiple mutations of the pathogen, therefore optimization of the treatment and non-
pharmaceutical measures are still the public concerns before updated vaccines are available [36,37].

In conclusion, the dynamics of COVID-19 incurred the intrinsic gap in the peak sizes and the peak dates between confirmed and unconfirmed cases. Evaluation of underestimated quantities could provide insight for meaningful decision-making when uncertainties and risks are not ruled out. Interventions based on pharmaceutical quarantine and non-pharmaceutical containment manifest a strong potential to lessen the peak sizes and the peak dates of the COVID-19 outbreak. Lowering and flattening of the pandemic peak is particularly important, as this reduces the acute pressure on the healthcare system as well as on the society as a whole. When it is difficult to pinpoint the next epidemic, the measures taken as of today would matter [16,38].

Declaration of Competing Interest

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

Acknowledgments

This research was partly funded by the China Ministry of Education Industry and Education Cooperation Project [grant number 202020041005] and Fujian Province Natural Science Funding Project [grant number 2020 J01892]. We thank the helpful comments from Sheng Wang.

References

[1] Deslandes A, Beriti V, Tandjoaquim-Lambotte Y, et al. SARS-COV-2 was already spreading in France before late December 2019. Antimicrob Agents Chemother 2020:3:106006. https://doi.org/10.1128/aac.2020.106006.
[2] WHO coronavirus report. https://www.who.int/emergencies/diseases/novel-coronavirus-2019?ncid=CyiWkCAAIIL-RAeEiWgoAIGC1wlYhs4kvP-D7sUHb6sX54N00pl2LUUs6s3B99P9D1sM4dAHAr1B0Cm1C6qAvD1BwL.
[3] Bedford J, Farrar J, Ihekweazu C, Kang G, Koopmans M, Neningang J. A new twenty-first century science for effective epidemic response. Nature 2019:575: 130-6. https://doi.org/10.1038/s41586-019-1717-y.
[4] Chen S, Yang J, Yang W, Wang C, Barnighausen T. COVID-19 control in China during mass population movements at New Year. Lancet 2020:395:766-7.
[5] Zhu N, Zhang D, Wang W, et al. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med 2020; published Feb 20. 10.1056/NEJMoa2001017.
[6] Rong XM, Yang L, Chu HD, Fan M. Effect of delay in diagnosis on transmission of COVID-19. Math Biosci Eng. 2020 Mar 11(17):2725-40. https://doi.org/10.3934/mbe.2020149.
[7] Li Q, Guan X, Wu P, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus Infection and Control Measures. Science 2020. 2020 Mar 21; 395 (10228):931-934.DOI:https://doi.org/10.1126/science.abb5793.
[8] Hellewell J, Abbott S, Gimma A, et al. Feasibility of controlling COVID-19 outbreaks in Europe through mass population movements at New Year. Lancet Infect Dis. 2020 Mar 5. pii: S1473-3099(20)30207-7.
[9] Wilder-Smith A, Chiew CJ, Lee VJ. Can we contain the COVID-19 outbreak with the same measures as for SARS? Lancet Infect Dis. 2020 Mar 31; 20 (10228):931-934.DOI:https://doi.org/10.1016/S1473-3099(20)30567-5.
[10] Peak CM, Childs LM, Grad YH, Buckee CO. Comparing nonpharmaceutical interventions for containing emerging epidemics. Proc Natl Acad Sci USA 2017; 114:4023-8. https://doi.org/10.1073/pnas.1614381114.
[11] Zhao S, Lin QY, Ran JJ, et al. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: a data-driven analysis in the early phase of the outbreak. Int J Infect Dis 2020;92:214-7. https://doi.org/10.1016/j.ijid.2020.01.050.
[12] Fang Y, Nie Y, Penny M. Transmission dynamics of the COVID-19 outbreak and effectiveness of government interventions: A data-driven analysis. J Med Virol. 2020. https://doi.org/10.1002/jmv.25756.
[13] Gurdasani D, Gudlaunnsdottir H, on the feasibility of simulation models in informing pandemic responses. Lancet Global Health. Published: April 30, 2020. 10.1016/s2214-109x (20)30219-9.