

2024 mpox outbreak: common analytics tasks and available R tools

James Azam, Hugo Gruson and Adam Kucharski

Published July 4, 2024

Citation

Azam, J., Gruson, H., & Kucharski, A. (2024, July 4). 2024 mpox outbreak: common analytics tasks and available R tools. *Epiverse-TRACE Developer Space*. <https://doi.org/10.59350/z78kb-qrz59>

Keywords

Mpox, Outbreak, Outbreak Analytics, DOI

Copyright

Copyright © James Azam et al. 2024. Distributed under the terms of the [Creative Commons Attribution 4.0 International License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Epiverse-TRACE developer space

There are ongoing outbreaks of mpox globally. The Democratic Republic of Congo (DRC) is so far the worst hit with a total of 7,851 cases and 384 deaths reported between January 1 and May 26, 2024 ¹. Before 2022, there were few reports of sustained mpox transmission globally. However, during the following year (Jan 1, 2022, and Jan 29, 2023), 110 countries in all six WHO Regions had reported a total of 85,473 confirmed cases and 89 deaths between them (Laurenson-Schafer et al. 2023).

Mpox is transmitted through respiratory droplets and direct contact with infected persons. The disease is characterized by fever, cough, and a rash, with the mean incubation period estimated to be about 7.8 days (Ward et al. 2022). Infected individuals may experience severe symptoms leading to hospitalisation or death. There are two genetic clades: clade I and clade II, which also has subclades IIa and IIb (Laurenson-Schafer et al. 2023).

Several analyses of the potential impact of outbreaks at country level have already emerged in 2024. The US CDC, for example, has analysed the potential size of outbreaks resulting from transmission within and between households ² and the risk of Clade 1 mpox outbreaks among some key populations associated with key transmission routes ³. Another group of researchers have estimated the transmissibility of mpox in the DRC from more recent (2010 - 2019) surveillance data to update existing estimates, which are based on old data (Charniga, McCollum, et al. 2024). However, tackling ongoing outbreaks around the world will require a coordinated response from the global health community.

The Epiverse-TRACE team is developing a set of analytical tools that could help support decision-makers during outbreaks. This post provides an overview of the tasks that such tools can be applied to in the context of the ongoing mpox outbreaks.

Common outbreak analytics tasks

Outbreak analytics in the context of the ongoing mpox outbreak involves several tasks that can be handled by existing and emerging R tools. Some of the tasks include estimating the transmission potential, forecasting infection dynamics, estimating severity, and assessing the impact of interventions.

Here, we briefly describe some common tasks, data required, and the ready R tools/packages developed by the Epiverse-TRACE team and the wider community.

Cleaning and validating data

Data cleaning is often the first task in outbreak analytics. This usually involves identifying and correcting errors in the data, standardizing the format of key variables, and ensuring that the data is in a format that is fit for analysis. Data validation is also important to ensure that the data is accurate.

`{cleanepi}` is useful for cleaning individual-level datasets, and `{linelist}` can be used to tag and validate key variables in datasets that might change over time. The `{numberize}` package

Epiverse-TRACE developer space

can also be used to convert numbers written as text. It currently has functionality for English, Spanish, and French.

Estimating transmission potential

A key initial question during emerging outbreaks is the transmission potential of the disease. This is typically quantified using parameters such as: the basic reproduction number, R_0 ; the time-varying reproduction number, R_t ; and k , which captures individual heterogeneity in transmission (i.e. “superspreading” potential). These quantities are useful to assess the potential for further spread of the disease and the impact of interventions.

Population-level transmissibility (R_0 and R_t)

The basic reproduction number, R_0 , is the average number of secondary cases produced by a single infected individual in a completely susceptible population. The time-varying reproduction number, R_t , on the other hand, is the average number of secondary cases produced by a single infected individual at time t in a partially susceptible population. R_t is a more useful quantity during an outbreak as it accounts for the impact of interventions and changes in population immunity.

If data is available on the daily number of reported cases, [{EpiNow2}](#) and [{EpiEstim}](#) can be used to estimate R_t . These packages require data on the time scale of transmission (i.e. the generation time, or the serial interval, which is [commonly used as a proxy](#) for this). While [{EpiEstim}](#) focuses on retrospective estimation of R_t , [{EpiNow2}](#) is designed for both retrospective and real-time estimation.

In estimating R_t , one practical consideration is the impact of various delays (biological and reporting) on the estimates (Charniga, Park, et al. 2024; Park et al. 2024; Katelyn M. Gostic 2020). [{EpiNow2}](#) adjusts for these delays in various ways. For example, it accounts for the symptom onset and reporting delays by taking the incubation period and reporting delay as inputs. Moreover, [{EpiNow2}](#) can estimate the reporting delay from the data if data on incidence by date of onset and report are available.

Furthermore, dedicated packages have emerged for estimating epidemiological delays from data using best practices. [{epidist}](#) offers the ability to estimate delay distributions, accounting for issues such as truncation (i.e., not all disease outcomes will yet be known in real-time).

If delay data are not available, published estimates of the incubation period and serial interval can be used. The [{epiparameter}](#) package collates a database of epidemiological distributions from the literature and provides functions for interacting with the database. You can view the [database](#) for currently available parameters (more entries are planned). Additionally, if only summary statistics are available (e.g. range and median), [{epiparameter}](#) can be used to extract the distribution parameters.

Epiverse-TRACE developer space

Individual-level transmissibility (superspreading)

The individual-level transmission heterogeneity (superspreading), often denoted as k , is an important measure for tailoring interventions at the individual level.

If we have data on the distribution of sizes of transmission clusters, the `{epichains}` package provides functions to set up the likelihood function to estimate R_0 and k . The user inputs the negative binomial offspring, which assumes individuals exhibit heterogeneity in transmission. The parameters of the negative offspring distribution can then be estimated using existing maximum likelihood or bayesian frameworks.

Furthermore, if we have individual-level transmission chain data, the `{superspreading}` package can be used to estimate R_0 and k from the offspring distribution. This package also provides functions to estimate the probability that an outbreak will not go extinct in its early stages because of randomness in transmission (e.g. if the primary spillover case(s) does not infect others).

If we have data on sexual contacts and the secondary attack rate, then we can also use `{superspreading}` to [calculate \$R_0\$ accounting for network effects](#).

Forecasting and nowcasting infection dynamics

Forecasting and nowcasting of infections are crucial for planning and resource allocation during an outbreak. Forecasting is the prediction of future cases, deaths, or other outcomes, while nowcasting is the prediction of the current outbreak situation. These predictions can help public health authorities to anticipate the trajectory of the outbreak and to implement timely interventions.

`{EpiNow2}` and `{epinowcast}` provide functions to forecast and nowcast the number of cases. The data required for `{EpiNow2}` has already been described in the previous section. The `{epinowcast}` package similarly requires data on the number of cases reported per date. `{epinowcast}` does not currently support forecasting but there are plans to add this functionality in future versions.

Estimating disease severity

The case fatality risk (CFR) is often used to assess the severity of a disease. CFR here refers to the proportion of deaths among confirmed cases.

With incidence data on the number of cases reported and the number of deaths reported, the `{cfr}` package can be used to estimate the case fatality rate and its uncertainty. Importantly, it accounts for the delay between the onset of symptoms and death, which is crucial for accurate estimation of the case fatality rate.

Here again, `{EpiNow2}` can be used to estimate the time-varying case fatality ratio using the same data as for the reproduction number. `{EpiNow2}` can estimate other severity metrics,

Epiverse-TRACE developer space

such as the case hospitalisation ratio, given data on cases and hospitalisations, and the hospitalisation fatality ratio, if data on hospitalisations and associated deaths are available.

Assessing the impact of interventions

mpox can be mitigated with behaviour change, treatment, and vaccination. Here, a few tools are available to assess the impact of intervention scenarios.

`{epidemics}` provides ready compartmental models to estimate the impact of vaccination and non-pharmaceutical interventions like behaviour change, which can conceptually be modelled as a reduction in the transmission rate through changes in the population contact structure.

If we want to explore population-level outbreak dynamics, `{epidemics}` allows for stratifying the population into arbitrary groups, specifying the contact structure between these groups, and rates of interventions. The data required to run these models include: population structure, contact structure, and timing and magnitude of interventions. Data on social contact matrices can be obtained from the `{socialmixr}` package.

Summary

In this post, we have outlined common outbreak analytics tasks relevant to the mpox outbreak, the data required, and R packages/tools that are currently available to facilitate these tasks. The tools described here are being developed by the Epiverse-TRACE team and the wider community, with the aim of ensuring high standards of research software development, and validation from end users, including epidemiologists, clinicians, and policy makers. The tools are designed to be user-friendly and well integrated, enabling one analysis task to easily feed into another. We [would therefore be keen to hear](#) from other groups interested in potentially collaborating or contributing on this growing ecosystem of tools.

Thanks to [Karim Mane](#) and [Chris Hartgerink](#) for their valuable comments on earlier drafts of this post.

References

- Charniga, Kelly, Andrea M McCollum, Christine M Hughes, Benjamin Monroe, Joelle Kabamba, Robert Shongo Lushima, Toutou Likafi, et al. 2024. "Updating Reproduction Number Estimates for Mpox in the Democratic Republic of Congo Using Surveillance Data." *The American Journal of Tropical Medicine and Hygiene* 110 (3): 561. <https://doi.org/https://doi.org/10.4269/ajtmh.23-0215>.
- Charniga, Kelly, Sang Woo Park, Andrei R Akhmetzhanov, Anne Cori, Jonathan Dushoff, Sebastian Funk, Katelyn M Gostic, et al. 2024. "Best Practices for Estimating and Reporting Epidemiological Delay Distributions of Infectious Diseases Using Public Health Surveillance and Healthcare Data." *arXiv Preprint arXiv:2405.08841*. <https://doi.org/https://doi.org/10.48550/arXiv.2405.08841>.

Epiverse-TRACE developer space

- Katelyn M. Gostic, Edward B. Baskerville, Lauren McGough. 2020. "Practical Considerations for Measuring the Effective Reproductive Number, R_t ." *PLoS Computational Biology* 16 (12): e1008409. <https://doi.org/https://doi.org/10.1371/journal.pcbi.1008409>.
- Laurenson-Schafer, Henry, Nikola Sklenovská, Ana Hoxha, Steven M Kerr, Patricia Ndumbi, Julia Fitzner, Maria Almiron, et al. 2023. "Description of the First Global Outbreak of Mpox: An Analysis of Global Surveillance Data." *The Lancet Global Health* 11 (7): e1012–23. [https://doi.org/https://doi.org/10.1016/S2214-109X\(23\)00198-5](https://doi.org/https://doi.org/10.1016/S2214-109X(23)00198-5).
- Park, Sang Woo, Andrei R Akhmetzhanov, Kelly Charniga, Anne Cori, Nicholas G Davies, Jonathan Dushoff, Sebastian Funk, et al. 2024. "Estimating Epidemiological Delay Distributions for Infectious Diseases." *medRxiv*, 2024–01. <https://doi.org/https://doi.org/10.1101/2024.01.12.24301247>.
- Ward, Thomas, Rachel Christie, Robert S Paton, Fergus Cumming, and Christopher E Overton. 2022. "Transmission Dynamics of Monkeypox in the United Kingdom: Contact Tracing Study." *Bmj* 379. <https://doi.org/http://dx.doi.org/10.1136/bmj-2022-073153>.

Footnotes

1. [WHO Disease Outbreak News↔](#)
2. [Modeling Household Transmission of Clade I Mpox in the United States↔](#)
3. [Risk of Clade 1 Mpox Outbreaks Among Gay, Bisexual, and Other Men Who Have Sex With Men in the United States↔](#)

Reuse

CC BY 4.0

Citation

BibTeX citation:

```
@online{azam2024,  
  author = {Azam, James and Gruson, Hugo and Kucharski, Adam},  
  title = {2024 Mpox Outbreak: Common Analytics Tasks and Available {R}  
    Tools},  
  date = {2024-07-04},  
  url = {https://epiverse-trace.github.io/posts/mpox-preparedness/},  
  doi = {10.59350/z78kb-qrz59},  
  langid = {en}  
}
```

For attribution, please cite this work as:

Azam, James, Hugo Gruson, and Adam Kucharski. 2024. "2024 Mpox Outbreak: Common Analytics Tasks and Available R Tools." July 4, 2024. <https://doi.org/10.59350/z78kb-qrz59>.