How should we present the epidemic curve for COVID-19?

Jean-Paul R. Soucy, Sarah A. Buchan, Kevin A. Brown

Affiliations: Division of Epidemiology, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada (Soucy, Buchan, Brown) and Public Health Ontario, Toronto, Ontario, Canada (Buchan, Brown)

Correspondence to: Jean-Paul R. Soucy (jeanpaul.soucy@mail.utoronto.ca)

Funding: This project received no funding.

Competing interests: The authors declare no competing interests.

Abstract

Epidemic curves are used by decision makers and the public to infer the trajectory of the COVID-19 pandemic and to understand the appropriateness of current response measures. Symptom onset date is commonly used to date cases on the epidemic curve in public health reports and dashboards. However, third-party trackers often plot cases on the epidemic curve by the date they were publicly reported by the public health authority. These two curves create very different impressions of epidemic progression. On April 1, the epidemic curve for Ontario, Canada based on public reporting date showed an accelerating epidemic, whereas the curve based on a proxy variable for symptom onset date showed a rapidly declining epidemic. This illusory downward trend (the “ghost trend”) is a feature of epidemic curves anchored using date variables earlier in time than the date a case was publicly reported, such as symptom onset date or sample collection date. This is because newly discovered cases are backdated, creating a perpetual downward trend in incidence due to incomplete data in the most recent days. Public reporting date is not subject to backdating bias and can be used to visualize real-time epidemic curves meant to inform the public and policy makers.

Keywords: Epidemics, COVID-19, SARS-CoV-2, Communicable Diseases, Epidemiology

Epidemic curves are used to infer the current trajectory of the COVID-19 pandemic and to inform policy surrounding the COVID-19 response, including physical distancing measures, as well as public perception of their necessity. When constructing epidemic curves, there are a variety of ways to date incident cases, such as by the symptom onset date or the date the case was publicly reported. Given the inevitable delay between when infections occur and when infections are detected and reported as cases, it is critical that the epidemic trajectory be communicated as quickly and as accurately as possible. The method selected to date incident cases for epidemic curves produced in real-time (rather than retrospectively, after the outbreak has ended) can profoundly alter the impression of the trajectory of the epidemic. In this analysis, we will demonstrate that epidemic curves based on symptom onset date (or a proxy thereof), as commonly presented in the media and by public health authorities across the world, are affected by an optimistic bias regarding the trajectory of COVID-19. They give the false impression that the epidemic is rapidly approaching an end, even when this is not the case.

What are epidemic curves?

The epidemic curve

An epidemic curve shows the trend in the incidence of a disease in a population through time (1). A key feature of epidemic curves is that they are generally descriptive and can be constructed in numerous different ways, such as through the choices of case definition, subpopulations or the method of dating incident cases. This contrasts with mathematical models such as the classic susceptible-infected-recovered model of Kermack and McKendrick (2), which are prescriptive, embedding assumptions and parameters regarding disease fundamentals. Epidemic curves can provide clues about the characteristics of novel diseases. For example, the epidemic curve of the previously unknown Legionnaires Disease showed cases tightly clustered in time, suggesting a common environmental source, ultimately identified as the air conditioning of the hotel hosting the Pennsylvania American Legion convention (3).

Epidemic curves have achieved newfound prominence during the COVID-19 pandemic. Early public health messaging focused on “flattening the curve”—referring to the epidemic curve—so that the daily number of new infections did not overwhelm the capacity of the healthcare system (4). More recently, epidemic curves have been used to guide further tightening or loosening of restrictions and to monitor the effects of these decisions on the trajectories of localized epidemics. Consequently, representations of the epidemic curve are a mainstay of public health and media reports and dashboards (5–9). This fact makes the choices made in how these curves are presented all the more important.

Heterogeneity in epidemic curves

An idealized epidemic curve would plot incident cases of disease based on the date of infection (on the x-axis), as this represents the true epidemiology of an infectious disease. However, date of infection is rarely known with certainty and is not available in real time, as the case must first be identified by public health authorities. Thus in practice, other dates, subsequent to the date of infection, are used to date incident cases. These dates include symptom onset date, sample collection date, lab reported date and public reporting date.

For the publication of real-time epidemic curves, public health authorities commonly select symptom onset date, since it is the closest measurable date to the date of infection. On the other hand, third-party trackers such as those run by Johns Hopkins University (7), The COVID-19 Tracking Project (10) and the COVID-19 Canada Open Data Working Group (11) use primarily public reporting date, in part

because more detailed date variables are not always available. These two dating methods produce very different epidemic curves, which can strongly affect the perceived trajectory of COVID-19 cases for reasons that will be explored in the subsequent section.

Why does the choice of epidemic curve matter?

Public health surveillance systems are continually updated as new information becomes available. A case is entered into the system after the public health authority receives the result from the laboratory, after which a case investigation begins and the data are publicly reported. Over the next few hours, days and weeks, additional information may be entered and previous information revised. This may include the date the case developed symptoms, if the case was ever symptomatic at all.

The ghost trend

Real-time epidemic curves can be subject to a “ghost trend”—an illusory downward bias to the epidemic trajectory in the most recent days of data. The cause of the ghost trend is backdating: cases reported today will often be assigned to earlier days on the epidemic curve, meaning few cases will appear in the most recent days. For example, a case may have a symptom onset date days or even weeks prior to the date the public health authority becomes aware of the case.

Bias due to backdating, leading to the ghost trend, is exacerbated by delays in administering tests, processing tests and communicating test results to public health authorities, since dates such as symptom onset date and date of sample collection will be commensurately further back in time relative to the public reporting date. Consequently, if a region is overwhelmed with new cases, the ghost trend bias will be magnified even further. However, the ghost trend is also an issue of natural history, human behaviour and healthcare systems, not just technology. Consider the following scenario: tests are processed instantaneously at point of collection, results are immediately transmitted directly into the public health surveillance system and symptom onset date is ascertained perfectly in real time. Epidemic curves based on symptom onset date would still show a ghost trend because of inevitable delays and variability between when an individual first experiences symptoms and when they seek and receive a test.

Symptom onset date has other practical limitations for constructing real-time epidemic curves due to delays and difficulties in pinpointing it. Delays can occur because it may take several days to conclude a case investigation (particularly during periods of high case load) (12). Recall issues and patient incapacitation or death also complicate measurement. For asymptomatic cases, the variable does not apply at all.

An example using the epidemic curve for Ontario, Canada

To circumvent the issues inherent to measuring symptom onset date, public health surveillance systems in Canada generally use a proxy variable for this date called the “episode date”. If symptom onset date is absent, the episode date is defined as the earliest available date in the following hierarchy: specimen collection date, laboratory testing date and the date the public health authority received the report from the laboratory (13). Unless the earliest date (either symptom onset date or sample collection date if the former cannot be ascertained) is the same as the public reporting date, all cases will be backdated relative to the public reporting date. During the early pandemic period in Ontario, symptom onset date was generally the final date entered into the system (if it was entered at all). This caused cases plotted on an epidemic curve according to episode date to be progressively backdated as more information

became available. This meant that while older cases on the epidemic curve did not change much over time, newer cases had their dates continuously pushed back as the episode date was replaced by earlier dates, ending with the true symptom onset date (if it could be ascertained). Thus, the most recent days on the epidemic curve will have few cases—creating the ghost trend.

Figure 1: Epidemic curve of COVID-19 in Ontario, Canada for March 1, 2020 to April 1, 2020 by episode date (left) and public reporting date (right) using three datasets extracted between April 1 and May 20. The pink curve represents the real-time epidemic curve, while the green and blue curves are those observed 1 and 7 weeks later, respectively. Note that “public reporting date” in this figure refers to the date the case was entered into the provincial surveillance system.

Consider Figure 1, which displays epidemic curves for COVID-19 in Ontario, Canada from March 1 to April 1, 2020 plotted by two different date variables: episode date (the proxy for symptom onset date) and public reporting date. We display these curves based on datasets extracted on three different dates: April 1 (pink, the real-time dataset), April 8 (green, 1 week later) and May 20 (blue, 7 weeks later, after which edits to cases from March have largely ceased). The real-time dataset contains 2,793 cumulative cases from January to April 1. However, the cumulative number of cases up to April 1 in the May 20 data extract differs greatly according which variable is used to date cases. Using public reporting date, the cumulative number of cases is 2,772, nearly unchanged from the real-time dataset. This is expected, since cases dated by public reporting date should only change to correct data entry errors and to remove duplicates and individuals no longer meeting the case definition. In contrast, the number of cumulative cases by episode date in this data extract is 5,922, more than double that of the real-time dataset. This is because many cases reported at the end of March had symptom onset dates in mid-March, and most cases with symptom onset dates in late March were not identified and reported until April.

The real-time epidemic curve by symptom onset date shows a pronounced ghost trend, creating the impression that incidence peaked mid-March and is rapidly declining toward the end of the month. The later data extracts correct this biased impression: daily incidence was climbing throughout the entire month of March. On the other hand, the real-time epidemic curve based on public reporting date reliably communicated the trend of increasing incidence throughout the month of March.

How are epidemic curves being used during COVID-19?

Epidemic curves for COVID-19 based on a proxy variable for symptom onset date are common. For example, the Public Health Agency of Canada's epidemiological summary (5) presents the curve using this date variable and it is commonly used for visualizing the epidemic curve in dashboards and reports from other public health organizations as well (8,14–19). In these figures, caveats are often given in the form of a shaded area on the graph covering recent dates, with text indicating that recent cases may not yet be reported.

When epidemic curves are used to assess disease trajectory in the midst of an outbreak, curves plotted by symptom onset date can be misleading. This possibility is not merely theoretical: these curves have been misinterpreted by public officials to justify loosening public health measures in at least two US states, Georgia (20) and Iowa (21). This phenomenon of incomplete, subsequently backdated incidence data is well known in the realm of mortality statistics, where manual coding of deaths often results in a lag before these deaths are counted in official datasets (22). It has also been discussed in the context of other diseases (23,24) and of charting COVID-19 deaths (25,26). Never has this phenomenon been so consequential as it is in our present moment.

Conclusion

Real-time epidemic curves are useful for visualizing the trajectory of the COVID-19 pandemic. Epidemic curves based on symptom onset date are important for understanding the epidemiology of an outbreak in retrospect, since they form an approximation of the infection curve. However, because newly discovered cases are backdated when plotted by symptom onset date, this leads to very few cases appearing in the most recent days of the epidemic curve, creating an illusory “ghost trend” of falling case numbers. This limits the usefulness of epidemic curves by symptom onset date for determining the disease trajectory in real time. By contrast, curves constructed using the date a case was publicly reported do not suffer from this optimistic bias.

The COVID-19 pandemic is far from over. As the virus continues to sweep across the globe and many regions experience resurgences, it is essential that epidemic curves reliably communicate the present trajectory of detected cases. Public reporting date can be used to visualize real-time epidemic curves meant to inform the public and policy makers.

Acknowledgements

The authors wish to thank Dr. Shelley L. Deeks and Dr. Matthew P. Fox for their feedback on this manuscript.

References

1. Rosenberg PS. Epidemic Curve. In: Wiley StatsRef: Statistics Reference Online. Hoboken, NJ: John Wiley & Sons; 2015.
2. Kermack WO, McKendrick AG, Walker GT. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London Series A, Containing Papers of a Mathematical and Physical Character*. 1927;115(772):700–21.
3. Fraser DW, Tsai TR, Orenstein W, Parkin WE, Beecham HJ, Sharrar RG, et al. Legionnaires' Disease: Description of an Epidemic of Pneumonia. *N Engl J Med*. 1977;297(22):1189–97.
4. Jones DS, Helmreich S. The Shape of Epidemics. *Boston Review*. 2020 [cited 2020 Sep 21]; Available from: <https://bostonreview.net/science-nature/david-s-jones-stefan-helmreich-shape-epidemics>
5. Public Health Agency of Canada. Epidemiological summary of COVID-19 cases in Canada. 2020 [cited 2020 Sep 21]. Available from: <https://health-infobase.canada.ca/covid-19/epidemiological-summary-covid-19-cases.html>
6. Government of the United Kingdom. Coronavirus (COVID-19) in the UK: Cases. 2020 [cited 2020 Sep 21]. Available from: <https://coronavirus.data.gov.uk/cases>
7. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases*. 2020;S1473309920301201.
8. State of Michigan. Coronavirus - Michigan Data. 2021 [cited 2021 Jan 16]. Available from: https://www.michigan.gov/coronavirus/0,9753,7-406-98163_98173---,00.html
9. CTV News. Tracking every case of COVID-19 in Canada. CTV News. 2021 [cited 2021 Jan 16]; Available from: <https://www.ctvnews.ca/health/coronavirus/tracking-every-case-of-covid-19-in-canada-1.4852102>
10. The COVID Tracking Project. The COVID Tracking Project. 2020 [cited 2020 Sep 21]. Available from: <https://covidtracking.com/>
11. Berry I, Soucy J-PR, Tuite A, Fisman D. Open access epidemiologic data and an interactive dashboard to monitor the COVID-19 outbreak in Canada. *CMAJ*. 2020 Apr 14;192(15):E420.
12. Ontario Agency for Health Protection and Promotion (Public Health Ontario). COVID-19 regional incidence and time to case notification in Ontario. Toronto, Canada: Queen's Printer for Ontario; 2020.
13. Ontario Agency for Health Protection and Promotion (Public Health Ontario). iPHIS User Guide: Enteric and Zoonotic Diseases. Toronto, Canada: Queen's Printer for Ontario; 2019.
14. City of Toronto. COVID-19: Status of Cases in Toronto. 2020 [cited 2020 Sep 21]. Available from: <https://www.toronto.ca/home/covid-19/covid-19-latest-city-of-toronto-news/covid-19-status-of-cases-in-toronto/>

15. BC Centre for Disease Control. British Columbia Weekly COVID-19 Surveillance Report: September 11–September 17, 2020. 2020 [cited 2020 Sep 21]. Available from: http://www.bccdc.ca/Health-Info-Site/Documents/BC_Surveillance_Summary_Sept_17_2020.pdf
16. Mississippi State Department of Health. Coronavirus COVID-19. 2021 [cited 2021 Jan 7]. Available from: <https://msdh.ms.gov/coronavirus>
17. Ohio Department of Health. COVID-19 Dashboard. 2021 [cited 2021 Jan 16]. Available from: <https://coronavirus.ohio.gov/wps/portal/gov/covid-19/dashboards/overview>
18. Ottawa Public Health. Daily COVID-19 Dashboard. 2020 [cited 2020 Sep 21]. Available from: <https://www.ottawapublichealth.ca/en/reports-research-and-statistics/daily-covid19-dashboard.aspx>
19. San Diego County. Daily COVID-19 Data Update 2021-1-6. 2021 [cited 2021 Jan 7]. Available from: https://www.sandiegocounty.gov/content/dam/sdc/hhsa/programs/phs/Epidemiology/COVID-19_Daily_Status_Update.pdf
20. Wooten N. COVID-19 data from Georgia’s Department of Health could be ‘misleading,’ experts say. Ledger Enquirer. 2020 [cited 2020 Sep 22]. Available from: <https://www.ledger-enquirer.com/news/coronavirus/article242416536.html>
21. Ojeda H. West Branch restaurants prefer carry-out despite lifted restrictions. Iowa City Press-Citizen. 2020 [cited 2020 Sep 22]. Available from: <https://www.press-citizen.com/story/news/2020/05/01/west-branch-restaurants-prefer-carry-out-despite-lifted-restrictions/3064806001/>
22. CDC. Technical Notes: Provisional Death Counts for Coronavirus Disease (COVID-19). 2020 [cited 2020 Sep 21]. Available from: https://www.cdc.gov/nchs/nvss/vsrr/covid19/tech_notes.htm
23. Finger F, Funk S, White K, Siddiqui MR, Edmunds WJ, Kucharski AJ. Real-time analysis of the diphtheria outbreak in forcibly displaced Myanmar nationals in Bangladesh. BMC Medicine. 2019;17(1):58.
24. Centers for Disease Control and Prevention. Interpretation of Epidemic (Epi) Curves during Ongoing Outbreak Investigations. 2018 [cited 2021 Jan 16]. Available from: <https://www.cdc.gov/foodsafety/outbreaks/investigating-outbreaks/epi-curves.html>
25. Walker P. Is There a Right Way to Chart COVID-19 Deaths Over Time? The COVID Tracking Project. 2020 [cited 2020 Sep 21]. Available from: <https://covidtracking.com/blog/is-there-a-right-way-to-chart-covid-19-deaths-over-time>
26. Mathieu E. Why do COVID-19 deaths in Sweden always appear to decrease in the last 10 days? Our World in Data. 2020 [cited 2021 Jan 7]. Available from: <https://ourworldindata.org/covid-sweden-death-reporting>