

RESEARCH

Open Access



Assessment of the fatality rate and transmissibility taking account of undetected cases during an unprecedented COVID-19 surge in Taiwan

Hsiang-Yu Yuan^{1,2*†}, M. Pear Hossain^{1†}, Tzai-Hung Wen³ and Ming-Jiuh Wang⁴

Abstract

Background: During the COVID-19 outbreak in Taiwan between May 11 and June 20, 2021, the observed fatality rate (FR) was 5.3%, higher than the global average at 2.1%. The high number of reported deaths suggests that many patients were not treated promptly or effectively. However, many unexplained deaths were subsequently identified as cases, indicating a few undetected cases, resulting in a higher estimate of FR. Whether the true FR is exceedingly high and what factors determine the detection of cases remain unknown. Estimating the true number of total infected cases (i.e. including undetected cases) can allow an accurate estimation of FR and effective reproduction number (R_t).

Methods: We aimed at quantifying the time-varying FR and R_t using the estimated true numbers of cases; and, exploring the relationship between the true case number and test and trace data. After adjusting for reporting delays, we developed a model to estimate the number of undetected cases using reported deaths that were and were not previously detected. The daily FR and R_t were calculated using the true number of cases. Afterwards, a logistic regression model was used to assess the impact of daily testing and tracing data on the detection ratio of deaths.

Results: The estimated true daily case number at the peak of the outbreak on May 22 was 897, which was 24.3% higher than the reported number, but the difference became less than 4% on June 9 and afterwards. After taking account of undetected cases, our estimated mean FR (4.7%) was still high but the daily rate showed a large decrease from 6.5% on May 19 to 2.8% on June 6. R_t reached a maximum value of 6.4 on May 11, compared to 6.0 estimated using the reported case number. The decreasing proportion of undetected cases was found to be associated with the increases in the ratio of the number of tests conducted to reported cases, and the proportion of cases that are contact traced before symptom onset.

Conclusions: Increasing testing capacity and contact tracing coverage without delays not only improve parameter estimation by reducing hidden cases but may also reduce fatality rates.

Keywords: COVID-19, Hidden cases, Fatality rate, Reproduction number, Test and trace

Introduction

Knowing the actual number of coronavirus disease 2019 (COVID-19) cases throughout an outbreak is critical in order to provide an accurate estimate of epidemiological parameters such as the fatality rate (FR) and effective reproduction number (R_t). These parameters aid in

*Correspondence: sean.yuan@cityu.edu.hk

†Hsiang-Yu Yuan and M. Pear Hossain contributed equally to this article

¹ Department of Biomedical Sciences, Jockey Club College of Veterinary Medicine and Life Sciences, City University of Hong Kong, Hong Kong, SAR, China

Full list of author information is available at the end of the article



making proper public health decisions, assessing health care system performance, and predicting the trend of COVID-19 spread. However, the number of undetected cases can be large and may vary during an outbreak. Limited capacities of contact tracing and testing often result in underestimation of true infections [1, 2]. The proportion of undetected cases may reduce after such capacities improve. Hence, estimating this constantly changing proportion of undetected cases throughout an outbreak is important.

After several months of zero confirmed community-acquired cases, quarantine exemption for flight crews, and super spreader events in tea parlors in Wanhua in Taipei in late April and early May 2021, triggered a fresh wave of local spread of the Alpha variant [3]. This resulted in 14,005 total reported cases between May 11 and June 20, 2021 [4]. Approximately 5% of cases resulted in death. This case fatality rate (CFR) was apparently higher than the average global rate (obtained by dividing the total number of deaths by the total number of cases worldwide), which has been consistently below 2.5% since November 16, 2020 [5]. Whether this high CFR was mainly due to insufficient hospital capacity and treatment, or a massive proportion of undetected cases is still not known.

Early in the outbreak, testing capacity was insufficient to cope with the rising cases among initial transmission clusters. The daily number of new cases grew to more than 200 within a week and continued to increase until reaching a plateau at the end of May 2021 (i.e., 596 cases on average per day from May 22 to 28). Because of the emerging outbreak, Taiwan had been under Level 2 alert since May 11, 2021 [6], followed by escalation to Level 3 restrictions on May 19, 2021 [7]. Under the stricter restrictions, people were required to wear masks outdoors; gatherings of more than four people indoors and more than nine people outdoors were banned; and all schools were closed. Social distancing measures reduced individual mobility [8–10] and effectively lowered R_t . At the same time, the daily number of tests conducted continued to increase, presumably allowing more cases to be identified.

During the outbreak, many confirmed cases failed to be detected when alive but were tested because of their death, indicating that a certain number of undetected cases existed. The number of undetected cases who eventually died (referred to as *undetected deaths*), together with the number of deaths who were known to have COVID-19 (referred to as *detected deaths*), can be used to infer the proportion of undetected cases if their fatality rates are known. Presumably, the probability of death among undetected cases is similar to that among

detected cases during the early period of the outbreak when hospital capacity or treatment is not sufficient.

Although knowing the numbers of detected and undetected deaths helps to estimate the proportion of undetected cases and hence to guide interventions, a challenge exists that many deaths from infection usually happen several weeks after symptom onset. This highlights the importance of early estimation of the true number of total cases and the number of associated deaths without reporting delay. Hence, it is important to know whether the changes in the proportion of detected deaths can be predicted by daily testing and tracing data.

We quantified time-varying FR and R_t by taking into account the proportion of undetected cases estimated using death data. We then developed a model based on logistic regression to predict the proportion of undetected cases using daily data related to testing and tracing capacity.

Methods

Data sources

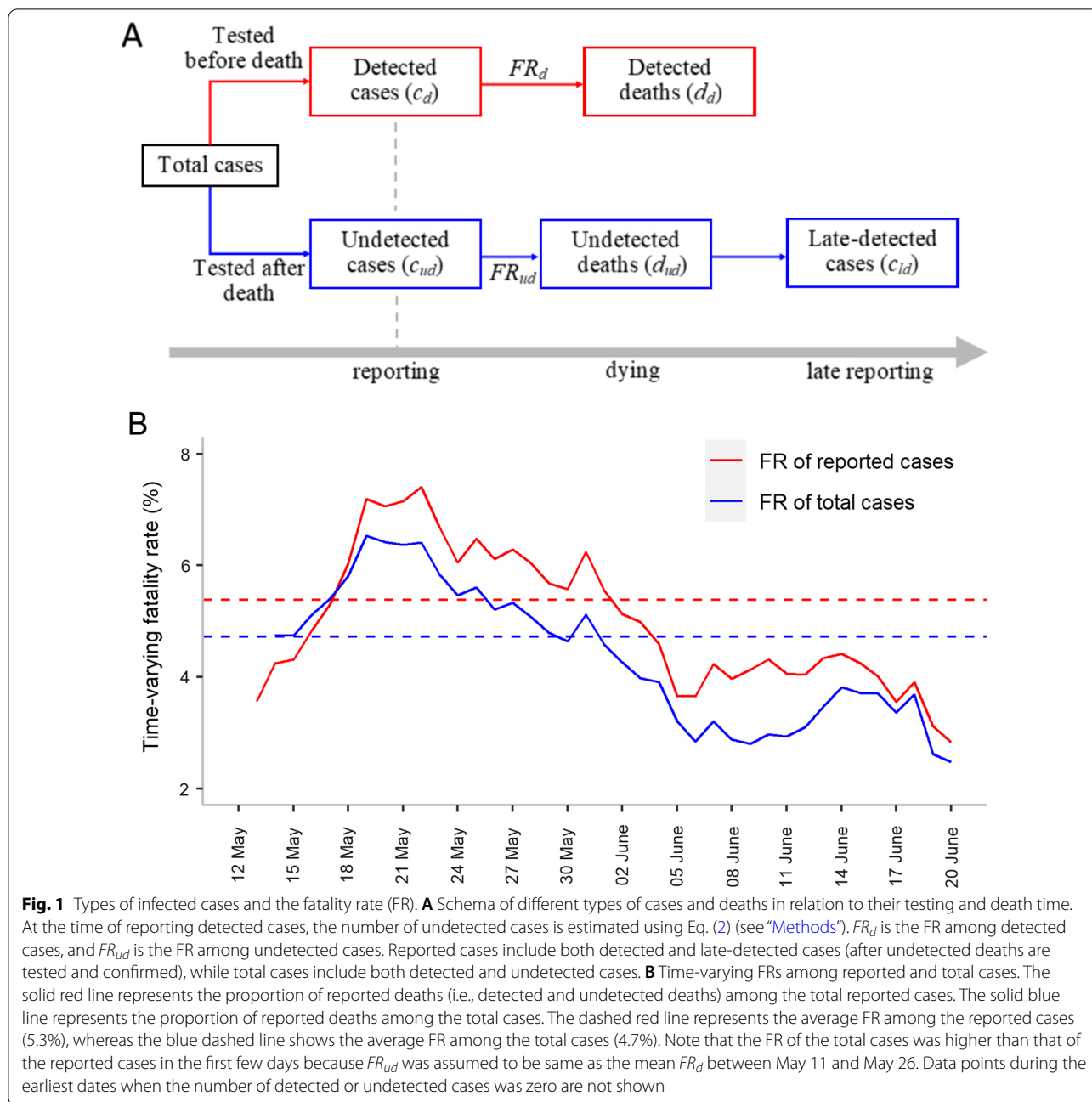
We collected the date of symptom onset time and testing date for each reported death of COVID-19 between May 28 and July 22, 2021 from Taiwan Centers for Disease Control [11]. The daily number of deaths reported before May 28 was obtained from the media [12, 13]. Daily number of confirmed cases was collected from Taiwan National Infectious Disease Statistics System [4]. We collected the daily number of tests conducted from the Government Information Open Platform, Taiwan [14, 15].

Estimating true total cases and fatality rate

Deaths from COVID-19 were classified into two categories, detected and undetected deaths, depending on whether testing was performed before the death or not, respectively (see the schema in Fig. 1A). To estimate the number of true total cases, we first considered the following ratio of undetected to detected deaths using the numbers of detected and undetected cases and their respective FR:

$$\frac{d_{ud}(t)}{d_d(t)} = \frac{c_{ud}(t) \times FR_{ud}}{c_d(t) \times FR_d(t)} \quad (1)$$

where d_d refers to the number of detected deaths, while d_{ud} refers to the number of undetected deaths; $c_d(t)$ and $c_{ud}(t)$ represent the number of cases that are detected and undetected at day t , respectively. Note that t refers to the reporting date for detected cases or detected deaths; For undetected cases or undetected deaths, t refers to the adjusted reporting date such that the reporting delay (i.e., the time elapsed between symptom onset and reporting) is adjusted to be the same as that of detected cases.



Thus, $d_d(t)$ represents the number of deaths among the detected cases who are reported at day t . Similarly, $d_{ud}(t)$ is the number of deaths among the undetected cases whose adjusted reporting date is at day t . $FR_d(t)$, which is likely to be affected by the change in hospital capacity or treatment, represents the daily FR among the detected cases at day t . FR_{ud} represents the FR among the undetected cases. FR_{ud} was assumed to be a constant, estimated as the average $FR_d(t)$ during the initial two weeks (from May 11 to May 24) of the outbreak when the

hospital capacity or treatment was not sufficient. Undetected deaths who are tested later are identified as “late-detected” cases (c_{ld}) (See Fig. 1A). We back-projected the number of late-detected cases from their late reporting time to their adjusted reporting date t [16], using the mean and standard deviation of the reporting delay among detected cases. Our aim was to estimate $c_{ud}(t)$. After rearrangement, the following formula was derived:

$$c_{ud}(t) = c_d(t) \times \frac{FR_d(t)}{FR_{ud}} \cdot \frac{d_d(t)}{d_{ud}(t)} \tag{2}$$

The value can be solved because all of the terms on the right are either known or can be estimated. We assumed that most of the undetected deaths were identified as “late-detected” cases (c_{ld}). Therefore, the number of undetected deaths was approximated by the number of late-detected cases ($d_{ud} \approx c_{ld}$) and then the ratio $\frac{d_d(t)}{d_{ud}(t)}$ was obtained. At the same time, the proportion of detected deaths (i.e., the detection ratio among death cases; $\frac{d_d(t)}{d_d(t)+d_{ud}(t)}$) was also calculated. Finally, the true number of total cases was derived empirically as the sum of detected and undetected cases (i.e., $c_d + c_{ud}$). Note that these ratios among deaths were also predicted by a regression model using data related to testing and tracing and hence a model-predicted number of total cases was obtained (see later sections).

The FRs of reported cases (including both detected and late-detected cases; $c_d + c_{ld}$) and total cases were estimated at the reporting time (or the adjusted reporting time for undetected cases) using the following equations.

$$FR_{reported}(t) = \left(\frac{d_d(t) + c_{ld}(t)}{c_d(t) + c_{ld}(t)} \right) \tag{3}$$

$$FR_{total}(t) = \left(\frac{d_d(t) + d_{ud}(t)}{c_d(t) + c_{ud}(t)} \right) \tag{4}$$

$FR_{reported}$ is commonly known as the case fatality rate, and FR_{total} is the infection fatality rate.

Estimating the proportion of detected deaths using a predictive model

We predicted the detection ratio among death cases using daily values of five indicators related to testing, tracing, and hospital capacities as candidate predictors. These indicators are: the *proportion of cases without contact tracing delay*, *ratio of the number of tests conducted*

to reported cases, *testing delay*, *reporting delay* and *death delay* (for definitions, see Fig. 2). We calculated the delay periods in testing, reporting and death by subtracting the date of symptom onset from the dates of these three events. Starting testing (the first test) earlier than or on the same day as symptom onset implied that cases were contact traced without delay. If cases were tested after symptom onset, they were either contact traced with delay or were not contact traced. The proportion of death cases that were contact traced without delay was calculated.

To investigate the factors that influence the proportion of detected deaths, we developed a logistic regression model. We assumed that the number of deaths that were previously detected on day t follows a binomial distribution, i.e. $d_d(t) \sim \text{binomial}(d(t), m(t))$, where $m(t) = \frac{d_d(t)}{d_d(t)+d_{ud}(t)}$ is the expected proportion of detected deaths on day t .

The full predictive model is:

$$\log \left(\frac{m(t)}{1 - m(t)} \right) = \alpha + \beta_1 R_{tc} + \beta_2 P_{ntd} + \beta_3 C_d + \beta_4 T_d + \beta_5 D_d \tag{5}$$

where R_{tc} is the daily ratio of tests conducted to reported cases; P_{ntd} represents the daily proportion of cases (among detected deaths) without contact tracing delay. C_d , T_d and D_d are daily reporting, testing and death delays, respectively. α is the intercept and β_i is the regression coefficient of each covariate. The proportion of undetected COVID-19 cases can be calculated using Eqs. (1) and (5) after $m(t)$ is estimated:

$$\frac{c_{ud}(t)}{c_{ud}(t) + c_d(t)} = 1 / \left(1 + \frac{m(t)}{1 - m(t)} \times \frac{FR_{ud}}{FR_d(t)} \right) \tag{6}$$

where $\frac{m(t)}{1 - m(t)} = \frac{d_d(t)}{d_{ud}(t)}$ is the odds of being detected.

(See figure on next page.)

Fig. 2 A Statuses of infection and testing of individual deaths. The gray bar represents the infection statuses of an infected case who later died after the start of infection. Orange and blue bars represent the flow of testing from the first test T_1 until the infected case was reported R . The infected case was categorized as *Detected* if the first test was performed before death. A case that was tested on the same date of or after death was categorized as *Undetected*. Among detected cases, we assumed that a case was contact traced without delay if the first test T_1 was performed before symptom onset O ; otherwise, contact traced with delay or not contact traced if the T_1 was performed after symptom onset. Testing delay refers to the time between symptom onset and the final (last) test T_f . Similarly, the reporting delay and death delay are defined as the time differences between symptom onset and reporting R , and death D , respectively. The reporting time among an undetected death was adjusted to an earlier time to have the same reporting delay as detected deaths. **B** Estimation of total number of COVID-19 cases (sum of detected and undetected) using a regression model. With the best-fitting model (see Table 2), we estimated the percentage of deaths that are detected, $m(t)$. Undetected proportion of cases was estimated based on the relationship between $m(t)$ and fatality rates (see Eq. 6). Gray dashed lines represent the predictors that were not included in the best-fitting model while estimating $m(t)$

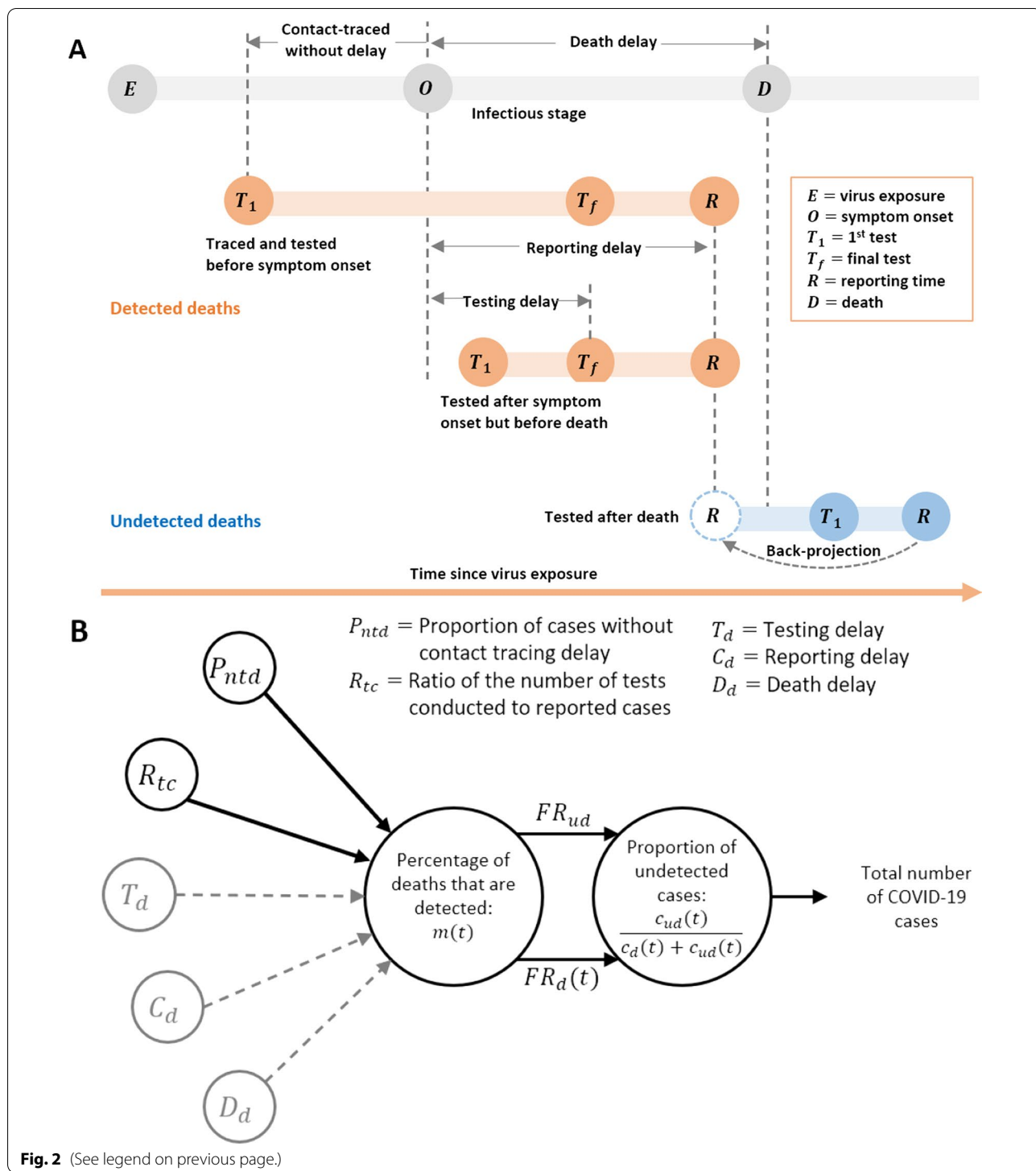


Fig. 2 (See legend on previous page.)

Model selection

To obtain the best model, the variables in Eq. 5 were added to the model iteratively. First, model fit was measured for each of the variables separately using the Akaike

information criterion (AIC) [17]. The model containing the lowest AIC value was selected as the best model candidate in this batch. Next, we added one additional variable to the candidate model from the remaining four variables in the next batch. Among the two-variable

models, the model with the lowest AIC value was selected as the best model candidate again. We obtained the best model candidates among three-variable, four-variable and full models. The final best model was obtained by comparing the best model candidates in different batches with the lowest AIC.

Model validation

To evaluate whether the predictive model achieved its intended purpose (i.e., to improve the accuracy of epidemiological parameter estimation), we explored the relationship between R_t estimated from the total cases predicted by the best model and daily mobility data. Cases were back-projected to infection time. The result was compared with R_t estimated using total cases that were empirically derived or using reported cases. R_t estimated from four scenarios of infections were compared:

Scenario 1 (S1) Total cases (at infection time) estimated using an empirical detection ratio—these cases include both the reported and undetected cases at their infection time. The number of undetected cases was estimated empirically assuming reporting delay was the same for detected and undetected cases.

Scenario 2 (S2) Total cases (at infection time) estimated from a model-predicted detection ratio—these cases include both the reported and undetected cases at their infection time. The number of undetected cases was estimated from the model assuming reporting delay was the same for detected and undetected cases.

Scenario 3 (S3) Reported cases (at infection time)—cases that were detected before death and late-detected after death at their infection time.

Scenario 4 (S4) Reported cases (at reporting time)—cases that were detected before death and late-detected after death at their reporting time. Late-detected cases were back-projected to the adjusted reporting time.

Estimating the effective reproduction number

The effective reproduction number R_t was estimated from the daily new cases of infection using the statistical package *EpiEstim* [18]. To estimate the daily number of new cases, we assumed that both the incubation time and reporting delay followed gamma distribution [19, 20]. The mean incubation time for the circulated strain in Taiwan was 3.53 days [21], and we estimated the mean reporting delay as 4.45 days. Assuming the standard deviations were equal for both the distributions (estimated as 3.93 days for the reporting delay), the distribution of time between infection and reporting was gamma distribution with a mean of 7.98 days and a standard deviation of 5.28 days. The mean of the distribution was estimated as the sum of mean incubation time and confirmation delay. In contrast, the standard deviation was obtained from

weighted means and pooled standard deviation for the period between infection and reporting using the following formula:

$$sd_{\text{gamma}} = sd_{\text{pooled}} \sqrt{\left(\frac{m_1 + m_2}{m_w} \right)} \quad (7)$$

where, m_1 and m_2 are mean incubation time and confirmation delay and m_w refers weighted mean of these two. sd_{pooled} represents the pooled standard deviation for the period between infection and reporting.

We then estimated total cases at infection time using the empirical detection ratio (S1) and the model-predicted detection ratio (S2), and reported cases at infection time (S3) using a back-projection method [16].

We set initial conditions for estimating R_t . Before May 11, we assumed that there were 15 cases each day between May 6 and 10, which was the average number of reported cases at infection time during this 5-day period.

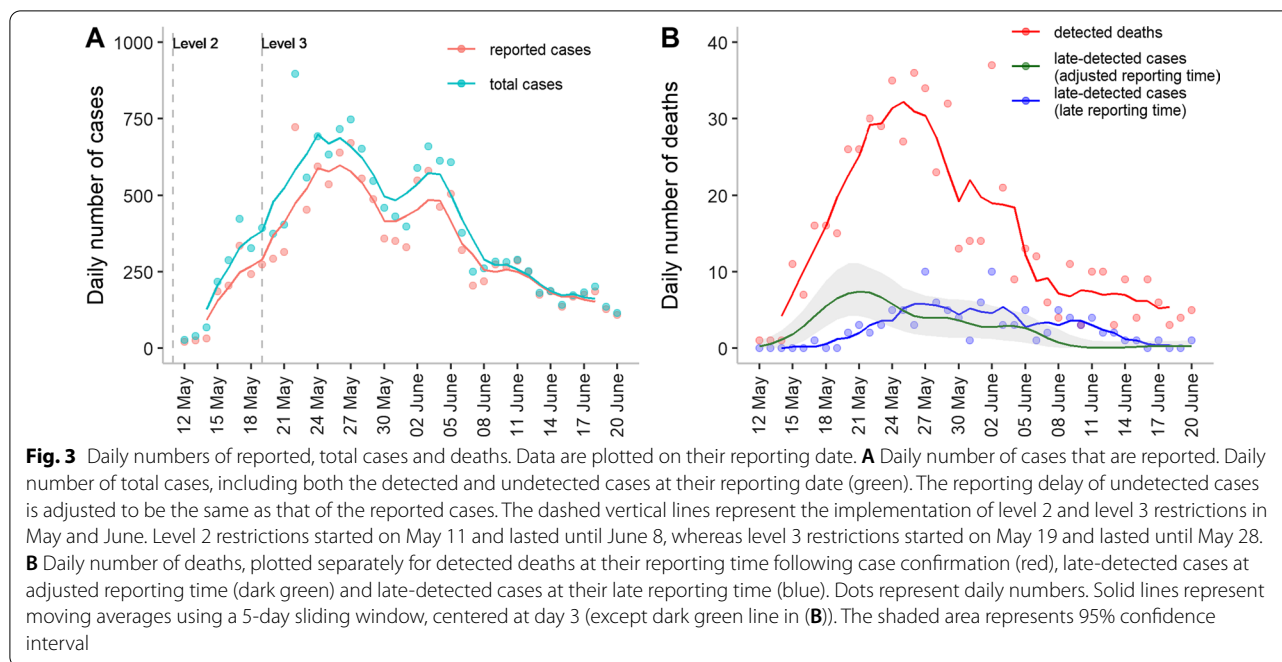
Results

Time-varying FR among true total cases (Eq. 4) was first quantified after taking into account undetected cases and was compared with that of reported cases. The number of total cases was also predicted using polymerase chain reaction (PCR) testing data (Eqs. 5 and 6). To assess the impact of including undetected cases, we investigated the relationship between R_t generated using total cases and mobility data and then determined whether the relationship improved, compared with R_t from reported cases.

After the number of undetected cases was considered, the estimated FR was lower than using reported cases but was still high during the initial period of the outbreak. The mean FR of total cases was estimated to be 4.7%, which was lower than the mean FR of 5.3% for reported cases (Fig. 1B). The FR increased rapidly from 4.7% and peaked at 6.5% on May 19, but then continued decreasing, reaching 2.8% on June 6. Since then, the rate was generally maintained.

From May 24 to June 3, the 5-day moving average numbers of reported cases reached a plateau and then declined thereafter (Fig. 3A). The estimated true daily case number at the peak of the outbreak on May 22 was 897, which was 24.3% higher than the reported number, indicating 19.5% of infections were not detected. The difference became less than 4% on June 9 and afterwards.

Until June 20, a total of 105 late-detected cases were reported, indicating many undetected deaths. Similarly, daily detected deaths also reached a plateau around May 24 (Fig. 3B). However, the number of late-detected cases (at adjusted reporting time), reached a peak (7 persons per day) on May 21 and started to decline immediately, approaching zero after June 8. This indicated the



improvement of the detected ratio among deaths. The detection ratio among deaths, which was about 50% initially, exceeded 95% after the end of May (Additional file 1: Figure S1B). This ratio was very different from the observed ratio (a V-shaped pattern) without back-projection (Additional file 1: Figure S1A).

Predicting detection ratio using testing data

We next investigated whether the improvement in the proportion of detected cases was related to the improved capacity of testing and tracing. The indicators of the capacity were explained by the schematic of individual infection and testing statuses of each case among deaths (for definitions, please refer to Fig. 2 and its legend). Depending on the time of testing, the case can be categorized as a detected death (contact traced without delay or tested after symptom onset but before death) or an undetected death (tested after death). More efficient contact tracing allowed more cases to be traced and tested before symptom onset and was indicated by the *proportion of cases without contact tracing delay*. This proportion fluctuated between 25 and 75% throughout the study period, with an increasing trend from late May (below 50%) to late June (above 60%) (Fig. 4A). The *testing delay* gradually increased, from approximately two days to up to 4–6 days, until June 14, a few weeks after the outbreak started to decline (Fig. 4B). The *reporting delay* from the day of symptom onset ranged mostly between 2.5 and 7.5 days (Fig. 4E), whereas the *death delay* continued increasing from 5 days to more than 18 days (Fig. 4C

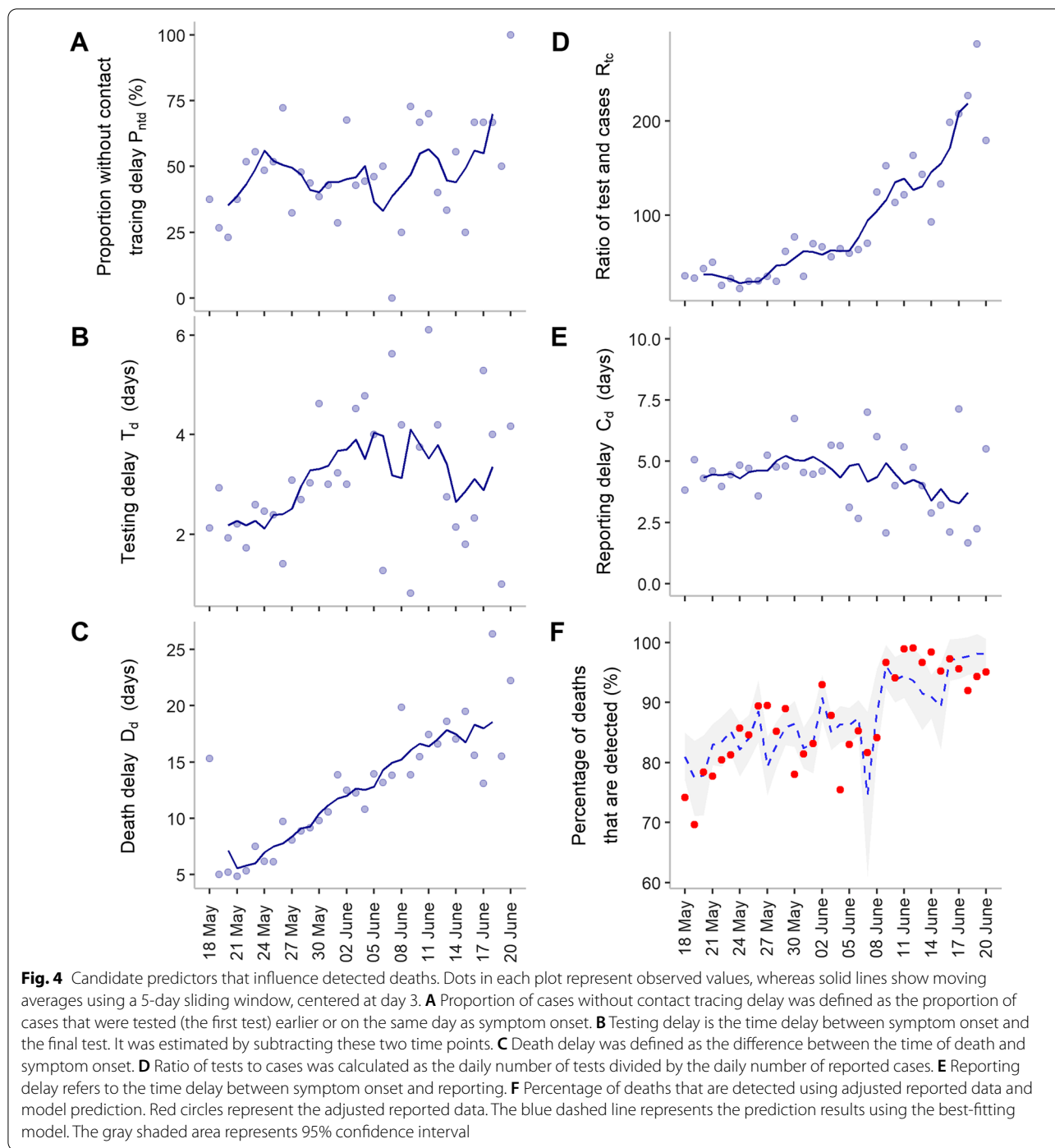
& Additional file 1: Figure S3B). The *ratio of the number of tests conducted to reported cases* increased from less than 50 to more than 200 (Fig. 4D), demonstrating the improvement in testing capacity throughout the outbreak.

We compared models starting from the most basic to more complex ones by their AIC values to identify the best-fitting model. The model with the predictor, the proportion of cases without contact tracing delay and the ratio of tests conducted to reported cases, was selected as the best-fitting model because of its lowest AIC value of 91.0 (Model 2 in Table 1).

The model successfully captured the trend in the proportion of detected deaths (Fig. 4F). 20 out of 34 daily values were successfully predicted within the confidence interval (CI). Among the values outside the interval, most of them were in the near distance; only two dots have errors larger than two times the intervals.

We further validated the best-fitting model by using past data as training sets (from the first 50, 60 to 70% of the full data sets) to predict future results. The model captured the observed trend of the number of detected deaths in each validation set (Additional file 1: Figure S5). Moreover, the model predicted most data points within CI. For 95% CI 89–93% of data were predicted correctly.

The results suggest that a higher detection ratio among deaths was determined by a larger proportion of cases who were contact traced without delay and a higher number of tests conducted relative to the number of cases (Table 2).



Comparing effective reproduction number and mobility index

Comparisons were made between R_t estimated using (i) total cases that were estimated using the empirical detection ratio; (ii) total cases that were estimated from the model-predicted detection ratio using testing data; and (iii) reported cases only (see Fig. 5A, B, Additional file 1:

Figure S2 and Methods). When the total case number was used, R_t was higher during the earlier dates. The number reached a maximum value of 6.4 on May 11, compared to 6.0 estimated using the reported case number. We further evaluated the relationship between R_t and mobility data during the period when R_t reduced from the maximum value to 1 (May 11 to May 24) (Additional file 1:

Table 1 Candidate models used to choose the best model

Models	Description	AIC
1	$\alpha + \beta_1 R_{tc}$	95.9
2	$\alpha + \beta_1 R_{tc} + \beta_2 P_{ntd}$	91.0
3	$\alpha + \beta_1 R_{tc} + \beta_2 P_{ntd} + \beta_3 C_d$	93.0
4	$\alpha + \beta_1 R_{tc} + \beta_2 P_{ntd} + \beta_3 C_d + \beta_4 T_d$	95.0
5	$\alpha + \beta_1 R_{tc} + \beta_2 P_{ntd} + \beta_3 C_d + \beta_4 T_d + \beta_5 D_d$	96.9

α and β s are model coefficients, whereas the proportion of contact tracing delay (P_{ntd}), the ratio of the number of tests conducted to reported cases (R_{tc}), the delay in testing (T_d), the delay in reporting (C_d), and the delay in deaths (D_d) are predictors. AIC represents the Akaike information criterion

Table 2 Parameter estimates of the best-fitting model (Model 2)

Covariates	Estimates	95% Confidence intervals	p-value
Ratio of number of tests conducted to reported cases (R_{tc})	0.009	0.002–0.018	0.0180
Proportion of cases without tracing delay (P_{ntd})	1.834	0.316–3.375	0.0185

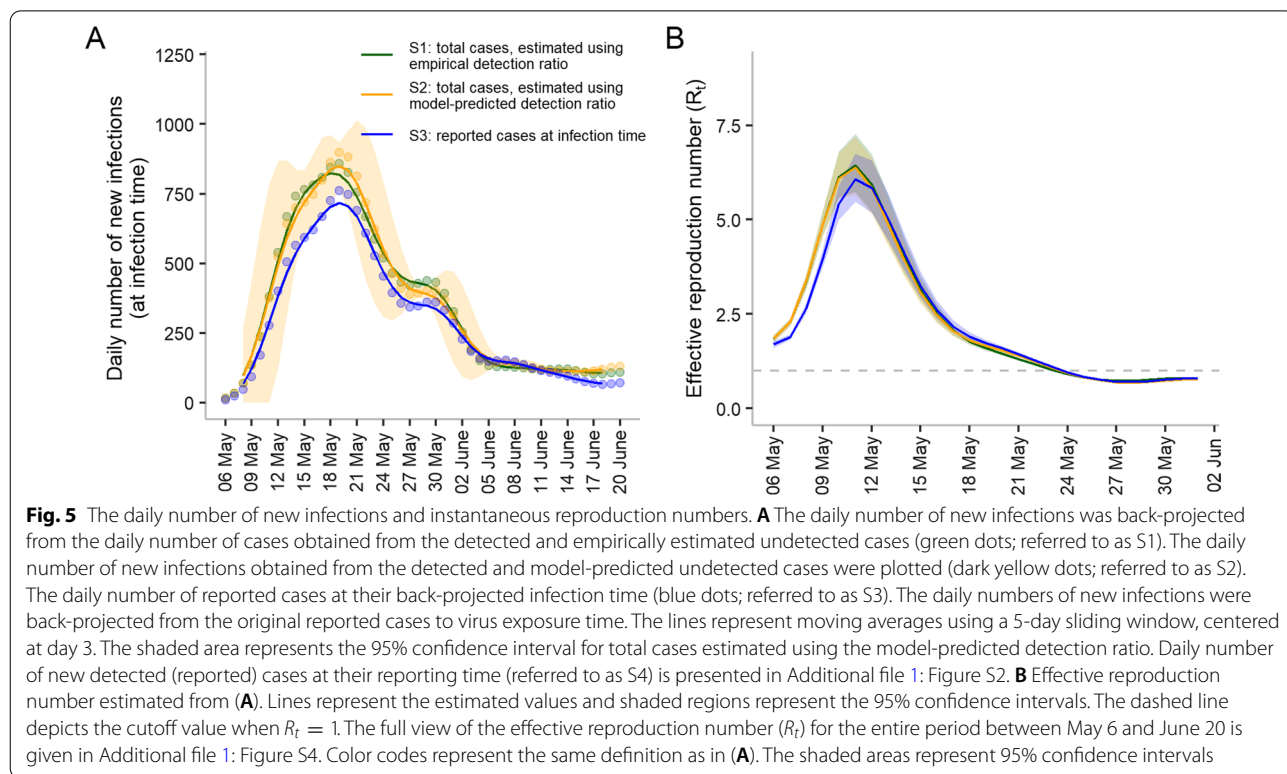
Table S1). We found that when the total case number was used (either estimated using the empirical detection ratio

or predicted using the testing data), a lower AIC was produced, indicating a better fit to the mobility data.

In summary, efficiencies of testing and contact tracing changed during the outbreak and were useful in predicting the proportion of undetected cases. After adding the undetected cases, a better estimate of R_t was made and a reduction in the FR was observed.

Discussion

Understanding whether a high FR observed in the recent largest COVID-19 outbreak in Taiwan was attributed to a higher number of undetected cases or insufficient health care capacity is important to guide interventions to reduce COVID-19 mortality in the future. An important observation is that even though the proportion of undetected cases was included, the average FR was only adjusted to 4.7% from 5.3%, which is still higher than the global average for the same time (i.e., 2.1% in May and June 2021 [5]). However, the daily FR reduced to 2.8% on June 6 and remained at this low level, similar to that in the United States (i.e., 2.8% in May and June 2021 [22]). The reduction from the initially high FR can be explained by the improvement in treatment to accommodate the sudden rise in cases. This is supported by the observation that the duration between symptom onset and death among detected deaths continued increasing from



approximately five days to more than two weeks in June (Fig. 4C).

Furthermore, the results highlight the importance of improving testing capacity and contact tracing efficiency. For example, if more cases can be traced or confirmed earlier (Fig. 4), many of them can be treated promptly, hence reducing fatality rates.

The number of hidden (undetected) COVID-19 cases often affects the estimation of transmissibility of the virus and the effectiveness of non-pharmaceutical interventions (NPIs) implemented. Even though the effects of contact tracing and testing on transmissibility have been studied [23, 24], how many hidden cases do they cause is unclear. We demonstrated that the time-varying detection ratios can be predicted using data on testing and contact tracing. As a result, a more accurate R_t can be obtained, which is likely to be explained by mobility data better (Additional file 1: Table S1). The guidance for implementing NPIs based on changes in mobility can be provided [10]. If detection ratio is low, R_t is likely to be underestimated. In our study, only about 20% of cases were undetected. Therefore, the difference of the estimation is only little. However, the impact of underestimation can be serious in countries with insufficient testing capacity or low tracing efficiency.

We found that the ratio of the number of tests conducted to reported cases, and the proportion of cases that are contact traced without delay can be used to “nowcast” the proportion of undetected cases. Because the number of tested samples can quickly reach the capacity limit when the case number is growing, many samples remain untested. Hence, each day, the number of confirmed cases depends largely on how many tests can be performed. A day delay in testing and confirming a case, leads to a day delay in tracing the close contacts of the case. Furthermore, a higher contact tracing coverage together with a shorter delay of being traced enables more cases to be identified earlier [23, 24]. These suggest increasing testing and tracing capacity to identify those infections earlier can reduce hidden cases more.

Modelling has been used to estimate the proportion of undetected COVID-19 cases using the observed case number during a specific period (e.g., before or after an intervention) of an outbreak [25, 26]. More recently, an approach through estimating under-ascertainment by directly comparing model-predicted death with excess deaths recorded was used [27]. We checked the number of deaths related to flu and pneumonia illness [11] and found no unusual excess deaths other than the reported COVID-19 deaths during this period. The proportion of undetected cases can also be calculated after incorporating seroprevalence data with false negative rates of tests

into models [28]. Overall, none of these methods estimate the constantly changing proportion of undetected cases.

Several criteria enabled us to make successful prediction using testing data. First, the number of deaths should be high. If this number is low, the uncertainty of estimating the number of undetected cases becomes high. Second, most of the deaths have to be tested eventually. Taiwan government has a strong directive to test all sudden death cases; for example, on June 18, it was announced that PCR tests would be performed for all sudden and unexplained deaths [29]. This may not likely be the case in countries with a large number of excess deaths associated with COVID-19.

In summary, predicting the number of undetected cases as early as possible using testing data can help obtain an R_t with a better relationship with mobility data, thus enabling policymakers to make timely public health decisions using mobility information to contain the outbreak.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12879-022-07190-z>.

Additional file 1. Figure S5.

Acknowledgements

We thank Meifen Feng from City University of Hong Kong for data collection. The authors thank National Taiwan University and City University of Hong Kong for providing excellent research facilities and thank colleagues at these institutes for their supports. The authors also acknowledge support from grants funded by Health and Medical Research Fund [COVID190215] and City University of Hong Kong [7200573 and 9610416].

Authors' contributions

HY and MH wrote the main manuscript text. HY, MW and TW designed the study and MH conducted the analyses. All authors read and approved the final manuscript.

Funding

Health and Medical Research Fund, COVID190215. City University of Hong Kong, 7200573. City University of Hong Kong, 9610416.

Availability of data and materials

The data that support the findings of this study are available in databases (references [4, 9–11]).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

No, I declare that the authors have no competing interests as defined by BMC, or other interests that might be perceived to influence the results and/or discussion reported in this paper.

Author details

¹Department of Biomedical Sciences, Jockey Club College of Veterinary Medicine and Life Sciences, City University of Hong Kong, Hong Kong, SAR, China. ²Centre for Applied One Health Research and Policy Advice, Jockey Club College of Veterinary Medicine and Life Sciences, City University of Hong Kong, Hong Kong, SAR, China. ³Department of Geography, National Taiwan University, Taipei City, Taiwan. ⁴Department of Anesthesiology, National Taiwan University Hospital, National Taiwan University, Taipei City, Taiwan.

Received: 1 November 2021 Accepted: 4 February 2022

Published online: 20 March 2022

References

- Grassly NC, Pons-Salort M, Parker EPK, White PJ, Ferguson NM. Comparison of molecular testing strategies for COVID-19 control: a mathematical modelling study. *Lancet Infect Dis*. 2020;20:1381.
- Contreras S, et al. The challenges of containing SARS-CoV-2 via test-trace-and-isolate. *Nat Commun*. 2021;12:1–13.
- Tan Y. COVID-19: what went wrong in Singapore and Taiwan? - BBC News. BBC News; 2021. <https://www.bbc.com/news/world-asia-57153195>.
- Taiwan National Infectious Disease Statistics System. Severe Pneumonia with Novel Pathogens (COVID-19); 2021. <https://nidss.cdc.gov.tw/en/nidss/disease?id=19CoV>.
- Our World in Data. Coronavirus (COVID-19) deaths—statistics and research; 2021. <https://ourworldindata.org/covid-deaths>.
- Taiwan Centers for Disease Control. CECC raises epidemic warning to Level 2 and implements related restrictions and measures, effective from May 11 to June 8, in response to increased risk of community transmission; 2021. <https://www.cdc.gov.tw/En/Bulletin/Detail/0jMlMmCVWTuhO9mfQCd-4g?typeid=158>.
- Taiwan Centers for Disease Control. CECC raises epidemic warning to Level 3 nationwide from May 19 to May 28; strengthened measures and restrictions introduced across Taiwan to reduce community transmission; 2021. https://www.cdc.gov.tw/En/Bulletin/Detail/VN_6yeoBTKhRkoSy2d0hJQ?typeid=158.
- Hâncean MG, Perc M, Lerner J. Early spread of COVID-19 in Romania: imported cases from Italy and human-to-human transmission networks. *R Soc Open Sci*. 2020;7:200780.
- Hâncean MG, Slavinec M, Perc M. The impact of human mobility networks on the global spread of COVID-19. *J Complex Networks*. 2021;8:1–14.
- Nouvellet P, et al. Reduction in mobility and COVID-19 transmission. *Nat Commun*. 2021;12:1–9.
- Taiwan Centers for Disease Control. COVID-19 (SARS-CoV-2 Infection). <https://www.cdc.gov.tw/En>.
- NOW健康. up to date! (A total of 850 cases) Taiwan's new coronary pneumonia death case book; 2021. https://healthmedia.com.tw/main_detail.php?id=45372.
- NOW健康. (Death Accumulation) Taiwan's New Coronary Pneumonia Death Cases and Events Book (47th to 88th death cases); 2021. https://healthmedia.com.tw/main_detail.php?id=49205.
- Government Information Open Platform. Taiwan's COVID-19 coronavirus test daily delivery number; 2021. <https://data.gov.tw/dataset/120451>.
- Taiwan daily confirmed local case data; 2021. https://docs.google.com/spreadsheets/d/12tQKRUaiBZfc9yDd6tmiOds62ke_4AcKmNJ6q4gdU/.
- Becker NG, Watson LF, Carlin JB. A method of non-parametric back-projection and its application to aids data. *Stat Med*. 1991;10:1527–42.
- Bozdogan H. Model selection and Akaike's Information Criterion (AIC): the general theory and its analytical extensions. *Psychometrika*. 1987;52:345–70.
- Cori A, Ferguson NM, Fraser C, Cauchemez S. A new framework and software to estimate time-varying reproduction numbers during epidemics. *Am J Epidemiol*. 2013;178:1505–12.
- Adam DC, et al. Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong. *Nat Med*. 2020;26:1714–9.
- Qin J, et al. Estimation of incubation period distribution of COVID-19 using disease onset forward time: a novel cross-sectional and forward follow-up study. *Sci Adv*. 2020;6:eabc1202.
- Homma Y, et al. The incubation period of the SARS-CoV-2 B.1.1.7 variant is shorter than that of other strains. *J Infect*. 2021;83:e15–7.
- Worldometer. COVID-19 coronavirus pandemic; 2021. <https://www.worldometers.info/coronavirus/>.
- Kretzschmar ME, et al. Impact of delays on effectiveness of contact tracing strategies for COVID-19: a modelling study. *Lancet Public Heal*. 2020;5:e452–9.
- Kucharski AJ, et al. Effectiveness of isolation, testing, contact tracing, and physical distancing on reducing transmission of SARS-CoV-2 in different settings: a mathematical modelling study. *Lancet Infect Dis*. 2020;20:1151–60.
- Liang J, Yuan H-Y, Wu L, Pfeiffer DU. Estimating effects of intervention measures on COVID-19 outbreak in Wuhan taking account of improving diagnostic capabilities using a modelling approach. *BMC Infect Dis*. 2021;21:1–10.
- Li R, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science*. 2020;368:489–93.
- Watson OJ, et al. Leveraging community mortality indicators to infer COVID-19 mortality and transmission dynamics in Damascus, Syria. *Nat Commun*. 2021;12:1–10.
- Bhattacharyya R, et al. Incorporating false negative tests in epidemiological models for SARS-CoV-2 transmission and reconciling with seroprevalence estimates. *Sci Rep*. 2021;11:1–14.
- Taiwan Centers for Disease Control. Report on the Press Conference after the National Epidemic Prevention Conference on June 18. 2021. <https://www.cdc.gov.tw/Bulletin/Detail/j5bFJQNgMHxaaQIK-j12w?typeid=9>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

