Nowcasting and Forecasting the 2022 U.S. Mpox Outbreak: Support for Public Health Decision Making and Lessons Learned

Kelly Charniga, PhD[1,2]*, Zachary J. Madewell, PhD[1,3], Nina B. Masters, PhD[4], Jason Asher, PhD[5], Yoshinori Nakazawa, PhD[2], Ian H. Spicknall, PhD[6]

[1] These first authors contributed equally to this article.

[2] Division of High Consequence Pathogens and Pathology, National Center for Emerging and Zoonotic Infectious Diseases, CDC

[3] Center for Global Health, CDC

[4] Epidemic Intelligence Service, CDC

[5] Center for Forecasting and Outbreak Analytics, CDC

[6] Division of Sexually Transmitted Disease Prevention, National Center for HIV, Viral Hepatitis, STD, & TB Prevention, CDC

* Corresponding author, email: ruq7@cdc.gov

Keywords: monkeypox virus, modeling, epidemic, real-time analysis, situational awareness

Abstract

Mpox is a zoonotic disease endemic in Central and West Africa. In May 2022, an outbreak of mpox characterized by human-to-human transmission was detected in multiple non-endemic countries. We performed nowcasting and forecasting for the 2022 mpox outbreak in the United States using the R package EpiNow2. We generated nowcasts/forecasts at the national level, by Census region, and for jurisdictions reporting the greatest number of mpox cases. Modeling results were shared for situational awareness within the Centers for Disease Control and Prevention (CDC) Mpox Response and publicly on the CDC website. We retrospectively evaluated forecast predictions at four key phases during the outbreak using three metrics, the weighted interval score, mean absolute error, and prediction interval coverage. We compared the performance of EpiNow2 with a naïve Bayesian generalized linear model (GLM). The EpiNow2 model had less probabilistic error than the GLM during every outbreak phase except for the early phase. We share our experiences with an existing tool for nowcasting/forecasting and highlight areas of improvement for the development of future tools. We also reflect on lessons learned regarding data quality issues and adapting modeling results for different audiences.

Background

The 2022 mpox (formerly known as monkeypox) outbreak is the first major infectious disease outbreak since the COVID-19 pandemic and was declared a Public Health Emergency of International Concern by the World Health Organization on July 23, 2022(World Health Organization, 2022). As of April 13, 2023, a total of 86,956 confirmed cases have been reported in 110 countries and territories(U.S. Centers for Disease Control and Prevention, 2022a). Unlike COVID-19, mpox is a disease known to be endemic in West and Central Africa for decades; it is caused by monkeypox virus (MPXV), a zoonotic orthopoxvirus(Beer and Rao, 2019). Classical symptoms involve fever, headache, muscle aches, fatigue, lymphadenopathy, and rash(U.S. Centers for Disease Control and Prevention, 2022e). Human-to-human MPXV transmission occurs through close contact with infectious material from skin lesions, respiratory secretions during prolonged face-to-face contact, and fomites, such as linens and bedding(WHO/ECDC, 2022). The 2022 mpox outbreak began in May and spread rapidly in non-endemic countries. This outbreak was characterized by human-to-human transmission of MPXV through close physical contact (often associated with sexual activities) and has disproportionately affected gay, bisexual, and other men who have sex with men(UK Health Security Agency, 2022).

During a public health crisis, such as the mpox outbreak, difficult and rapid decisions with limited available data are often required(Thompson et al., 2022). Infectious disease models may assist with informing policy by predicting the magnitude and duration of an outbreak or epidemic, evaluating characteristics of pathogen transmission, such as transmissibility, and designing vaccination strategies, among others(Buchwald et al., 2020; Madewell et al., 2021b). However, infectious disease models are often complex, integrating data from heterogenous sources with many parameter assumptions that are subject to uncertainty. These aspects make it challenging to effectively implement such models and communicate the results and potential limitations to decision makers, other public health partners, and the general public(Lloyd-Smith et al., 2015; Madewell et al., 2021a).

During the COVID-19 pandemic, the state of the art of outbreak analysis advanced considerably(Chakraborty et al., 2022). Methods and tools for estimating key epidemiological parameters,

such as the effective reproduction number, R_t, were developed and shared in real-time(Abbott et al., 2020). Monitoring R_t, the average number of secondary cases caused by a single infected individual in a large population, during an outbreak is useful for assessing transmission dynamics and evaluating the effectiveness of public health measures (e.g., vaccination, contact tracing, isolation, and quarantine)(Cori et al., 2013). Nowcasts and forecasts were produced by numerous research groups around the globe(Arslan et al., 2021; Birrell et al., 2021; Cramer et al., 2022; Spreco et al., 2022), the results of which were instrumental for decision makers weighing possible control measures(Biggerstaff et al., 2021) such as lockdowns and social distancing. Outbreak forecasting predicts specific outcomes, (e.g., number of cases, deaths, or hospitalizations) at some specific future times (e.g., weeks, months, etc.), whereas nowcasting estimates those outcomes for the current time, accounting for delays in reporting.

In this manuscript, we share our experience nowcasting and forecasting the mpox outbreak, describe challenges faced vis-a-vis data quality, parameter estimation, and model application and propose recommendations to improve nowcasting and short-term forecasting efforts for future outbreaks.

Methods

Nowcasting/forecasting the mpox outbreak

We used data on probable and confirmed mpox cases in the United States (see “Case definition” in Supplementary methods) reported to CDC by state and local public health jurisdictions from May 17, 2022, through March 16, 2023. Data were submitted in several different formats throughout the outbreak period. These formats included: a CDC-operated call center through the Emergency Operations Center (EOC), a long and a short case report form (CRF), and the National Notifiable Diseases Surveillance System (NNDSS). Cases could have data submitted via more than one source, and jurisdictions could update data on cases after initial submission (Supplementary methods). All reported data were processed in CDC’s Data Collation and Integration for Public Health Event Response (DCIPHER) platform, an instance of Palantir Foundry (Palantir Technologies Inc, Denver, CO). DCIPHER is a secure, cloud-based data integration, analytics, and situational awareness platform used by the Centers for Disease Control and Prevention (CDC) and other federal, state, tribal, local, and territorial public health jurisdictions to

collect, collaborate on, and share public health data.(U.S. Department of Health and Human Services, 2020). DCIPHER collates data of differing origin, structure, and purpose to provide near real-time insights into public health problems, with the goal of providing a complete picture of situational awareness.

We considered three approaches for estimating R_t which are implemented in the R packages EpiEstim (version 2.2-4)(Cori et al., 2021), earlyR (version 0.0.5)(Jombart et al., 2020), and EpiNow2 (version 1.3.2)(Abbott et al., 2021a) (Supplementary methods). Initially, we used all three methods to estimate R_t at the national level as well as for jurisdictions reporting the highest incidence of mpox. Although estimates of the historical range of the serial interval of mpox were available at the start of the outbreak, they were based on data from the Democratic Republic of Congo, which reflected largely non-sexual household spread(Fenner et al., 1988). We considered these historical parameter estimates as a starting point for early outbreak analysis, using them (along with sensitivity analyses) until new estimates were generated. Updated estimates characterized by the mean and standard deviation were needed for the global outbreak given the novel mode of transmission. In June 2022, we were able to use an estimated mean serial interval (i.e., the period of time between symptom onset in the primary case and symptom onset in the secondary case) of 9.8 days (95% credible interval [CrI]: 5.9 – 21.4) from 17 case pairs reported by the United Kingdom(UK Health Security Agency, 2022). At that time, symptom onset date was available for most reported cases, and imported cases were still contributing to a high proportion of MPXV transmission. EpiEstim results were considered the most appropriate at this stage of the outbreak because this method accounts for imported vs. locally acquired cases, has a stable codebase, is widely used, and is computationally efficient(Nash et al., 2022).

In July 2022, we started using EpiNow2, which uses a similar approach as EpiEstim (a branching process model, Supplementary methods), but it better accounts for reporting delays and incorporates multiple sources of uncertainty(Abbott et al., 2020); for example, it removes noise associated with weekend effects and uses random walks for temporal smoothing. Forecasting is supported internally for R_t, number of infections, cases by date of report, and growth rate. Unlike EpiEstim, EpiNow2 does not

distinguish between imported vs. locally acquired infections. EpiNow2 is the most computationally expensive approach, requiring longer model run times (Supplementary methods). The model assumes that testing procedures, surveillance effort, and reporting delays remain constant over the estimation period. To use EpiNow2, cases by date of report must be provided as well as the generation time distribution (the time between infection of a primary and secondary case), incubation period distribution (the time between infection and symptom onset in a case), and any other delay distributions (e.g., the delay between symptom onset and report date). The model estimates the number of new cases by date of report, number of cases by their date of infection, R_t, and time-varying growth rate. Estimates over the last 16 days of the time-series are based on partial data due to the presumption of reporting delays. Input parameters and methods for adjusting for right-truncation evolved as we learned more about the outbreak.

Communication methods

R_t estimates were shared internally through Situational Reports and leadership meetings and publicly through CDC's Center for Forecasting and Outbreaks Analytics' (CFA's) Technical Reports(U.S. Centers for Disease Control and Prevention, 2022c) and CDC's public-facing mpox website(U.S. Centers for Disease Control and Prevention, 2022d). Estimates were generated for distribution at least once per week. The technical reports were intended for scientific audiences. The purpose of sharing these results was to improve understanding of the outbreak and inform further scientific inquiry.

Performance assessment methods

We chose eight dates during four key outbreak phases to retrospectively evaluate our short-term (one-week-ahead) forecasts of reported mpox cases generated from EpiNow2: 1. one month into the outbreak prior to exponential growth (June 13 and June 27), 2. during exponential growth (July 5), 3. at the outbreak peak (July 27), 4. during the declining phase (September 6, September 19, October 11, and December 5). Ideally, the same day of the week would be used, but some historical versions of the dataset were not available for this analysis and some dates fell on national holidays when new cases were not reported. Like the real-time analyses, we used rash onset date as the first reference date to define the

reporting delay distribution for all eight time points, while the second reference date changed over time (Table S1).

We used three metrics to evaluate the forecasts. Our primary metric was the weighted interval score (WIS). For each of the eight time points considered, WIS was computed for each daily prediction and averaged across the seven-day forecast. The WIS measures the consistency of a group of prediction intervals with an observed value (probabilistic accuracy). The WIS is positive, and lower values correspond to smaller error (Supplementary methods)(Cramer et al., 2022). To evaluate the error in the forecast's point estimate, we used the mean absolute error (MAE), which was computed as $MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|$, where y_i is the observed number of mpox cases on day i, $\hat{y}_i$ is the median forecast on day i, and N = 7(Cramer et al., 2022). We also used prediction interval (PI) coverage rates, which check the degree to which the model provides calibrated predictions. Coverage rates are calculated by determining the proportion of times the 50% or 90% PIs included the observed value (for example, a well-calibrated forecast would have a 50% PI coverage close to 0.50. Also see Supplementary methods)(Cramer et al., 2022).

We compared the performance of EpiNow2 with a naïve Bayesian generalized linear model (GLM, Supplementary methods). We calculated the relative WIS and relative MAE for EpiNow2 and the GLM as $\theta_{EpiNow2,GLM} = \frac{mean\ score\ of\ EpiNow2}{mean\ score\ of\ GLM}$, where the mean score is the average of the models' performance (WIS or MAE) over all eight dates evaluated. If $\theta_{EpiNow2,GLM}$ was less than 1, that indicated the forecasts generated by EpiNow2 had less error than the GLM, whereas $\theta_{EpiNow2,GLM} > 1$ indicated EpiNow2 performed worse.

For both EpiNow2 and the GLM, we removed recent cases (reported in the last 3 – 5 days) from the time series to adjust for right truncation of the data for all four outbreak phases (Table S1). Forecasts were evaluated using mpox data as of March 16, 2023. We used an updated report date ("modeling date", see below) as the basis for the comparison.

Results

Challenges of nowcasting/forecasting

While jurisdictions are required to report every person with a positive mpox lab result to CDC, only minimal data are required to satisfy case reporting requirement (e.g., case ID and reporting jurisdiction). CDC also changed the requested data variables over time to try and reduce reporting burden. While CDC asks jurisdictions to collect and report additional data variables to achieve situational awareness and surveillance goals, jurisdictional case surveillance systems may not have aligned to CDC's requested data variables, not all cases could be reached and fully investigated, and finally not all data variables could be shared with CDC if restricted by local laws or practices. Received data were subjected to additional manual data cleaning to standardize formats and correct obvious data entry errors. As a result, even key data variables such as demographic characteristics (e.g., age, race/ethnicity, HIV status, vaccination) was inconsistent across jurisdictions, precluding detailed sub-analyses. For example, out of 29,921 cases in DCIPHER through December 31, 2022, 21,480 (71.8%) were missing HIV status, 16,474 (55.1%) were missing smallpox vaccination, 3,913 (13.1%) were missing gender identity, 3,172 (10.6%) were missing race, 2,928 (9.8%) were missing ethnicity, and 250 (0.8%) were missing age. The timing and frequency of data submission varied between jurisdictions and changed over the course of the outbreak. Some jurisdictions reported case data in near real-time whereas others submitted a bolus of cases all at once, the latter of which caused large, artificial spikes in the time-series. There were instances of duplicate cases being reported from several jurisdictions which may be attributed to the changes in reporting processes. Spurious cases at the end of the time series had to be investigated (and usually removed) because they artificially inflated the nowcasts/forecasts. These data issues required us to monitor the model output closely and modify the code as needed.

In early July 2022, during the transition from the long CRF to the short CRF, reporting of mpox cases to CDC started to lag in some jurisdictions, especially those most affected by the outbreak. Jurisdictions were publicly reporting more cases on their websites than what CDC had received reports for. This led to a lengthy case reconciliation process during which case data uncertainty prevented it from being used for nowcasting/forecasting at the national level. Also in July, an increasing proportion of cases

were reported with missing symptom onset dates (from 26% to 53% for rash onset date between June 13 and July 5). To ensure each case had a date associated with it for plotting epidemic curves, a new report date field was created which we started using for nowcasting/forecasting. The new date was determined by the following hierarchy: orthopoxvirus test date, date of call to the call center, date the short CRF was created, and the long CRF timestamp (Figure S1). In September 2022, the definition of report date changed, redefined as the earliest among all available dates for a case, including symptom onset, which facilitated improved visualization of epidemic curves. The change was also motivated by issues with the "created at" dates and the new date fields available for jurisdictions using NNDSS. It was a combined decision vetted with all teams. However, this new date presented challenges for its use in the EpiNow2 framework because the delay from symptom onset to report date would be zero for some cases.

Successes of nowcasting/forecasting

During the case reconciliation process in July, publicly available data through health department websites was used for subnational analyses (e.g., California(California Department of Public Health, 2022) and New York City(New York City Department of Health and Mental Hygiene, 2022)). We used WebPlotDigitizer(Rohatgi, 2021) to extract time series data from pdfs when the underlying data were not available for download.

Given the problems stated previously with the report date field, we worked to create a new report date field specifically for nowcasting/forecasting, deemed "modeling date." At the time, most mpox cases were submitted to DCIPHER via the short CRF, which included orthopoxvirus test date and the date the case was first received by DCIPHER. We tried to align dates from the other sources with orthopoxvirus test date, finding that the date in which the record was created was least preferred due to the spikes in the time series caused by bulk uploading data. Thus, the new "modeling date" was determined by a hierarchy across the different data streams where priority was given to diagnosis date, orthopoxvirus test date, orthopoxvirus test confirmation date, case investigation start date, orthopoxvirus sample collection date, date of call to CDC call center, the earliest among report date (to public health department, county, or state), date CDC announced case, and the earliest date the case was entered into DCIPHER.

In October, we updated estimates of the serial interval for rash onset of 7.0 days (95% CrI 5.8 – 8.4) from 40 case pairs and incubation period for rash onset of 7.5 days (95% CrI 6.0 – 9.8) from 35 U.S. case-patients and used those as model inputs for EpiNow2(Madewell et al., 2023). These data were obtained through the collaboration of several U.S. jurisdictions on a special study. The estimated serial interval for the 2022 outbreak was on the lower end of the historical range observed in the Democratic Republic of Congo (7 – 23 days)(Fenner et al., 1988).

Adapting model output and communicating nowcasts/forecasts

We adapted the presentation of our results for a scientific/technical audience and the general public. The default plots from EpiNow2 include three panels: cases by date of report, cases by date of infection, and R_t. Green represents estimates based on complete data, orange represents estimates based on partial data, and purple represents the forecast (the default is seven days). Gray bars in the top panel show the actual time series of reported cases, while gray bars in the middle panel show the back-calculated infection time series. For the Technical Reports, Situational Reports, and IM meetings, we removed the middle panel (Figure 1)(U.S. Centers for Disease Control and Prevention, 2022c). For the website, we only showed R_t, removed the forecast, and removed the 20% credible intervals to minimize confusion (Figure 2)(U.S. Centers for Disease Control and Prevention, 2022d). We included a simple description of the plot that could be understood by non-experts. In accordance with CDC's Data Modernization Initiative(U.S. Centers for Disease Control and Prevention, 2022f), a national effort aimed at modernizing state and national core data and surveillance infrastructure, data for the underlying plots were made available for download as comma-separated values (csv) files with the Technical Reports.

Sub-national analyses revealed some differences between regions regarding the start of the outbreak, when it peaked, and how long it lasted (Figures 3 – 4). For example, Figure 4 demonstrates a later introduction date and slightly longer tail for Texas compared to other jurisdictions.

Performance assessment

The GLM had lower WIS compared to EpiNow2 for early phase of the outbreak (Table 1); however, during all other phases, EpiNow2 had a slight advantage. The relative WIS was 0.89 over all

eight time points considered, indicating that EpiNow2 had on average 11% less probabilistic error than the GLM.

EpiNow2 had lower MAE than the GLM for six out of eight time points, but performance was similar: the relative MAE was 0.96. In other words, EpiNow2 had only 4% less point error than the GLM.

Overall, predictions were moderately well calibrated. For the 90% PI, EpiNow2 achieved coverage rates within 10% of the desired coverage level for seven out of eight time points compared to six out of eight for the GLM. For the 50% PI, EpiNow2 achieved coverage rates within 10% of the desired coverage level for only two out of eight time points compared to five out of eight for the GLM.

Qualitative results of the nowcasts/forecasts are shown in Figure S2. The 90% CrIs from EpiNow2 were very wide for the early phase of the outbreak, while the GLM had large uncertainty around the outbreak's peak. Both models underestimated reported mpox cases for the seventh time point on October 11 which could be due to a discrepancy in the data available at the time versus the ground-truth data (Figure S3). This time point had the lowest PI coverage rates.

Discussion

We performed nowcasting/forecasting to inform the U.S. response to the 2022 mpox outbreak in real-time. Validation showed that the method implemented in EpiNow2 predicted case counts reasonably well, but improvements are needed around key time periods such as the outbreak's peak. One reason that the nowcasts/forecasts did not always align with reality is because the definition of report date changed over time, while the ground truth data were constructed using modeling date. We found a higher WIS for EpiNow2 in the early phase of the outbreak compared to the GLM which could be due to choices of priors for parameters (e.g., wide intervals for R_t).

Subnational analyses allowed us to better understand the spatial heterogeneity of the epidemic which may be attributed to differences between jurisdictions in terms of composition (e.g., population age structure, density, and contact patterns) and public health activities (e.g., vaccination, surveillance methods, frequency of testing)(De Nicola et al., 2022) as well as the timing and frequency of case reporting. One limitation of nowcasting/forecasting at the subregional or jurisdictional level is that the

effects of bulk uploads are more apparent, resulting in greater uncertainty (wider credible intervals). Another limitation is that movement between jurisdictions could have a greater impact on subnational estimates, as mobility is not accounted for in our approach. Finally, some jurisdictions stopped reporting rash onset date, which decreased the sample size available for estimating the reporting delay distribution over time.

Recommendations Regarding Data Quality

Nowcasting/forecasting methods perform best when the underlying surveillance data are accurate, timely, and complete, but they are often sub-optimal and variable as the outbreak evolves: while data may improve as an outbreak progresses, they may re-deteriorate once the outbreak slows and intensity of effort is low. Fortunately, the quality and frequency of data improved over the course of the U.S. mpox outbreak. Communicating with specific jurisdictions about our priority dates for modeling improved data quality. Close collaboration between epidemiologists/modelers and informaticians, including the use of an issue tracking system in DCIPHER, also facilitated quick investigation and resolution of data errors.

EpiNow2 Limitations

The main limitation of EpiNow2 is its steep learning curve due to limited documentation of package functions and few reports of its application to other outbreaks. Increasing commenting in the code, creating more tutorials or vignettes, and developing a graphical user interface could help. At the same time, we acknowledge that this tool was developed during the COVID-19 pandemic by academics supporting that response. In the aftermath of the COVID-19 pandemic, lack of funding and current incentive structures in academia (i.e., rewarding high-impact publications and citations) may preclude these important developments.

Another limitation is the long computing time required for the analyses. We were able to increase computational efficiency by running the model on multiple cores in parallel, but the processing time became particularly cumbersome if an analysis needed to be repeated. In the future, cloud-based computing could be used to obtain more consistent and faster model run times.

There were also instances of unusually long run times whereby the first two Markov chain Monte Carlo (MCMC) chains performed as expected, but subsequent chains never finished processing. Some MCMC convergence issues were resolved by reducing the parameter fitting period (e.g., truncating the beginning of the time series). One study reported that EpiNow2 estimates are more reliable when case numbers at each time step are large and there are at least 14 timepoints without zeroes(Binny et al., 2022). Large daily fluctuations and limited case counts could substantially affect model estimates, which should be interpreted with caution.

Another limitation is that the method we used does not account for under-ascertainment, which occurs when not all infections are diagnosed and reported as cases of the disease to the surveillance system. The under-ascertainment rate is needed to understand the true burden of disease caused by the outbreak; however, current estimates for the U.S. mpox outbreak are lacking. Indirect evidence from a recent modeling study(Clay et al., 2023) suggests that 65% of mpox infections were diagnosed and reported in Washington, D.C. However, the model was not designed to measure the under-ascertainment rate(Patrick Clay, personal communication, March 10, 2023), and this quantity should be assessed by other methods (e.g., models specifically designed to assess under-ascertainment, serological surveys, and community-based surveys).

Strategies for Forecasting the Next Outbreak

For the next outbreak, it is important to develop better strategies for regularly capturing and storing snapshots of surveillance data which remain easily accessible for systematic analysis. Most analyses of disease surveillance data are performed only after the data have undergone a rigorous and routine reconciliation and closeout process by data submitters with further validation by CDC surveillance epidemiologists, but timely outbreak response decision support does not allow for such processes. Instead, jurisdictions are asked to submit available case data, albeit often incomplete data, in a timely manner and submit additional data or corrections to data entry errors as time and resources permit. While some snapshots of the surveillance data over time were saved (by exporting data on a particular day and saving a csv file locally), we do not have a complete history of the data, especially around key

points in the outbreak, such as the peak. A complete history would help to understand key delay distributions and other quirks (e.g., backfilling and revision of reference dates) involved in the data-generating process. Understanding the data generating process is crucial for the improvement of methods and tools for nowcasting/forecasting.

Ensemble models have been used for a variety of infectious disease outbreaks, such as COVID-19(Cramer et al., 2022; Dean et al., 2020), Zika(Oidtman et al., 2021), influenza(Reich et al., 2019), and Ebola(Viboud et al., 2018). Ensembles combine predictions from several models that use different methodology and sometimes input data. Because some models overpredict, while others underpredict, ensemble models often outperform individual models over time. We were only able to work on one model for most of the outbreak, given the small size of our team. In the future, we may consider using at least two simpler models and comparing them.

One potentially useful addition to EpiNow2 and other currently available tools for nowcasting/forecasting outbreaks would be flexibility in handling dates. We frequently encountered missing dates for cases in the mpox surveillance data. Ideally, a method or tool would be able to keep track of multiple dates for a case and estimate missing dates based on the full distribution of dates across all cases. Epinowcast is a new hierarchical nowcasting package that enables more flexibility in adjusting for truncated data(Abbott et al., 2021b). Novel nowcasting approaches use hierarchical generalized additive models, which can provide even more flexibility to modify the model in real-time to the evolving data environment(Overton et al., 2023). Another improvement would be to reduce the time required to run the analyses. Rather than focusing on the efficiency of the MCMC algorithm, computation time could be reduced if the model only needed to be run on the new data. Finally, the imputed time series of cases by symptom onset date would be a useful data visualization output that is not currently available in EpiNow2. As described above, defining the date field for the presentation of epidemic curves was a challenge in the mpox outbreak and having an imputed symptom onset date for each case would have been useful for comparison purposes.

CFA played an important advisory role in our nowcasting/forecasting efforts. In the future, CFA plans to create new tools for outbreak analysis and modeling. CFA could also serve as a link between CDC modelers and jurisdictions with modeling capacity to share experiences and code. Technical Reports represent a new way for CDC to share timely information with the federal government, state and local leaders, and scientists in academia and industry. Technical Reports have been well received within and outside CDC(Daskalakis, 2022; Fox, 2022; Rivers, 2022) and their publication aligns with CDC's current restructuring efforts aimed at making the agency more response ready, including sharing science and data faster(U.S. Centers for Disease Control and Prevention, 2022b).

Conclusion

Real-time estimation of R_t as well as nowcasting/forecasting is important for determining the extent to which current public health measures are effective and/or need to be modified. The quality and timeliness of reported data pose challenges to these analyses. Ease of use, model computing time, and ability to handle multiple dates are priorities for consideration in the development of future nowcasting/forecasting tools. A naïve model may be superior to a complex one, such as EpiNow2, during the early phase of an outbreak when data scarcity causes R_t to be largely unconstrained, especially once reporting delays are considered. Future outbreak response activities could be enhanced through inclusion of clear and consistent communication about modeling outputs as well as close collaboration between modeling and informatics/data teams.

Tables

Table 1. Evaluation of short-term (one-week-ahead) forecasts of reported mpox cases during the 2022 U.S. outbreak. For WIS, bold indicates where one model performed better than the other.

		EpiNow2				**Bayesian GLM with negative binomial**			
Forecast date	**Outbreak phase**	**90% PI coverage**	**50% PI coverage**	**WIS**	**MAE**	**90% PI coverage**	**50% PI coverage**	**WIS**	**MAE**
Monday, June 13	Early	1	0.71	33.8	4.3	0.86	0.57	**6.4**	5.1
Monday, June 27	Early	0.86	0	34.0	29.0	0.86	0.14	**24.7**	26.9
Tuesday, July 5	Exponential growth	0.86	0.71	**35.6**	32.4	1	0.57	38.5	34.1
Wednesday, July 27	Peak	1	0.57	**137.0**	107.9	1	0.57	184.3	94.4
Tuesday, September 6	Decline	0.86	0.71	**107.6**	84.3	0.86	0.43	117.1	86.8
Monday, September 19	Decline	0.86	0.43	**61.2**	52.4	0.71	0.43	81.8	65.9
Tuesday, October 11	Decline	0.43	0.29	**30.6**	33.4	0.57	0.14	38.3	41.9
Monday, December 5	Decline	1	0.71	**4.3**	3.1	1	0.71	6.4	4.6

PI: prediction interval; WIS: weighted interval score; MAE: mean absolute error; GLM: generalized linear model

Figures

Figure 1. Effective reproduction number estimates for the U.S. 2022 mpox outbreak intended for a technical/scientific audience. The top panel shows estimates of cases by date of report with actual cases shown by gray bars. The bottom panel shows estimates of the effective reproduction number by date. In all panels, shaded regions reflect 90%, 50%, and 20% credible intervals in order from lightest to darkest. Green shows estimates, red shows estimates based on partial data, and purple shows forecasts. Report date is determined by a hierarchy across the different data streams where priority is given to diagnosis date, orthopoxvirus test date, orthopoxvirus test confirmation date, case investigation start date, orthopoxvirus sample collection date, date of call to CDC call center, report date (to public health department, county, or state), date CDC announced case, and the date the case was entered into DCIPHER, in that order.

Figure 2. Effective reproduction number (R_t) estimates for the U.S. 2022 mpox outbreak intended for the general public. The graph shows the R_t estimation over time based on complete data (gray) or partial data (blue). The most recent data are considered incomplete due to delays in reporting mpox cases. As a result, there is more uncertainty associated with the most recent R_t estimates. $R_t > 1$ means the epidemic is growing. $R_t < 1$ means the epidemic is shrinking. Shading represents the 50% and 90% credible intervals (uncertainty in the estimates)

Figure 3. Effective reproduction number estimates for the 2022 mpox outbreak in four U.S. Census regions. The left panels show estimates of cases by date of report with actual cases shown by gray bars. The right panels show estimates of the effective reproduction number by date. In all panels, shaded regions reflect 90%, 50%, and 20% credible intervals in order from lightest to darkest. Green shows estimates, red shows estimates based on partial data, and purple shows forecasts. Report date is determined by a hierarchy across the different data streams where priority is given to diagnosis date, orthopoxvirus test date, orthopoxvirus test confirmation date, case investigation start date, orthopoxvirus sample collection date, date of call to CDC call center, report date (to public health department, county, or state), date CDC announced case, and the date the case was entered into DCIPHER, in that order.

Figure 4. Effective reproduction number estimates of the 2022 mpox outbreak for the six jurisdictions in the U.S. with the highest case counts. The left panels show estimates of cases by date of report with actual cases shown by gray bars. The right panels show estimates of the effective reproduction number by date. In all panels, shaded regions reflect 90%, 50%, and 20% credible intervals in order from lightest to darkest. Green shows estimates, red shows estimates based on partial data, and purple shows forecasts. Report date is determined by a hierarchy across the different data streams where priority is given to diagnosis date, orthopoxvirus test date, orthopoxvirus test confirmation date, case investigation start date, orthopoxvirus sample collection date, date of call to CDC call center, report date (to public health department, county, or state), date CDC announced case, and the date the case was entered into DCIPHER, in that order.

Data availability

Data and code to run the nowcasts/forecasts and perform model validation will be available on GitHub following publication in a peer-reviewed journal.

Acknowledgements

We thank all public health professionals involved in reporting mpox cases to CDC. We also acknowledge the Mpox Response Data Analytics and Visualization Task Force Informatics Team for data management and support.

Declaration of interest

The authors report no conflicts of interest.

Disclaimer

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention, U.S. Department of Health and Human Services.

Ethics statement

This activity was reviewed by CDC and was conducted consistent with applicable federal law and CDC policy (45 C.F.R. part 46, 21 C.F.R. part 56; 42 U.S.C. Sect. 241(d); 5 U.S.C. Sect. 552a; 44 U.S.C. Sect. 3501 et seq).

Funding

No specific funding was obtained for this work.

Author contributions

Conceptualization (KC, YN), data curation (KC, JA), formal analysis (KC, ZJM), investigation (KC, ZJM), methodology (KC, ZJM, JA), project administration (KC), software (KC, ZJM, NBM), supervision (JA, YN, IHS), validation (KC, ZJM), visualization (KC, ZJM), writing – original draft preparation (KC, ZJM), writing – review & editing (KC, ZJM, NBM, JA, YN, IHS).

References

Abbott, S., Hellewell, J., Sherratt, K., Gostic, K., Hickson, J., Badr, H., DeWitt, M., EpiForecasts, Funk, S., 2021a. EpiNow2: Estimate real-time case counts and time-varying epidemiological parameters, 1.3.2 ed. EpiForecasts.

Abbott, S., Hellewell, J., Thompson, R., Sherratt, K., Gibbs, H., Bosse, N., Munday, J., Meakin, S., Doughty, E., Chun, J., 2020. Estimating the time-varying reproduction number of SARS-CoV-2 using national and subnational case counts [version 2; peer review: 1 approved, 1 approved with reservations]. Wellcome Open Research 5, 112.

Abbott, S., Lison, A., Funk, S., 2021b. epinowcast: Flexible hierarchical nowcasting. Zenodo. doi: 10.5281/zenodo.5637165.

Arslan, S., Ozdemir, M., Ucar, A., 2021. Nowcasting and Forecasting the Spread of COVID-19 and Healthcare Demand in Turkey, a Modeling Study. Front Public Health 8, 575145.

Beer, E., Rao, V., 2019. A systematic review of the epidemiology of human monkeypox outbreaks and implications for outbreak strategy. PLoS Negl Trop Dis 13, e0007791.

Biggerstaff, M., Slayton, R., Johansson, M., Butler, J., 2021. Improving Pandemic Response: Employing Mathematical Modeling to Confront Coronavirus Disease 2019. Clinical Infectious Diseases 74, 913-917.

Binny, R., Lustig, A., Hendy, S., Maclaren, O., Ridings, K., Vattiato, G., Plank, M., 2022. Real-time estimation of the effective reproduction number of SARS-CoV-2 in Aotearoa New Zealand. PeerJ 10, e14119.

Birrell, P., Blake, J., van Leeuwen, E., Gent, N., De Angelis, D., 2021. Real-time nowcasting and forecasting of COVID-19 dynamics in England: the first wave. Philos Trans R Soc Lond B Biol Sci 376, 20200279.

Buchwald, A., Adams, J., Bortz, D., Carlton, E., 2020. Infectious Disease Transmission Models to Predict, Evaluate, and Improve Understanding of COVID-19 Trajectory and Interventions. Annals of the American Thoracic Society 17, 1204-1206.

California Department of Public Health, 2022. Monkeypox Data in California

Chakraborty, T., Ghosh, I., Mahajan, T., Arora, T., 2022. Nowcasting of COVID-19 Confirmed Cases: Foundations, Trends, and Challenges, in: Azar, A., Hassanien, A. (Eds.), Modeling, Control and Drug Development for COVID-19 Outbreak Prevention. Springer, Cham. doi: https://doi.org/10.1007/978-3-030-72834-2_29.

Clay, P., Asher, J., Carnes, N., Copen, C., Delaney, K., Payne, D., Pollack, E., Mermin, J., Nakazawa, Y., Still, W., 2023. Modelling the impact of vaccination and sexual behavior change on reported cases of mpox in Washington D.C. medRxiv.

Cori, A., Cauchemez, S., Ferguson, N., Fraser, C., Dahlqwist, E., Demarsh, P., Jombart, T., Kamvar, Z., Lessler, J., Li, S., 2021. EpiEstim: Estimate Time Varying Reproduction Numbers from Epidemic Curves, 2.2-4 ed, CRAN.

Cori, A., Ferguson, N., Fraser, C., Cauchemez, S., 2013. A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics. American Journal of Epidemiology 178, 1505-1512.

Cramer, E., Ray, E., Lopez, V., Bracher, J., Brennen, A., Rivadeneira, A., Gerding, A., Gneiting, T., House, K., Huang, Y., 2022. Evaluation of individual and ensemble probabilistic forecasts of COVID-19 mortality in the United States. PNAS 119, e2113561119.

Daskalakis, D., 2022. You know that you are a #publichealth nerd when you wake up on a gloomy Saturday to read @CDCgov

Twitter, p. [Tweet].

De Nicola, G., Schneble, M., Kauermann, G., Berger, U., 2022. Regional now- and forecasting for data reported with delay: toward surveillance of COVID-19 infections. Adv Stat Anal 106, 407-426.

Dean, N., Pastore Y Piontti, A., Madewell, Z., Cummings, D., Hitchings, M., Joshi, K., Kahn, R., Vespignani, A., Halloran, M., Longini, I., 2020. Ensemble forecast modeling for the design of COVID-19 vaccine efficacy trials. Vaccine 38, 7213-7216.

Fenner, F., Henderson, D., Arita, I., Jezek, Z., Ladnyi, I., 1988. Smallpox and its Eradication. World Health Organization, Geneva.

Fox, S., 2022. These monkeypox technical reports from the CDC are [three fire emojis]. Twitter, p. [Tweet].

Jombart, T., Cori, A., Nouvellet, P., Skarp, J., Kamvar, Z., Taylor, T., 2020. earlyR: Estimation of Transmissibility in the Early Stages of a Disease Outbreak, 0.0.5 ed. CRAN.

Lloyd-Smith, J., Funk, S., McLean, A., Riley, S., Wood, J., 2015. Nine challenges in modelling the emergence of novel pathogens. Epidemics 10, 35-39.

Madewell, Z., Charniga, K., Masters, N., Asher, J., Fahrenwald, L., Still, W., Chen, J., Kipperman, N., Bui, D., Shea, M., 2023. Serial interval and incubation period estimates of monkeypox virus infection in 12 U.S. jurisdictions, May – August 2022. Emerg Infect Dis 29, 818-821.

Madewell, Z., Dean, N., Berlin, J., Coplan, P., Davis, K., Struchiner, C., Halloran, M., 2021a. Challenges of evaluating and modelling vaccination in emerging infectious diseases. Epidemics 37, 100506.

Madewell, Z., Pastore Y Piontti, A., Zhang, Q., Burton, N., Yang, Y., Longini, I., Halloran, M., Vespignani, A., Dean, N., 2021b. Using simulated infectious disease outbreaks to inform site selection and sample size for individually randomized vaccine trials during an ongoing epidemic. Clinical Trials 18, 630-638.

Nash, R., Nouvellet, P., Cori, A., 2022. Real-time estimation of the epidemic reproduction number: Scoping review of the applications and challenges. PLOS Digit Health 1, e0000052.

New York City Department of Health and Mental Hygiene, 2022. 2022 Monkeypox Surveillance Data.

Oidtman, R., Omodei, E., Kraemer, M., Castañeda-Orjuela, C., Cruz-Rivera, E., Misnaza-Castrillón, S., Cifuentes, M., Rincon, L., Cañon, V., 2021. Trade-offs between individual and ensemble forecasts of an emerging infectious disease. Nature Communications 12, 5379.

Overton, C., Abbott, S., Christie, R., Cumming, F., Day, J., Jones, O., Paton, R., Turner, C., Ward, T., 2023. Nowcasting the 2022 mpox outbreak in England. arXiv.

Reich, N., McGowan, C., Yamana, T., Tushar, A., Ray, E., Osthus, D., Kandula, S., Brooks, L., Crawford-Crudell, W., 2019. Accuracy of real-time multi-model ensemble forecasts for seasonal influenza in the U.S. PLOS Computational Biology 15, e1007486.

Rivers, C., 2022. Changing trends in monkeypox epidemiology. Substack.

Rohatgi, A., 2021. WebPlotDigitizer, Vers. 4.5 ed, Pacifica, California, USA.

Spreco, A., Jöud, A., Eriksson, O., Soltesz, K., Källström, R., Dahlström, Ö., Eriksson, H., Ekberg, J., Jonson, C., 2022. Nowcasting (Short-Term Forecasting) of COVID-19 Hospitalizations Using Syndromic Healthcare Data, Sweden, 2020. Emerg Infect Dis. 28, 564-571.

Thompson, J., McClure, R., Scott, N., Hellard, M., Abeysuriya, R., Vidanaarachchi, R., Thwaites, J., Lazarus, J., Lavis, J., Michie, S., 2022. A framework for considering the utility of models when facing tough decisions in public health: a guideline for policy-makers. Health Research Policy and Systems 20, 107.

U.S. Centers for Disease Control and Prevention, 2022a. 2022 Monkeypox Outbreak Global Map.

U.S. Centers for Disease Control and Prevention, 2022b. CDC Moving Forward Summary Report.

U.S. Centers for Disease Control and Prevention, 2022c. Monkeypox Technical Reports.

U.S. Centers for Disease Control and Prevention, 2022d. Outbreak Reproduction Number Estimates.

U.S. Centers for Disease Control and Prevention, 2022e. Signs and Symptoms.

U.S. Centers for Disease Control and Prevention, 2022f. What is the Data Modernization Initiative?

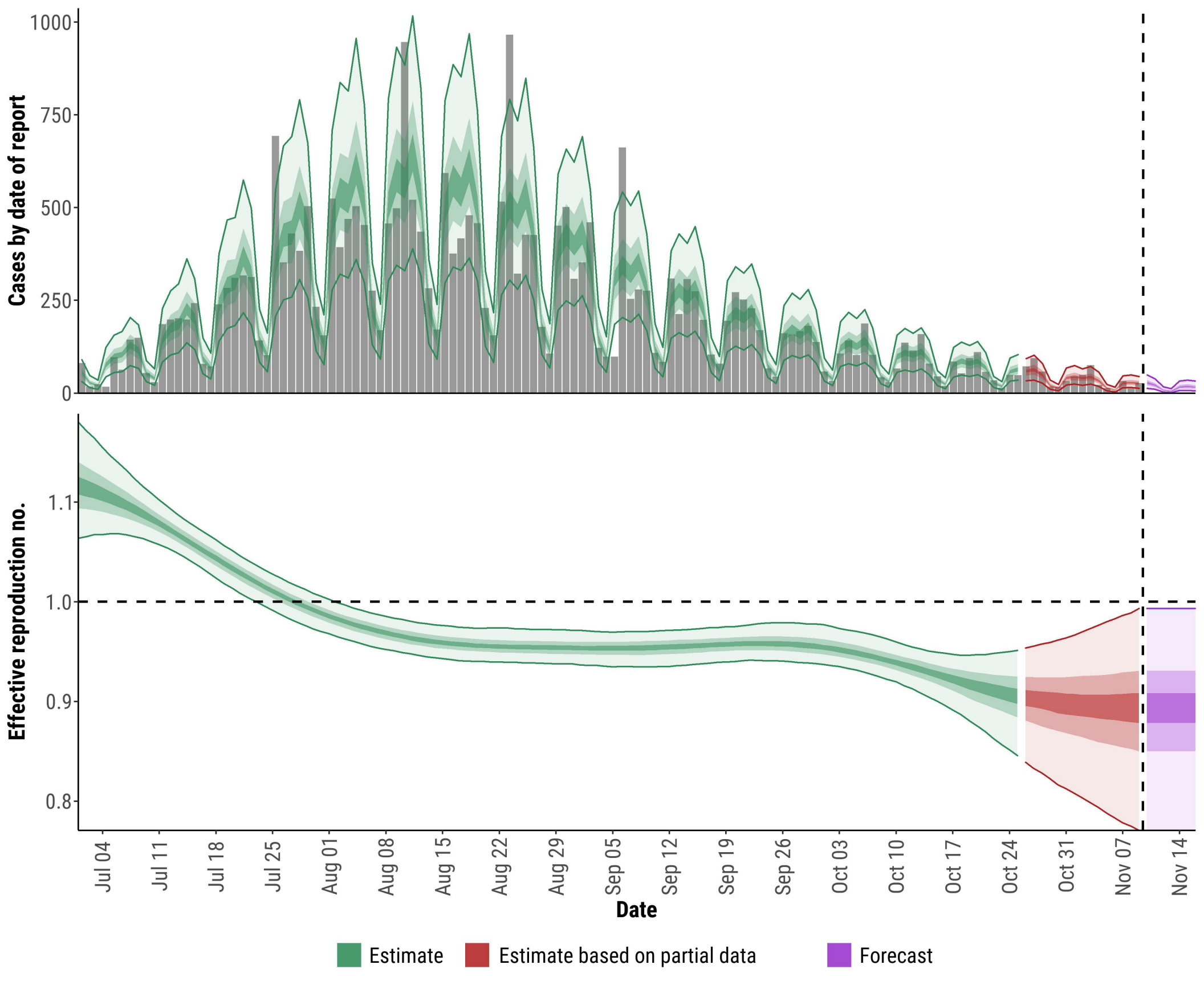
U.S. Department of Health and Human Services, 2020. Privacy Impact Assessment.

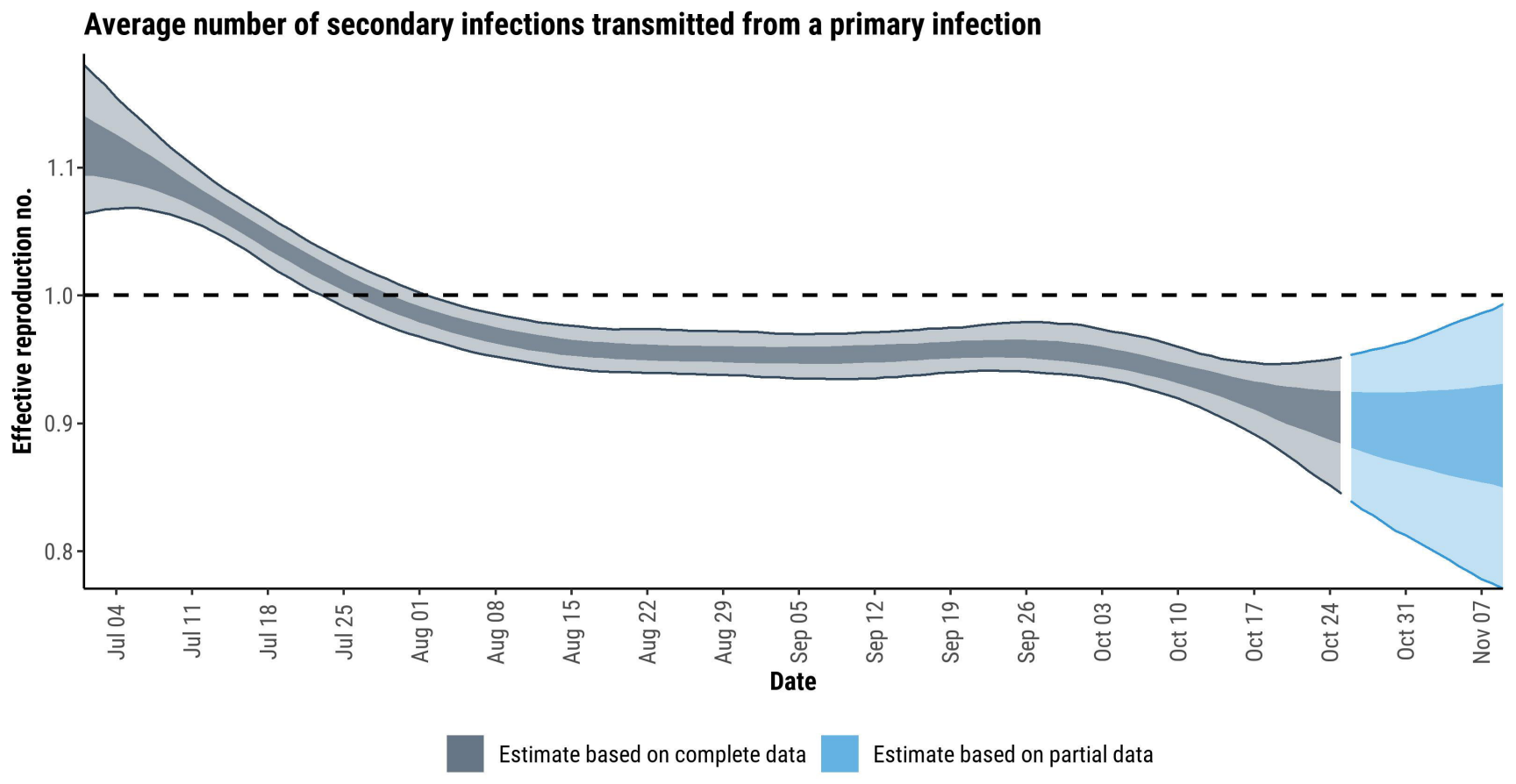
UK Health Security Agency, 2022. Investigation into monkeypox outbreak in England: technical briefing 1.

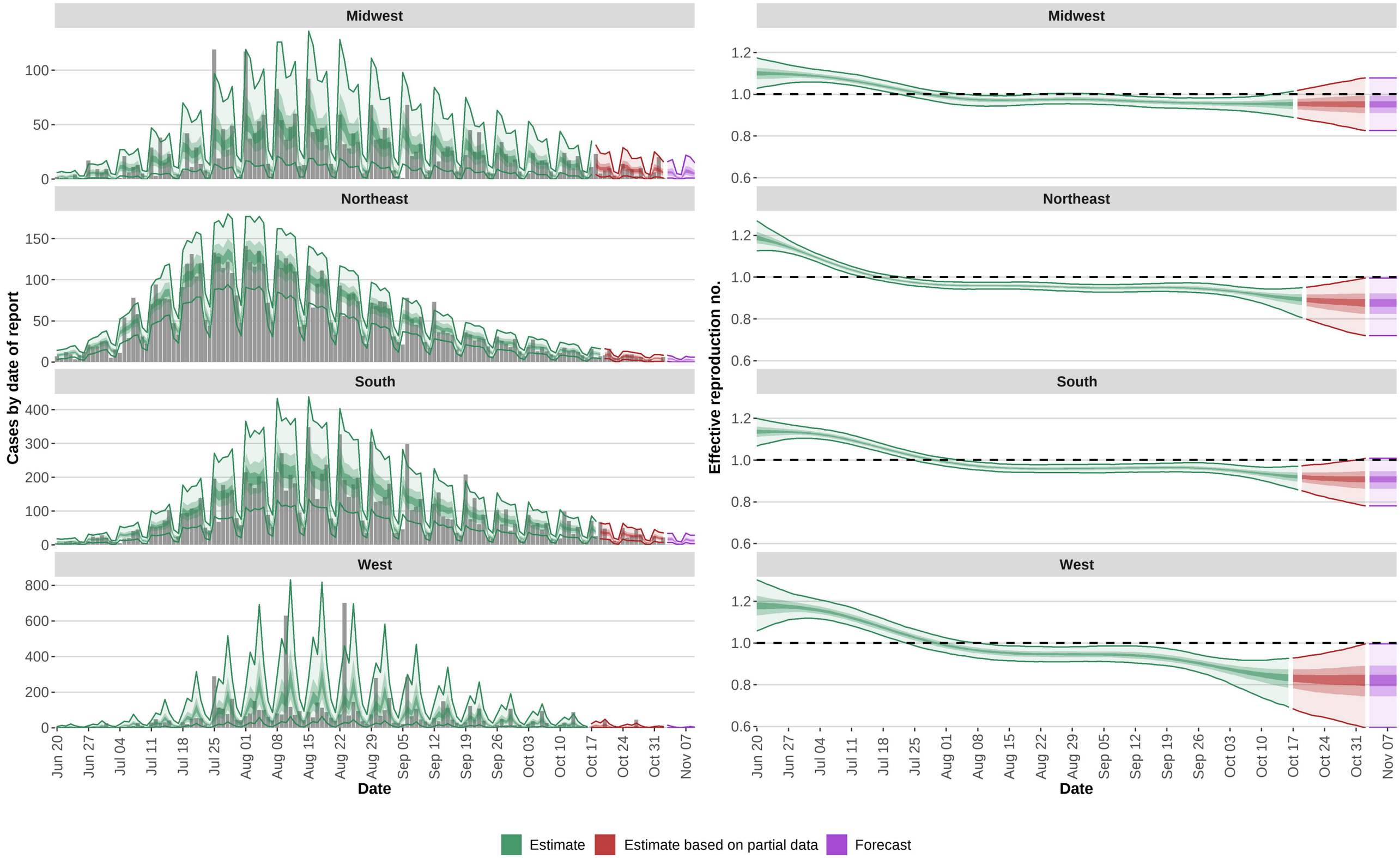
Viboud, C., Sun, K., Gaffey, R., Ajelli, M., Fumanelli, L., Merler, S., Zhang, Q., Chowell, G., Simonsen, L., Vespignani, A., 2018. The RAPIDD ebola forecasting challenge: Synthesis and lessons learnt. Epidemics 22, 13-21.

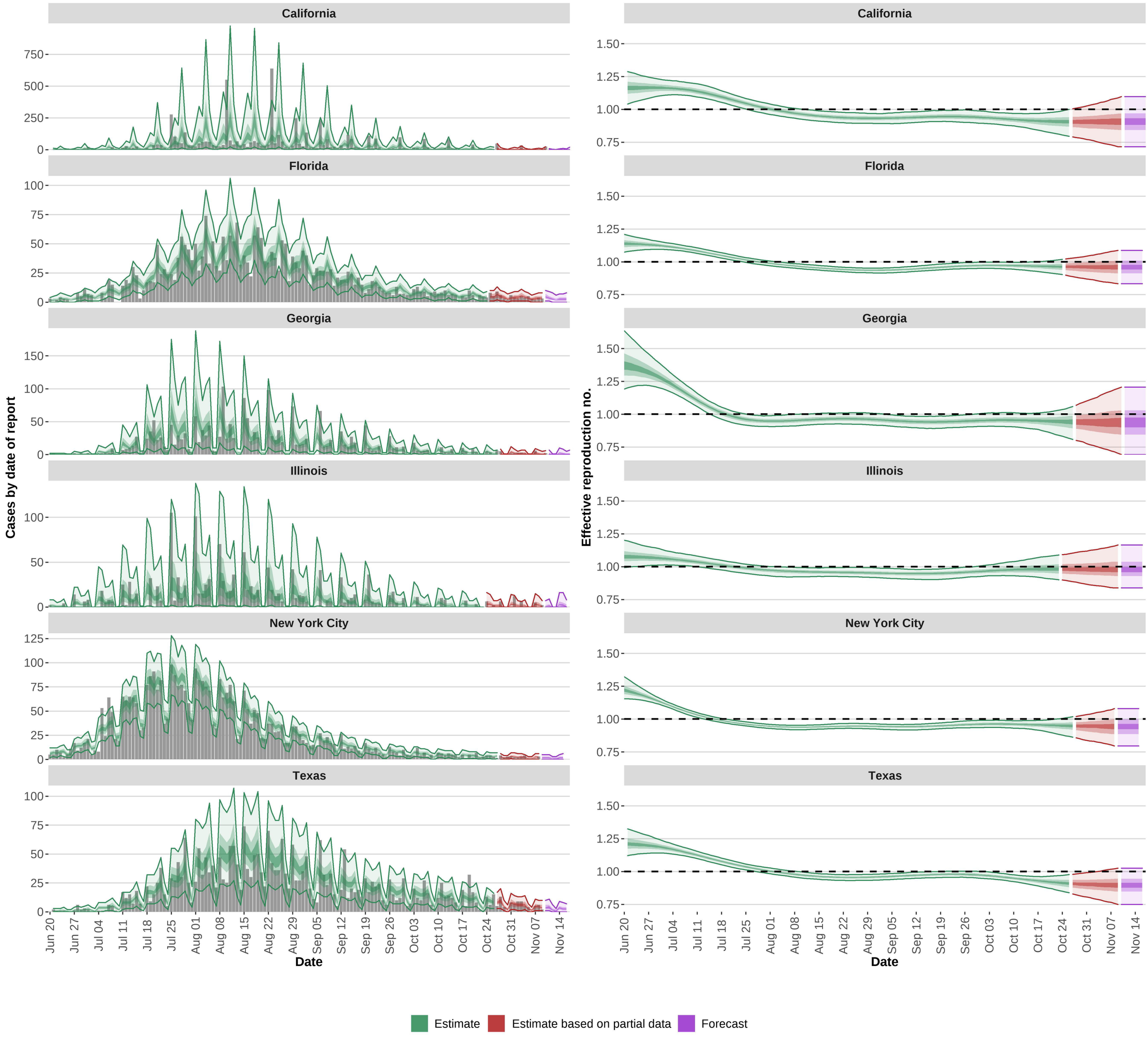
WHO/ECDC, 2022. Interim advice for public health authorities on summer events during the monkeypox outbreak in Europe, 2022.

World Health Organization, 2022. WHO Director-General declares the ongoing monkeypox outbreak a Public Health Emergency of International Concern.


Cases by date of report
1000
750
500
250
0
Effective reproduction no.
1.1
1.0
0.9
0.8
Jul 04
Jul 11
Jul 18
Jul 25
Aug 01
Aug 08
Aug 15
Aug 22
Aug 29
Sep 05
Sep 12
Sep 19
Sep 26
Oct 03
Oct 10
Oct 17
Oct 24
Oct 31
Nov 07
Nov 14
Date
Estimate
Estimate based on partial data
Forecast


Average number of secondary infections transmitted from a primary infection
Effective reproduction no.
1.1
1.0
0.9
0.8
Jul 04
Jul 11
Jul 18
Jul 25
Aug 01
Aug 08
Aug 15
Aug 22
Aug 29
Sep 05
Sep 12
Sep 19
Sep 26
Oct 03
Oct 10
Oct 17
Oct 24
Oct 31
Nov 07
Date
Estimate based on complete data
Estimate based on partial data


Midwest
Northeast
South
West
Cases by date of report
Effective reproduction no.
Date
Jun 20
Jun 27
Jul 04
Jul 11
Jul 18
Jul 25
Aug 01
Aug 08
Aug 15
Aug 22
Aug 29
Sep 05
Sep 12
Sep 19
Sep 26
Oct 03
Oct 10
Oct 17
Oct 24
Oct 31
Nov 07
Estimate
Estimate based on partial data
Forecast


California
Florida
Georgia
Illinois
New York City
Texas
Cases by date of report
Effective reproduction no.
Date
Estimate
Estimate based on partial data
Forecast