Now-casting the COVID-19 epidemic: The use case of Japan, March 2020

Stephan Glöckner PhD[1,2], Prof Gérard Krause MD, [1,2], Prof Michael Höhle PhD[3]

Affiliations

1 Helmholtz-Centre for Infection Research, Department of Epidemiology, Braunschweig, Germany

2 German Center for Infection Research (DZIF),

3 Department of Mathematics, Stockholm University, Stockholm, Sweden

Corresponding Author
Stephan Glöckner PhD MSc
Department of Epidemiology, Helmholtz Centre for Infection Research
Braunschweig 38124
Germany
Email: stephan.gloeckner@helmholtz-hzi.de
Phone: +49 531 6181 3125

Abstract

Background

Reporting delays in disease surveillance impair the ability to assess the current dynamic of an epidemic. In continuously updated epidemic curves, case numbers for the most recent epidemic week or day usually appear to be lower than the previous, suggesting a decline of the epidemic. In reality, the epidemic curve may still be on the rise, because reporting delay prevents the most recent cases to appear in the case count. In context of the COVID-19 epidemic and for countries planning large international gatherings, such as the Summer Olympic Games in Japan 2020, the ability to assess the actual stage of an epidemic is of outmost importance.

Methods

We applied now-casting onto COVID-19 data provided by the nCoV-2019 Data Working Group to evaluate the true count of cases, by taking into account reporting delays occurring between date of symptom onset and date of confirmation.

Findings

We calculated a decrease of reporting delay, from a median delay of ten days in calendar week four 2020 to six days in calendar week eight, resulting in an overall mean of 4.3 days. The confidence intervals of the now-casting indicated an increase of cases in the last reporting days, while case country in that same time period suggested a decline.

Interpretation

As a specific use case this tool may be of particular value for the challenging risk assessment and risk communication in the context of the Summer Olympic Games in Japan 2020 and similar situations elsewhere.

Introduction

Reporting delay of cases is a common obstacle for real-time risk assessment and management of epidemics. In continuously updated epidemic curves, case numbers for the most recent epidemic week or day usually appear to be lower than the previous, suggesting a decline of the epidemic. In reality however, the epidemic curve may still be on the rise, because reporting delay prevents the most recent cases to already appear in the case count. This phenomenon often introduces misperceptions in the general public and in media coverage as can again be seen in the current COVID-19 epidemic.[1]

Reporting delay can be decomposed into several parts: (a) the time passing from symptom onset until the patient seeks medical service, (b) the time passing between this service and initiation of a laboratory test, (c) the time between this test initiation and completion of its result, (d) the time between test completion and its registration at the local health authority level and (e) the time passing from transmitting the case report from that local level to the (usually national) health authority level that will generate and publish the epidemic curve.[2]

Technology and rapid procedures can minimize the delay occurring from the first contact with the health care system and appearance of a case in the daily case count at national level, but they can hardly reduce the delay between appearance of first symptoms and contact to the health care service.[2] The latter is subject to a variety of factors such as severity of the disease, public awareness for it, individual benefit to seek health services, accessibility to those services, some of which may also be subject to change during an epidemic. For all those reasons, reporting delay is not only specific to a country and its surveillance and health care system, but to the disease in questions and the phase of the epidemic.

For the COVID-19 in China, The World Health Organization (WHO) published a median reporting delay of 12 days (range 8-18 days) at the start of the outbreak and three days (range 1-7 days) in early February.[3] Disease transmission models report a reporting delay larger than ten days during the start of the novel coronavirus outbreak and four days delay after the implementation of the quarantine procedures on January 27th.[4] A study monitoring transmission potential of COVID-19 in Singapore estimated a reporting delay of 6.9 days.[5] In contrast, the 2012 MERS-COV outbreak in Saudi-Arabia had a four day reporting delay[6] and during the 2018 Measles outbreak, Japan showed a mean reporting delay of 4.5 days.[7]

Now-casting uses previous data to estimate the delay distribution while taking into account that cases with long-delays are currently not observed due to right-truncation with the aim to adjust epidemic curves for "observed-but-not-yet-reported cases".[8,9] On the occasion of the 2011 STEC/HUS outbreak in Germany the now-casting method was further optimized for the specific situation of being in the midst of an acute outbreak with interventions to actively reduce the delay.[10]

We report here on an improved concept of now-casting, that allows the adjustment of epidemic curves by taking into account that information about symptom onsets may not be available for a relevant proportion of the case reports. We demonstrate the approach with data from the COVID-19 outbreak in Japan, because of its high case counts, accessibility to data details and the particular importance of such an approach for the risk assessment for the upcoming Summer Olympic Games, hosted in Tokyo, July to August 2020. The aim of this approach is to reduce and transparently display uncertainty on the real time status of epidemic curves.

Methods

For the purpose of this research, we define the reporting delay to be the time between the onset of disease symptoms of a case and the time the case report is available at the public health authority that generated the epidemic curve.[2]

We used the line-list dataset provided by the nCoV-2019 Data Working Group[11] and WHO situation reports,[12] to allow for a comparison between the two sources. On March 6th 2020, nCoV-2019 Data Working Group contained 204 cases (excluding all cases from the Diamond Princess and with the last case reported on March 2nd 2020), of which 131 (64%) contained date of symptom onset and date of case confirmation. For the remaining 73 (36%) cases without information on symptom onset, we assume that the case still had a symptom onset date and it is just missing, because it could not be readily identified. We imputed the respective date using a generalized additive Weibull regression model for both the scale and shape parameter with week of confirmation as a covariate.[13]

The now-casting approach consists of a Bayesian approach for modelling delay by which case reports appear in the continuously updated epidemic curve (also known as reporting triangle) and provides full predictive distributions for the daily delay-adjusted number of symptom onsets.[10] We used a sliding window of the past 14 days to estimate the time-varying delay distribution, in order to account for the assumption that the reporting delay decreases as the outbreak progresses, due to aforementioned changes in the health care seeking and reporting procedures. Uncertainty of the delay imputation was transported by repeating the imputation 100 times and averaging the resulting now-casting quantities.

We developed this approach within R[14] using the open-source R package surveillance[15] and developed a Shiny application[16,17], allowing users to perform now-casting without need for sophisticated programming skills and made it available at:

https://helmholtz-epid.shinyapps.io/nowcasting_dashboard/ .

Results

The comparison of the nCoV-2019 Data Working Group line-list data and the data gathered from the WHO Situation reports for date of confirmation shows similar trends, while the WHO data seems to lag one day behind, starting February 14, 2020. WHO reported for February 5th a cumulative total of 33 cases, but set the case count back on February 6th to a cumulate total of 25 cases. We set the cases of February 6th to unknown and restarted the plotting of the epidemic curve at this day (see Figure 1).
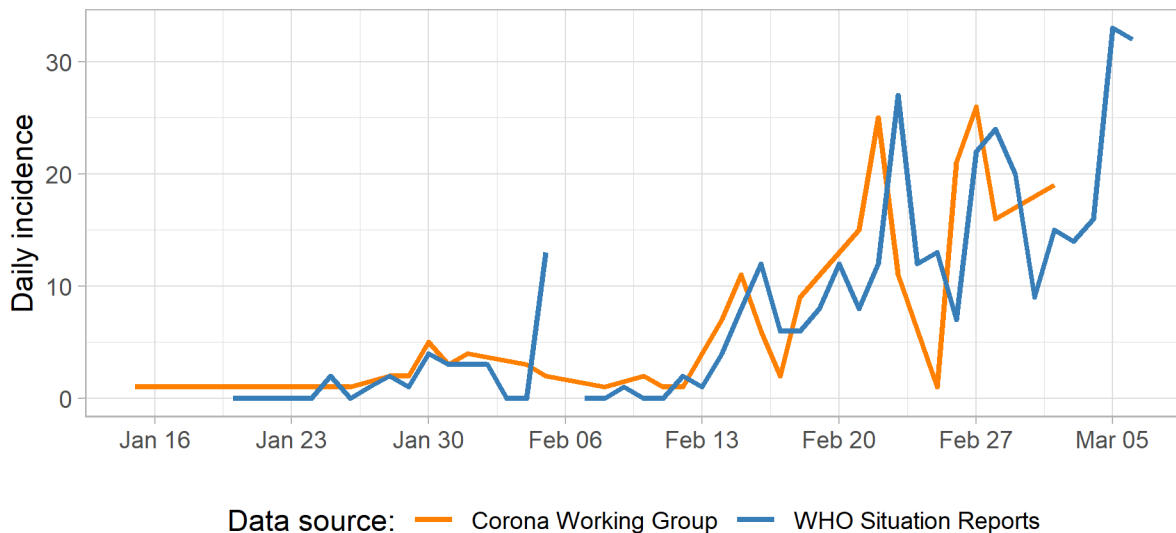

Figure 1 COVID-2019 Epidemic curve of confirmed cases for Japan as of March 6th 2020, comparing data from nCoV-2019 Data Working Group and WHO.

Our modelling of symptom onsets show a decrease of reporting delay over the time (see Figure 2), from a median of ten days in calendar week four to six days in calendar week eight.
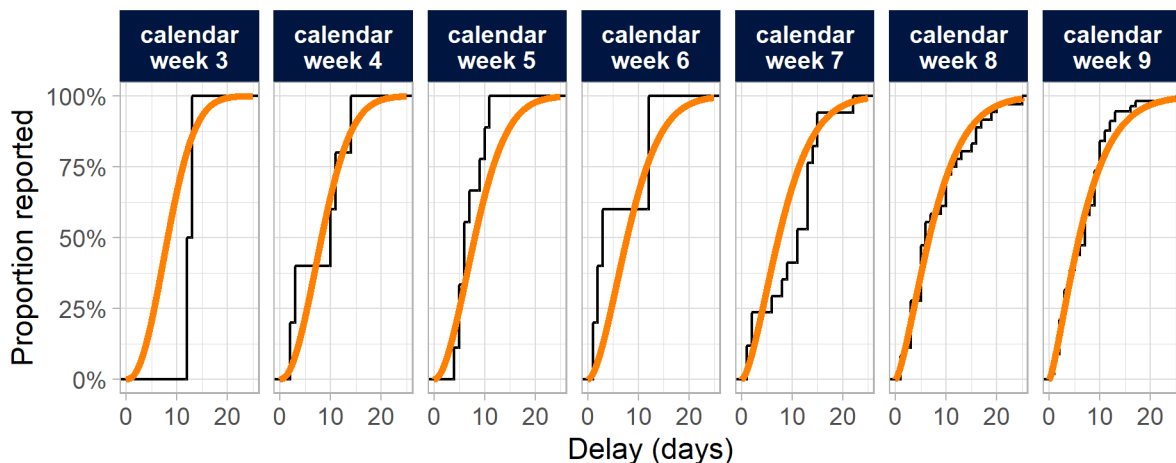

Figure 2 Empirical (black line) and model fitted (orange line) cumulative distribution function for the reporting delay for each calendar week.

Figure 3 shows the results of the now-cast for the line-list data on March 2nd 2020 and done for the previous 14 days. Case counts more than nine days ago receive little adjustment, because the delay distribution is such that almost all cases are available. As an example, the number of now-casted cases on February 29th 2020 is 6.2 on top of the observed 1.9 with a 95% prediction interval of 0.5 to 23.2. From the adjusted cases between February 28th and March 1st it becomes clear that despite the observed

counts being low there appears to be an increasing trend. This trend is confirmed by the WHO data on confirmed cases for March 2nd and beyond.
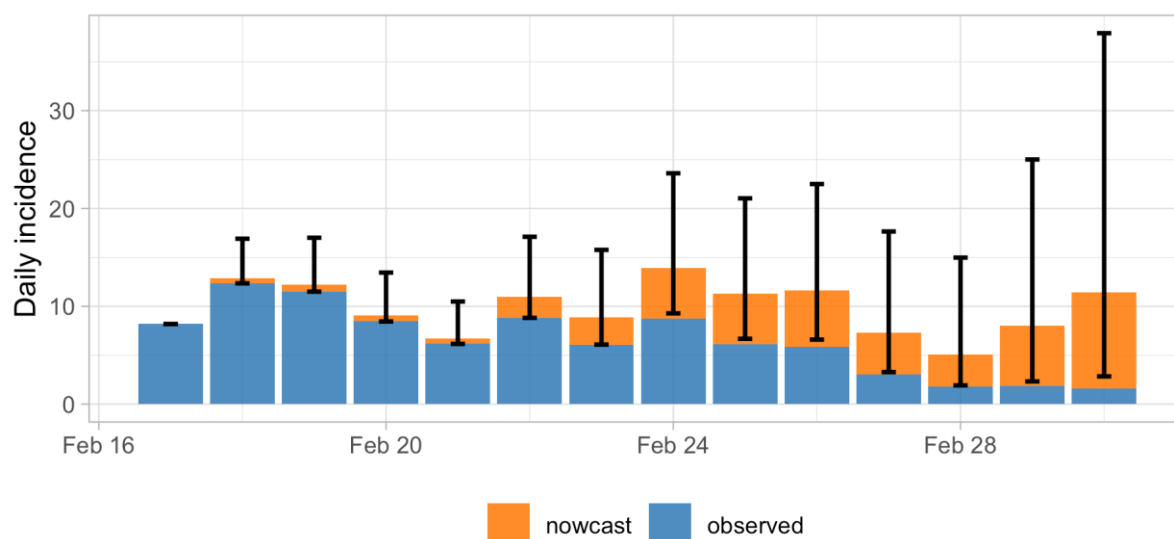

Figure 3 Now-cast for the epidemic curve of COVID-19 symptom onsets in Japan performed on March 2nd 2020 for the previous 14 days. The orange area indicates the median of the predictive distribution for the number of cases missing. Furthermore, 95% prediction intervals for the total counts are shown.

Discussion

Our approach helps to assess whether the apparent decline is imminent to any real-time display of epidemic curves, is likely to reflect a true decline or more likely to reflect an artefact resulting from reporting delays. This artifact is immanent to any surveillance system and its resulting epidemic curves if continuously updating all data available at the time of publication. This approach also visualizes easily the level of uncertainty with this regard and therefore facilitates transparent public health risk communication.

In contrast to prediction models aiming to estimate the future development of an outbreak, the now-cast approach has the more humble, but no less valuable, objective to estimate whether an outbreak is still on the rise or already declining. In comparison to previous now-casting approaches, the method presented here, takes into account that a) reporting delay changes over time during an epidemic and also that b) for a relevant proportion of case reports not all data are available for computation of case specific reporting delays.

Public awareness is likely to have a major impact on surveillance sensitivity and also timeliness[18]: The WHO reported the first Corona case in Japan on January 20th 2020, with four additional cases on January 25th. During this time, the media had already reported intensively about COVID-19 and therefore public awareness was likely to have been already high. When the first case of SARS-CoV-2 was reported in China on December 1st 2019, the Japanese National Institute of Diseases established a real time PCR for this type of virus until January 16th 2020. Furthermore, a telephone hotline for professionals within the health system and public communications were activated.[19] These and other activities may explain why the reporting delay in Japan was fairly short to begin with and diminished further as the outbreaks progressed, compared to COVID-19 outbreaks in other countries.[3] The delays displayed for calendar week nine as part of the imputation might experience a small downwards bias due to cases with long delays not being available at the date of query, however, this is then later taken into account by now-casting.

The added value of this now-casting approach increases with the length of the reporting delay – the larger the delays the more important is the adjustment. As illustrated per 95% prediction intervals in figure 3, uncertainty of now-casts is obviously highest for the days are closest to the actual day of the now-cast itself. We used publicly available data for our analysis, which tends to be less complete than the data available to national surveillance agencies. This slightly limits the precision of our now-cast, but also indicates the even higher value of now-casting when used within a nation's surveillance system. For this reason, we are currently building this tool as a permanent functionality into the dashboard of the Surveillance Outbreak Response Management System (www.sormas.org) to further enhance its role in COVID-19 response.[20–23]

Via the open-source R package surveillance applied within the user-friendly shiny application, public health surveillance institutions can easily apply this now-cast approach in their country for COVID-19 or any other disease. No system or country specific configurations are needed. As a specific use case this tool may be of particular value for the challenging risk assessment and risk communication in the context of the Summer Olympic Games in Tokyo July to August 2020 and similar situations elsewhere.

Acknowledgements

References

1. The New York Times. New Cases in China Appear to Be Slowing. 2020 Feb 18 [cited 2020 Mar 11]; Available from: https://www.nytimes.com/2020/02/18/world/asia/china-coronavirus.html

2. Swaan C, van den Broek A, Kretzschmar M, Richardus JH. Timeliness of notification systems for infectious diseases: A systematic literature review. Lau EH, editor. PLOS ONE. 2018 Jun 14;13(6):e0198845.

3. WHO. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19) [Internet]. Geneva: World Health Organisation; [cited 2020 Mar 4] p. 1–40. Report No.: 16-24 February 2020. Available from: https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf

4. Tuite AR, Fisman DN. Reporting, Epidemic Growth, and Reproduction Numbers for the 2019 Novel Coronavirus (2019-nCoV) Epidemic. Ann Intern Med [Internet]. 2020 Feb 5 [cited 2020 Feb 29]; Available from: https://annals.org/aim/fullarticle/2760912/reporting-epidemic-growth-reproduction-numbers-2019-novel-coronavirus-2019-ncov

5. Tariq A, Lee Y, Roosa K, Blumberg S, Yan P, Ma S, et al. Real-time monitoring the transmission potential of COVID-19 in Singapore, February 2020 [Internet]. Epidemiology; 2020 Feb [cited 2020 Mar 5]. Available from: http://medrxiv.org/lookup/doi/10.1101/2020.02.21.20026435

6. Ahmed AE. Diagnostic delays in 537 symptomatic cases of Middle East respiratory syndrome coronavirus infection in Saudi Arabia. Int J Infect Dis. 2017 Sep;62:47–51.

7. Akhmetzhanov AR, Lee H, Jung S-M, Kinoshita R, Shimizu K, Yoshii K, et al. Real Time Forecasting of Measles Using Generation-dependent Mathematical Model in Japan, 2018. PLoS Curr. 2018 Oct 15;10.

8. Donker T, Boven M van, Ballegooijen WM van, Klooster TMV, Wielders CC, Wallinga J. Nowcasting pandemic influenza A/H1N1 2009 hospitalizations in the Netherlands. Eur J Epidemiol. 2011;26(3):195–201.

9. Zeger SL, See LC, Diggle PJ. Statistical methods for monitoring the AIDS epidemic. Stat Med. 1989 Jan;8(1):3–21.

10. Höhle M, an der Heiden M. Bayesian nowcasting during the STEC O104: H4 outbreak in Germany, 2011. Biometrics. 2014;70(4):993–1002.

11. nCoV-2019 Data Working Group. Epidemiological Data from the nCoV-2019 Outbreak: Early Descriptions from Publicly Available Data [Internet]. GitHub. 2020 [cited 2020 Mar 10]. Available from: https://github.com/beoutbreakprepared/nCoV2019

12. Brown E. data2019nCoV: Data on the COVID-19 Outbreak. 2020.

13. Rigby RA, Stasinopoulos DM. Generalized additive models for location, scale and shape. J R Stat Soc Ser C Appl Stat. 2005;54(3):507–554.

14. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2019. Available from: https://www.R-project.org/

15. Salmon M, Schumacher D, Höhle M. Monitoring Count Time Series in R : Aberration Detection in Public Health Surveillance. J Stat Softw [Internet]. 2016 [cited 2020 Mar 10];70(10). Available from: http://www.jstatsoft.org/v70/i10/

16. Chang W, Cheng J, Allaire JJ, Xie Y, McPherson J. shiny: Web Application Framework for R [Internet]. 2019. Available from: https://CRAN.R-project.org/package=shiny

17. Borges B, Allaire JJ. flexdashboard: R Markdown Format for Flexible Dashboards [Internet]. 2017. Available from: https://CRAN.R-project.org/package=flexdashboard

18. Jha A. Stories, data, and plans: Preparing for the next pandemic [Internet]. Sustainable Cities. 2020 [cited 2020 Mar 4]. Available from: https://blogs.worldbank.org/sustainablecities/stories-data-and-plans-preparing-next-pandemic

19. Gallego V, Nishiura H, Sah R, Rodriguez-Morales AJ. The COVID-19 outbreak and implications for the Tokyo 2020 Summer Olympic Games. Travel Med Infect Dis. 2020 Feb;101604.

20. Krause G, Poll B. Global Goods Addressing Outbreaks: SORMAS and COVID-19 [Internet]. Digital Square. 2020 [cited 2020 Mar 10]. Available from: https://digitalsquare.org/blog/2020/3/4/global-goods-addressing-outbreaks-sormas-and-covid-19

21. Krause G, Walter C, Dörrbecker J. Mobile Seuchenüberwachung: Dem Weg der Viren auf der Spur. Dtsch Arztebl Int. 2020;117(10):A-494-.

22. Tom-Aba D, Silenou BC, Doerrbecker J, Fourie C, Leitner C, Krause G. Global Goods Maturity Model Indicates Surveillance Outbreak Response Management & Analysis System (SORMAS) as a Mature Digital Health "Global Goods" (Preprint) [Internet]. JMIR mHealth and uHealth; 2019 Sep [cited 2020 Mar 10]. Available from: http://preprints.jmir.org/preprint/15860

23. Silenou BC, Tom-Aba D, Adeoye O, Arinze CC, Oyiri F, Suleman AK, et al. Use of Surveillance Outbreak Response Management and Analysis System for Human Monkeypox Outbreak, Nigeria, 2017–2019. Emerg Infect Dis. 2020 Feb;26(2):345–9.

Figure 1 Epidemic Curve
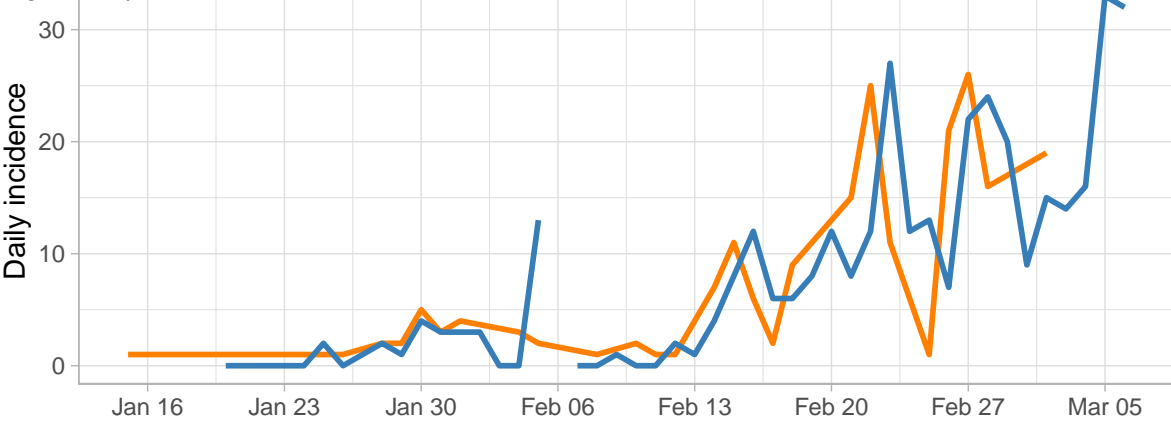
Figure 2 Delay Proportion

Figure 3 Nowcast