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.

Effect of changing case definitions for COVID-19 on the epidemic curve and transmission parameters in mainland China: a modelling study

Tim K Tsang, Peng Wu, Yun Lin, Eric H Y Lau, Gabriel M Leung, Benjamin J Cowling

Summary

Background When a new infectious disease emerges, appropriate case definitions are important for clinical diagnosis and for public health surveillance. Tracking case numbers over time is important to establish the speed of spread and the effectiveness of interventions. We aimed to assess whether changes in case definitions affected inferences on the transmission dynamics of coronavirus disease 2019 (COVID-19) in China.

Methods We examined changes in the case definition for COVID-19 in mainland China during the first epidemic wave. We used exponential growth models to estimate how changes in the case definitions affected the number of cases reported each day. We then inferred how the epidemic curve would have appeared if the same case definition had been used throughout the epidemic.

Findings From Jan 15 to March 3, 2020, seven versions of the case definition for COVID-19 were issued by the National Health Commission in China. We estimated that when the case definitions were changed, the proportion of infections being detected as cases increased by 7·1 times (95% credible interval [CrI] 4·8–10·9) from version 1 to 2, 2·8 times (1·9–4·2) from version 2 to 4, and 4·2 times (2·6–7·3) from version 4 to 5. If the fifth version of the case definition had been applied throughout the outbreak with sufficient testing capacity, we estimated that by Feb 20, 2020, there would have been 232 000 (95% CrI 161 000–359 000) confirmed cases in China as opposed to the 55 508 confirmed cases reported.

Interpretation The case definition was initially narrow and was gradually broadened to allow detection of more cases as knowledge increased, particularly milder cases and those without epidemiological links to Wuhan, China, or other known cases. These changes should be taken into account when making inferences on epidemic growth rates and doubling times, and therefore on the reproductive number, to avoid bias.

Funding Health and Medical Research Fund, Hong Kong.

Lancet Public Health 2020; 5: e289–96

Published Online April 21, 2020 https://doi.org/10.1016/S2468-2667(20)30089-X

WHO Collaborating Centre for Infectious Disease Epidemiology and Control, School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong Special Administrative Region, China (T K Tsang PhD, P Wu PhD, Y Lin BM, E H Y Lau PhD, Prof G M Leung MD, Prof B J Cowling PhD)

Correspondence to: Dr Peng Wu, School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong Special Administrative Region, China **pengwu@hku.hk**

Introduction

When a newly emerging infectious disease is first identified, specifying appropriate case definitions can help to identify individuals who are infected in an efficient manner.[1] Often a hierarchy of case definitions will be used, so that a suspected case can be defined based on broad epidemiological and clinical criteria—eg, patients with particular exposures or in particular geographical locations, with particular signs or symptoms, at a particular time. A confirmed case can be defined as a suspected case in which the pathogen of interest is identified or isolated with a specific laboratory test. Epidemiological and clinical information for patients who meet a case definition can inform the source or sources of infections, potential modes of transmission, transmission dynamics, and severity of the infection. All this information is important for establishing optimal control measures.

Coronavirus disease 2019 (COVID-19) is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The novel virus was first identified in a cluster of patients with atypical pneumonia in Wuhan, China, in December, 2019.[2,3] At the end of January, 2020, it became clear that infection was spreading efficiently from person to person, and also that there was a broader clinical spectrum of infections.[4] As a consequence of the evolving information on the epidemiological and clinical spectrum of infections, there have been several revisions to the case definition for COVID-19 in mainland China.

Here, we review the various COVID-19 case definitions that have been used in mainland China as of March 13, 2020, and examine the implications of changes in case definitions on the epidemiology of COVID-19, aiming to quantify the effect of changes in the case definition on inferences about transmission parameters based on the epidemic curve.

Methods

Sources of data

We obtained the officially published guidelines on diagnosis and treatment of COVID-19 from the National Health Commission and other public sources. The first two editions were not originally released publicly, while

Research in context

Evidence before this study

Coronavirus disease 2019 (COVID-19) case numbers increased throughout January, 2020, in China. As more information became available on disease spectrum, and laboratory testing capacity was expanded, the case definitions were also changed. We searched PubMed for studies published in English from database inception up until April 4, 2020, reporting the effect of changing case definitions on the epidemic curve for COVID-19 using keywords including "COVID-19", "2019-nCoV", "novel coronavirus-infected pneumonia", "SARS-CoV-2", and "case definition".

We examined 19 studies in detail and found three studies that were relevant to the change of case definition. One study estimated the incubation period in the early stage of the outbreak, which can be helpful for modifying the case definitions. Another modelling study allowed for the change of case definitions in Wuhan via additional parameters for the change in case detection probabilities. A study also noted that their analyses were based on earlier case definitions, and stated that the estimated effective reproductive number was an upper bound because later case definitions could capture more cases. We found no study directly estimating the effect of case definitions on epidemic curves except two studies that briefly summarised the changes of case definitions in the guidelines on the diagnosis and treatment of patients with COVID-19 in China.

Added value of this study

We collected publicly available information on epidemic curves in China, and summarised the changes in seven versions of case definitions. We found that changes in the case definitions of COVID-19 had a substantial effect on the proportion of infections that were detected as cases. We estimated that if changing case definitions were unaccounted for, the growth rate would be overestimated. We also estimated the total number of cases if a broader case definition had been applied at the early stage of the epidemic and if there had been sufficient laboratory capacity. With these assumptions, we estimated that approximately as many as 232 000 infections could have been confirmed as COVID-19 cases in China by Feb 20, 2020, around four times more than the 55 508 cases identified by that date.

Implications of all the available evidence

Changes of case definitions or laboratory testing capacity should be accounted for when analysing an epidemic curve. In China, broadening the case definitions over time allowed a greater proportion of infections to be detected as cases. Taking into account these changes, we estimated that there were at least 232 000 infections in the first epidemic wave of COVID-19 in mainland China. The true number of infections could still be higher than that currently estimated considering the possibility of under-detection of some infections, particularly those that were mild and asymptomatic, even under the broadest case definitions.

the third edition onwards have been released by the National Health Commission.[5] Epidemic curves by onset date and report date from Dec 2, 2019, to Feb 20, 2020, in China were extracted from the data presented in the report of the WHO-China Joint Mission on Coronavirus Disease 2019 in February, 2020.[6]

Statistical analysis

We reviewed the case definitions and highlighted the key changes in sequential updates. We fitted an exponential growth model to the incidence of cases to quantify the effect of changing case definitions on the epidemic curve for laboratory-confirmed cases (appendix p 2). In the model, we assumed that each change in case definition increased the proportion of cases that would be detected among all infections. Also, we assumed the effect of changing case definition was the same for all regions in China. To account for the control measures, such as the lockdown in Wuhan and other cities in China on Jan 23, 2020, and the subsequent days,[7] we allowed the growth rate to change on this date. Because the interventions acted to prevent infections but the epidemic curve was based on date of symptom onset in our analysis, the effect of the interventions would be expected to have a slightly delayed effect on the epidemic curve, which we accounted for by incorporating the incubation period distribution (appendix p 2). The incubation period was assumed to follow a log-normal distribution with a mean of 5·2 days (SD 3·9).[8]

See Online for appendix

When changing the case definition, there could be a backfill of cases that fulfilled the new case definition around the change time. We allowed for backfill up to 10 days before each change in case definition by assuming that a change in case definition could have a partial effect on incidence before the change date t, accounting for the reporting delay, which was estimated from the onset time series and report time series (appendix p 2). We estimated the growth rate as one of the model parameters, and we estimated the doubling time using log(2) divided by the estimated growth rate. We fitted separate models for Wuhan, Hubei province excluding Wuhan, and the rest of mainland China excluding Hubei province, to account for the regional differences in growth rates, epidemic timing, and potential transmissibility. We estimated the basic reproductive number R_0, corresponding to the mean number of secondary infections from one case at the start of the outbreak, using the formula: 1 divided by $M(-r)$,[9] where r was the growth rate and M was the moment generating function of the generation time distribution. We assumed the generation time distribution followed the same gamma distribution as a

Figure 1: **Evolution of case definitions for COVID-19**

Seven editions of the National Guideline for Diagnosis and Treatment of COVID-19 have been published in China since Jan 15, 2020. COVID-19=coronavirus disease 2019. SARS-CoV-2=severe acute respiratory syndrome coronavirus 2. *Version 4 referred to travel history to or residence in an area with sustained local transmission of SARS-CoV-2 infection. †Individuals with symptoms were considered as those showing fever and respiratory symptoms in versions 2 and 3, and fever or respiratory symptoms in other versions of the definition. ‡In version 4, a patient was either one of a cluster of patients or epidemiologically linked to a confirmed COVID-19 case. §Clustering events were further clarified in version 7 as "2 or more cases with fever and/or respiratory symptoms found in a small area within 2 weeks", but not in previous versions. ¶Multiple rows indicate alternative options for meeting the case definition.

previously estimated serial interval distribution with a mean of 7·5 days (SD 3·4).[8] As a sensitivity analysis, we used another estimated serial interval distribution with the mean of 4·7 days (SD 2·9).[10] In addition, we did a sensitivity analysis allowing backfill for up to 15 days before each change in case definitions.

To account for the uncertainty in estimates of the onset-to-reporting interval, and to allow us to quantify the uncertainty in model parameters including the growth rates, we did our analysis in a Bayesian framework and constructed a Markov chain Monte Carlo algorithm,[11] which allowed joint parameter estimation (appendix p 3). Substantial differences in parameters (eg, growth rate and doubling time) were defined as the non-overlap of their credible intervals (CrI), meaning that the probability that the two parameters were the same was less than 0·05. On the basis of the modelling results, we estimated the number of cases if version 5 of the case definition had been applied throughout the outbreak (appendix p 4). All statistical analyses were done using R version 3.5.2. All data and code required to reproduce the analysis are available online (appendix p 4).

For **data and code** see https://github.com/timktsang/covid19_casedef

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

We analysed the changes of the case definition for COVID-19 applied in China from Jan 15 to March 3, 2020. Before Jan 15, 2020, we were unable to identify the case definition that was used in Wuhan

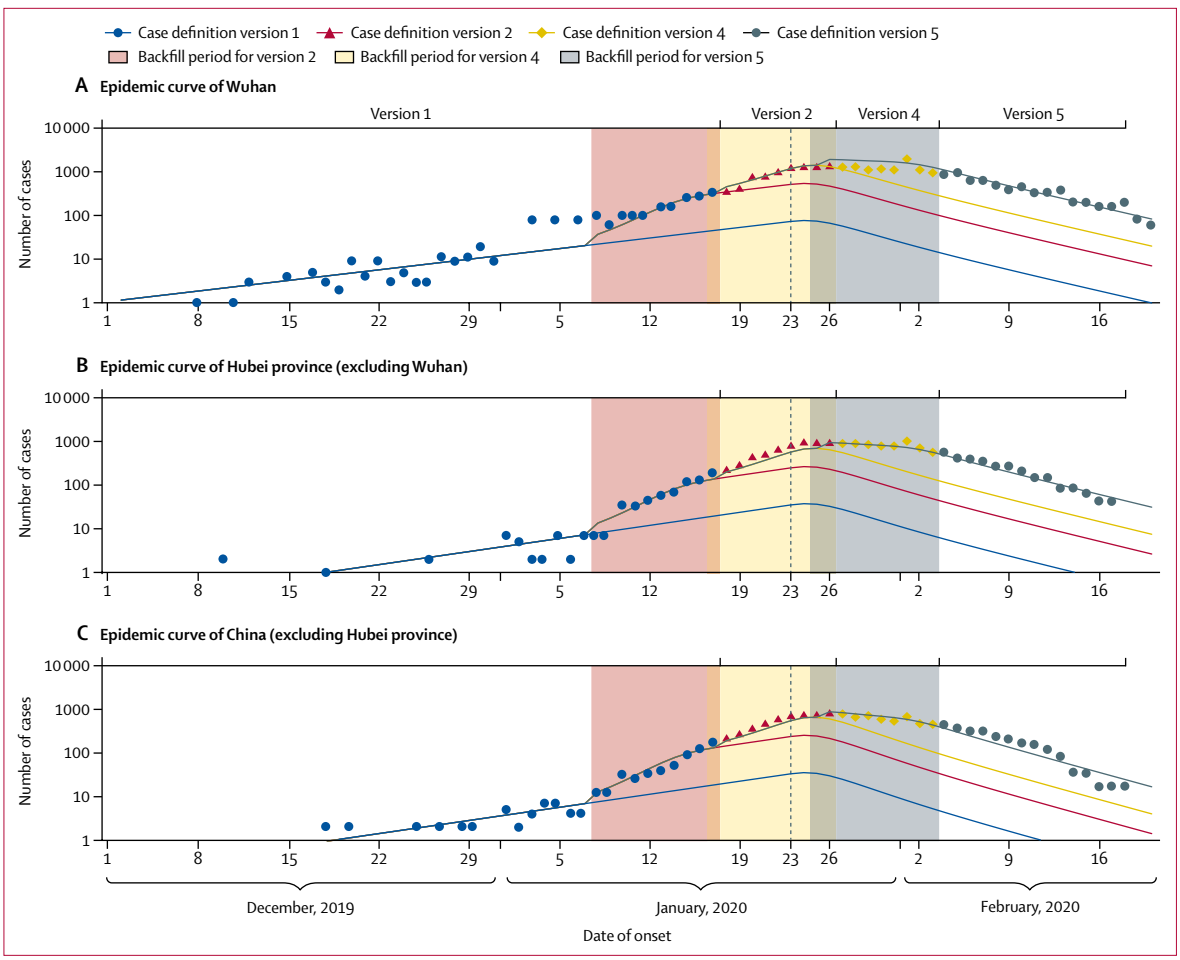

Figure 2: **Reported COVID-19 cases by date of onset and the the modelled exponential growth of daily numbers of cases by application of different versions of case definitions**

Data are assuming that the version of the case definition was applied throughout the study period in mainland China, as of Feb 20, 2020. Symbols and lines show daily numbers of reported and estimated cases, and colours indicate cases in line with the different versions of COVID-19 case definitions. The coloured shading areas reflect that changing case definitions were adjusted earlier to reflect the assumption that there was a backfill of symptomatic cases who had not yet presented for diagnosis up to 10 days before each change in case definition, and therefore the effect of changing case definition would appear to modify the proportion of infections captured as cases before the actual day of change. The vertical dashed line indicates the implementalion of control measures.
COVID-19=coronavirus disease 2019.

to identify the earliest 41 confirmed cases. The first national guideline for diagnosis and treatment was issued on Jan 15, 2020, and required six specific criteria to be met for a patient to be a confirmed case of COVID-19 (figure 1, appendix p 6). Notably, patients needed to have an epidemiological link to Wuhan or a wet market in Wuhan and had to fulfil four clinical conditions indicative of viral pneumonia to be identified as suspected cases. They then had to have a respiratory specimen tested by full genome sequencing showing a close homology with SARS-CoV-2 for the final confirmation of COVID-19. In the following days and weeks, several revisions were made to the case definitions, allowing gradually greater sensitivity in the criteria required for case confirmation (figure 1). We present the seven versions of cases definitions in the appendix (pp 6–18).

The second edition of the case definitions removed the requirement for failure of antibiotic treatment to identify suspected cases and allowed PCR confirmation in addition to whole genome sequencing. There was no change in case definitions in the third edition, but classifications of severe and critical cases were modified and clarified. The fourth edition allowed patients to have an epidemiological link to other areas with reported cases, instead of being restricted to Wuhan, and suspected cases required only two, instead of all three, types of clinical manifestations in addition to an epidemiological link. The greatest change was in the fifth edition, which introduced a new category of cases (ie, clinically confirmed cases), specifically for Hubei province, which was the epicentre of the outbreak and had the largest number of cases identified in the country. Here, clinically confirmed cases were patients that met clinical criteria and had radiological evidence of

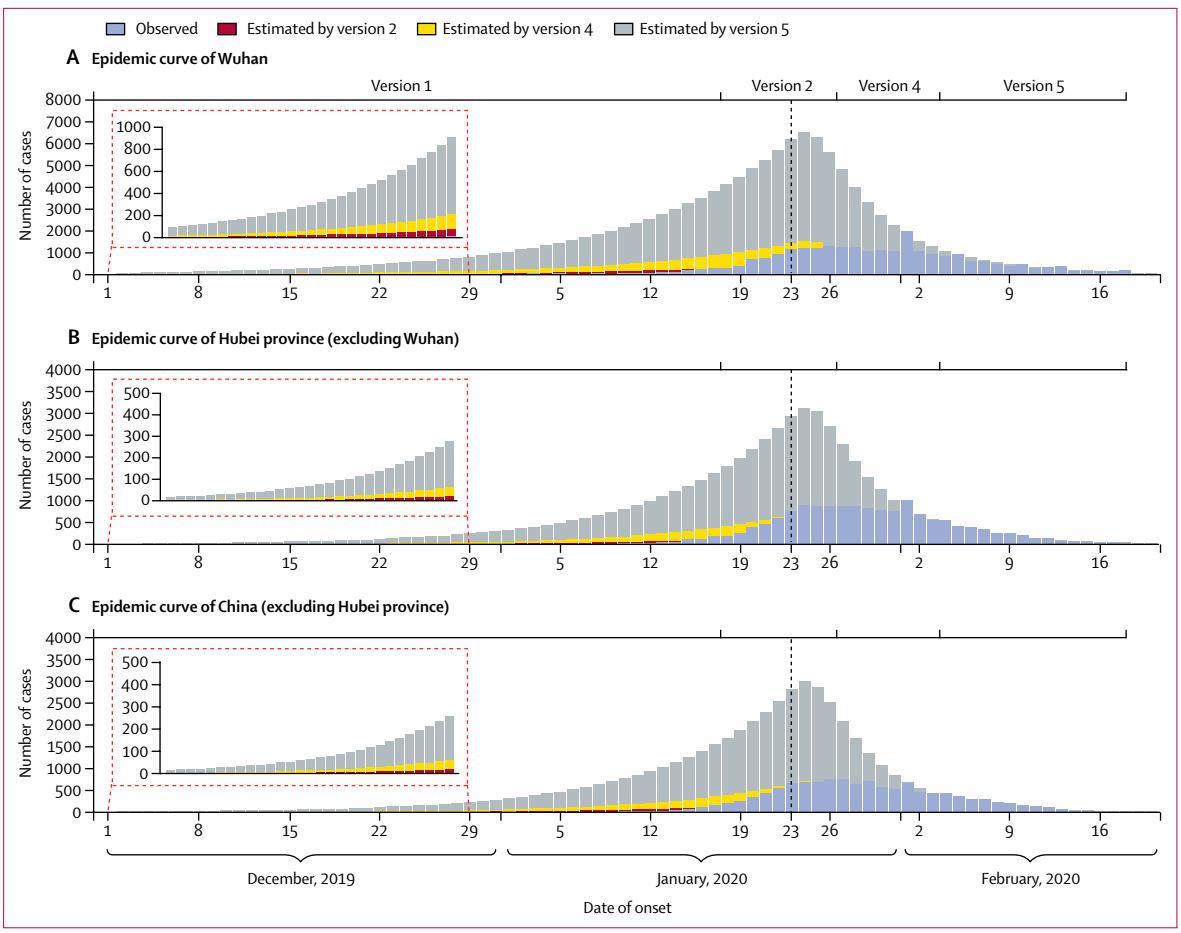

Figure 3: **Occurrence of COVID-19 cases by different case definitions**
COVID-19 cases by date of illness onset in Wuhan (A), Hubei province excluding Wuhan (B), and other provinces in mainland China excluding Hubei province (C). Observed cases are indicated with blue bars. Red bars indicate estimates for case definition version 2, yellow bars for case definition version 4, and grey bars for case definition version 5. COVID-19=coronavirus disease 2019.

pneumonia with or without a certain epidemiological link but did not need to have a virological confirmation of infection. In the sixth edition, this criterion for diagnosis of clinically confirmed cases was removed and no distinction was made between cases inside or outside Hubei province. In the seventh edition, serology was added as an additional option for laboratory confirmation.

We modelled the effects of changes in case definition from version 1 to version 2, from version 2 to 4, and from version 4 to 5. We did not explore the effects of changing from version 2 to 3 because version 3 applied the same definitions for suspected and confirmed COVID-19 cases as version 2 but only included updates to the severity classifications and therefore had no effect on the incidence or the epidemic curve. We were not able to explore the change after version 5 as we only analysed data up to Feb 20, 2020, which included just the first 2 days after the release of version 6. We were not able to find publicly available information on incidence of cases by illness onset date after Feb 20, and had to censor our analysis at that point.

The changes in case definitions had a clear effect on the proportion of infections that were identified and counted as confirmed cases. As of Feb 20, 2020, there were 55 508 confirmed cases in China, among which 27 000 were from Wuhan, 16 000 were from the rest of Hubei province, and 13 000 were from the rest of China. We estimated that the mean onset-to-reporting delay was 8·6 days (95% CrI 7·4–10·1) and the 95th percentile of this distribution was 15·7 days (13·0–20·1). Allowing for a 10 day backfill of cases, we estimated that when the case definitions were changed from version 1 to 2, version 2 to 4, and version 4 to 5, the proportion of infections being identified as COVID-19 cases was increased by 7·1 times (95% CrI 4·8–10·9) from version 1 to 2, 2·8 times (1·9–4·2) from version 2 to 4, and 4·2 times (2·6–7·3) from version 4 to 5 (figure 2).

Based on the model, we estimated that if the case definitions from version 5 had been applied throughout the outbreak, and there had been sufficient availability of laboratory testing with RT-PCR from the early phase of the epidemic, 232 000 cases (95% CrI 161 000–359 000)

	Wuhan	Hubei province excluding Wuhan	China excluding Hubei province
Growth rate, per day			
With adjustment for changes in case definitions	0·08 (0·06–0·10)	0·10 (0·08–0·12)	0·10 (0·08–0·12)
Without adjustment	0·15 (0·14–0·17)	0·18 (0·13–0·28)	0·19 (0·16–0·24)
Doubling time, days			
With adjustment for changes in case definitions	8·7 (7·3–10·8)	7·0 (5·8–8·8)	7·0 (5·8–8·7)
Without adjustment	4·5 (4·1–4·8)	3·9 (2·4–5·3)	3·6 (2·9–4·3)

Data are growth rates and doubling times with 95% credible intervals.

Table: **Estimates of the epidemic growth rate and doubling time before Jan 23, 2020, with or without adjustment for changes in case definitions**

could have met the case definition and could have been detected by Feb 20, 2020, of which 127 000 cases (86 000–198 000) were from Wuhan, 55 000 (38 000–86 000) were from the rest of Hubei province excluding Wuhan, and 50 000 (34 000–78 000) were from the rest of China excluding Hubei (figure 3). Among the 127 000 cases that we estimated in Wuhan by Feb 20, we estimated that there could have been approximately 11 000 infections (95% CrI 7000–21 000) that met version 5 of the case definition with illness onset by Jan 1, 2020. In the observed data, there were 114 confirmed COVID-19 cases with illness onset by Jan 1, 2020, corresponding to around 1% of our estimated total. Before Jan 23, we estimated that 92% (95% CrI 88–95) of cases were undetected.

We estimated that after implementation of control measures on Jan 23, the growth rate declined substantially to less than 0, from 0·08 to –0·15 in Wuhan, which was a change of –0·23 (95% CrI –0·27 to –0·20). The corresponding changes in growth rate were –0·26 (–0·30 to –0·22) for the rest of Hubei province excluding Wuhan, and –0·28 (–0·32 to –0·25) for the rest of China excluding Hubei. These findings suggested that the control measures were very effective, reducing the effective reproductive number to well less than 1. Specifically, using a mean serial interval of 7·5 days,[8] the effective reproductive numbers were reduced to 0·21–0·28 for the three regions, while the estimates were reduced to 0·36–0·44 with a mean serial interval of 4·7 days.[10]

After adjusting for the changes in case definitions, we estimated that the epidemic growth rate before Jan 23, 2020, was around 0·08 to 0·10 and the doubling time was around 7·0 to 8·7 days for these three geographical areas, and the differences among them were not substantial (table). If instead the change in case definitions was unaccounted for, the growth rate would have been substantially overestimated and the doubling time would have been substantially underestimated (table). Using a growth rate of 0·08–0·10 with a mean serial interval of 7·5 days[8] would lead to R_0 estimates in the range of 1·8–2·0. If we instead used the growth rate estimates of 0·15–0·19 (table), we would obtain R_0 estimates in the range of 2·8–3·5. In a sensitivity analysis, using a mean serial interval of 4·7 days,[10] R_0 estimates using a growth rate of 0·08–0·10 were in the range of 1·4–1·5, while using the growth rate estimates of 0·15–0·19 (table) we would obtain R_0 estimates in the range of 1·9–2·2.

In a sensitivity analysis allowing for 15 days of backfill each time the case definition changed, the proportion of infections being identified as COVID-19 cases was increased by 3·0–8·8 times. We estimated that 253 000 cases (95% CrI 158 000–436 000) would have met the case definition and could have been detected by Feb 20, 2020. These estimates were slightly higher, but as expected, given the backfill period was longer.

Discussion

We estimated that changes in case definitions of COVID-19 in China led to stepwise increases in the proportion of all infections identified as cases, by 7·1 times from version 1 to 2, 2·8 times from version 2 to 4, and 4·2 times from version 4 to 5. Overall, we estimated that around 232 000 cases could have been confirmed in the first wave of COVID-19 in China by late February, 2020, if, hypothetically, version 5 of the case definitions had been used throughout as opposed to the 55 508 confirmed cases reported. The number of individuals who were infected is likely to be greater than 232 000 because many mild cases were not tested or confirmed, and some infections were asymptomatic.[12] We estimated that many cases were undetected when using an earlier case definition, which is consistent with the study by Li and colleagues, which estimated that around 85% of cases were undetected before Jan 23, when case definition 2 was used.[13] Our results were also consistent with another modelling study indicating clear differences between earlier and later cases identified in Wuhan.[14] The estimated case numbers were considerably higher than the observed case numbers, suggesting a so-called clinical iceberg phenomenon, which is common for a disease that can cause both mild and severe illnesses like COVID-19.[15] As a result, when case definitions were broadened, more mild cases could be detected.

The introduction of clinically confirmed cases in the fifth version of the case definitions allowed many individuals who were highly suspected to be cases but who did not receive a virological test due to insufficient testing capacity to be isolated and treated in time, allowing reallocation of laboratory testing resources for identifying and then isolating cases in the community as part of the containment efforts. This category was removed within a week in the sixth edition of case definitions because laboratory testing was deemed sufficient to confirm all cases and the clinically confirmed category was unnecessary.[16] However, confirmation of viral pneumonia with radiological evidence could be an important alternative for diagnosis and surveillance of

COVID-19 in locations with limited laboratory testing capacity, and could also be a good option if or when a surge in COVID-19 consultations exceeds local laboratory capacity. This method could be combined with testing a portion of the clinically confirmed cases to correct the actual case numbers afterwards.[17]

Case definitions are often developed for outbreak investigations in which the objective is to identify the source of infections,[18] while case definitions are used for surveillance only later if an epidemic occurs. In the case of the COVID-19 epidemic in China, the initial case definitions for COVID-19 allowed investigation of potential animal exposures and infections epidemiologically linked with the epicentre, Wuhan, but might not capture cases linked with wider areas potentially affected by COVID-19.[19] Similarly, the earlier case definitions had more specific requirements for clinical manifestations given the limited knowledge of the novel virus, leading to a low sensitivity for case identification including an under-detection of milder infections.[19,20] As evidence for the clinical spectrum of COVID-19 became available, the case definition was updated to account for this information. Changes in the availability and use of testing can also lead to a similar effect. For example, when the USA increased their laboratory testing capacity from less than 300 to more than 10 000 cases tested per day from late-February to early-March,[21] the total case numbers increased rapidly.

Our analysis suggests that estimates of key epidemiological parameters using epidemic curves could be biased if they do not account for such changes in case definitions. Specifically, we found that if we had estimated the exponential growth in the epidemic curve without accounting for the changes in case definitions, we would have substantially overestimated the growth rate and substantially underestimated the doubling time (table). There are several high estimates of growth rates and R_0, some of which might suffer from this particular bias,[22,23] although one study divided their analysis by case definition from version 1 to 4 and noted that the estimated reproductive number was an upper bound.[24] Other high estimates of growth rates or R_0 based on epidemic curves by reporting date might have overestimated transmissibility because of the shortening in onset-to-reporting delays as the epidemic progressed. It should be noted that our estimates of reproductive numbers were based on growth rates and serial intervals, and therefore they were sensitive to the assumption of serial intervals. However, the estimates of effect of changing case definitions were based on exponential growth models, which were insensitive to changes in the serial interval.

Our findings also suggest caution might be needed for analyses of the trajectories of epidemic curves elsewhere. Epidemics could appear to be growing faster than they actually are, because of rapid expansions in testing practices. The availability of and resolve for laboratory testing will also be a major factor shaping epidemic curves,[25] which will be important to guide the public health responses. Because of the limited capacity for confirmation tests, Switzerland, for example, might have stopped testing mild cases and restricted tests to those who are more ill,[26] and other countries might need to adopt the same approach as case numbers increase—radiological confirmation could be a potential alternative to track the incidence of hospitalised cases. In addition to accounting for changes in case definitions or testing capacity, analyses of epidemic trajectories should also take into account the implementation of public health measures.

Our study has some limitations. A first limitation is that we did not formulate an individual-based mechanistic transmission model but used a simple model with exponential growth and then exponential decay (figure 2). Future research could explore more complex dynamic models to account for the other factors that are potentially affecting transmission. One example is to allow for the marginal effects of different types of interventions, such as lockdowns and other distancing interventions that were introduced at different times towards the end of January, 2020, in addition to accounting for the changes in case definitions. As a consequence, analyses of the effects of interventions in China should be evaluated with caution if they do not account for the changes in case definitions. Second, we only explored the effect of changing case definitions on the epidemic curve of all cases because there is no publicly available data for the epidemic curves by severity. There were changes in the classification of severe cases, and future studies are needed to explore their effect on estimates of fatality risk. Third, we were only able to collect data for the epidemic curve up to Feb 20, 2020. Therefore, we cannot evaluate the effect of changes in the case definition from version 5 to 6 and from version 6 to 7, although case numbers have declined substantially since Feb 20, 2020.

In conclusion, we have shown that changes in case definitions had a substantial effect on the proportion of all infections identified as cases as time progressed, and therefore also had a substantial effect on the epidemic curve. We estimated that there could have been 232 000 cases by Feb 20, 2020, if, hypothetically, version 5 of the case definitions had been used throughout the epidemic. Still, this would be an underestimate of the number of infections up to that point because it would not have captured some mild or asymptomatic cases. Serological studies will be useful to estimate the cumulative incidence of infections.

Contributors

TKT, PW, GML, and BJC were responsible for study design. PW, YL, and EHYL were responsible for data collection. TKT, PW, and YL were responsible for data analysis. TKT, PW, EHYL, GML, and BJC were responsible for data interpretation. TKT wrote the first draft of the manuscript. All authors contributed to the final draft.

Declaration of interests

BJC reports honoraria from Sanofi Pasteur and Roche. All other authors declare no competing interests.

For **data and code** see https://github.com/timktsang/covid19_casedef

Data sharing

The data and computer code (in R languages) for the data analysis can be downloaded online.

Acknowledgments

This project was supported by a commissioned grant from the Health and Medical Research Fund, Food and Health Bureau, Government of the Hong Kong Special Administrative Region.

Editorial note: The *Lancet* Group takes a neutral position with respect to territorial claims in published maps and institutional affiliations.

References

1 Gregg MB. Field Epidemiology. New York, NY: Oxford University Press, 2002.

2 Zhu N, Zhang D, Wang W, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med* 2020; **382:** 727–33.

3 Wuhan Municipal Health Commission. Report on unexplained viral pneumonia. Jan 5, 2020. http://wjw.wuhan.gov.cn/front/web/showDetail/2020010509020 (accessed March 12, 2020; in Chinese).

4 Wu P, Hao X, Lau EHY, et al. Real-time tentative assessment of the epidemiological characteristics of novel coronavirus infections in Wuhan, China, as at 22 January 2020. *Euro Surveill* 2020; **25**.

5 National Health Commission of the People's Republic of China. Release of 7th edition of case defintions. 2020. http://www.nhc.gov.cn/yzygj/s7653p/202003/46c9294a7dfe4cef80dc7f5912eb1989.shtml (accessed March 12, 2020; in Chinese).

6 WHO. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). 2020. https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf (accessed March 12, 2020).

7 Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet* 2020; **395:** 689–97.

8 Li Q, Guan X, Wu P, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N Engl J Med* 2020; **382:** 1199–207.

9 Wallinga J, Lipsitch M. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc Biol Sci* 2007; **274:** 599–604.

10 Nishiura H, Linton NM, Akhmetzhanov AR. Serial interval of novel coronavirus (COVID-19) infections. *Int J Infect Dis* 2020; **93:** 284–86.

11 Gilks WR, Richardson S, Spiegelhalter D. Markov chain Monte Carlo in Practice. London: Chapman & Hall, 1996.

12 The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19)—China, 2020. http://weekly.chinacdc.cn/en/article/id/e53946e2-c6c4-41e9-9a9b-fea8db1a8f51 (accessed April 14, 2020).

13 Li R, Pei S, Chen B, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science* 2020; published online March 16. DOI:10.1126/science.abb3221.

14 Kucharski AJ, Russell TW, Diamond C, et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect Dis* 2020; published online March 11. http://dx.doi.org/10.1016/S1473-3099(20)30144-4.

15 Wong JY, Kelly H, Ip DK, Wu JT, Leung GM, Cowling BJ. Case fatality risk of influenza A (H1N1pdm09): a systematic review. *Epidemiology* 2013; **24:** 830–41.

16 CRJ online. The capacity of PCR testing in Hubei has been increased. 2020. http://news.cri.cn/20200221/1f2057ea-de3b-707b-bfdc-e467ef7a596d.html (accessed March 12, 2020; in Chinese).

17 Lipsitch M, Hayden FG, Cowling BJ, Leung GM. How to maintain surveillance for novel influenza A H1N1 when there are too many cases to count. *Lancet* 2009; **374:** 1209–11.

18 United States Centers for Disease Control and Prevention. Principles of Epidemiology in Public Health Practice. 2011. https://www.cdc.gov/csels/dsepd/ss1978/index.html (accessed March 12, 2020).

19 Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020; **395:** 497–506.

20 Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. *JAMA* 2020; **323:** 1239.

21 Centers for Disease Control and Prevention of United States. Testing in the U.S. 2020. https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/testing-in-us.html (accessed April 4, 2020).

22 Liu T, Hu J, Kang M, et al. Transmission dynamics of 2019 novel coronavirus (2019-nCoV). *bioRxiv* 2020; published online Jan 26. DOI:10.2139/ssrn.3526307 (preprint).

23 Sanche S, Lin YT, Xu C, Romero-Severson E, Hengartner N, Ke R. High contagiousness and rapid spread of severe acute respiratory syndrome coronavirus 2. *Emerg Infect Dis* 2020; published online April 7. DOI:10.3201/eid2607.200282.

24 Zhang J, Litvinova M, Wang W, et al. Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study. *Lancet Infect Dis* 2020; published online April 2. https://doi.org/10.1016/S1473-3099(20)30230-9.

25 Lipsitch M, Swerdlow DL, Finelli L. Defining the epidemiology of covid-19—studies needed. *N Engl J Med* 2020; **382:** 1194–96.

26 Unisanté, University Center for General Medicine and Public Health in Lausanne. The current Swiss strategy for responding to the new coronavirus epidemic. 2020. https://coronavirus.unisante.ch/en (accessed April 10, 2020).