



HAL
open science

Best practices for estimating and reporting epidemiological delay distributions of infectious diseases

Kelly Charniga, Sang Woo Park, Andrei R Akhmetzhanov, Anne Cori,
Jonathan Dushoff, Sebastian Funk, Katelyn M Gostic, Natalie M Linton,
Adrian Lison, Christopher E Overton, et al.

► To cite this version:

Kelly Charniga, Sang Woo Park, Andrei R Akhmetzhanov, Anne Cori, Jonathan Dushoff, et al.. Best practices for estimating and reporting epidemiological delay distributions of infectious diseases. 2024. hal-04572940v2

HAL Id: hal-04572940

<https://hal.science/hal-04572940v2>

Preprint submitted on 10 Oct 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

1 Best practices for estimating and reporting epidemiological delay distributions of infectious
2 diseases

3

4

5 Kelly Charniga¹, Sang Woo Park², Andrei R. Akhmetzhanov³, Anne Cori⁴, Jonathan Dushoff^{5,6,7},
6 Sebastian Funk^{8,9}, Katelyn M. Gostic¹⁰, Natalie M. Linton¹¹, Adrian Lison¹², Christopher E.
7 Overton^{13,14,15}, Juliet R.C. Pulliam¹⁰, Thomas Ward¹⁴, Simon Cauchemez¹, Sam Abbott^{8,9}

8

9

10 ¹ Mathematical Modelling of Infectious Diseases Unit, Institut Pasteur, Université Paris Cité,
11 CNRS UMR 2000, Paris, France

12 ² Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey,
13 USA

14 ³ College of Public Health, National Taiwan University, Taipei, Taiwan

15 ⁴ MRC Centre for Global Infectious Disease Analysis, School of Public Health, Imperial College
16 London, London, UK

17 ⁵ Departments of Mathematics & Statistics and Biology, McMaster University, Hamilton, Ontario,
18 Canada

19 ⁶ Department of Biology, McMaster University, Hamilton, Ontario, Canada

20 ⁷ M. G. DeGroot Institute for Infectious Disease Research, McMaster University, Hamilton,
21 Ontario, Canada

22 ⁸ Department of Infectious Disease Epidemiology and Dynamics, London School of Hygiene &
23 Tropical Medicine, London, UK

24 ⁹ Centre for Mathematical Modelling of Infectious Diseases, London School of Hygiene &
25 Tropical Medicine, London, UK

26 ¹⁰ Center for Forecasting and Outbreak Analytics, U.S. Centers for Disease Control and
27 Prevention, Atlanta, Georgia, USA

28 ¹¹ Graduate School of Medicine, Hokkaido University, Sapporo-shi, Hokkaido, Japan

29 ¹² Department of Biosystems Science and Engineering, ETH Zurich, Zurich, Switzerland

30 ¹³ Department of Mathematical Sciences, University of Liverpool, Liverpool, UK

31 ¹⁴ All Hazards Intelligence, Infectious Disease Modelling Team, Data Analytics and Surveillance,
32 UK Health Security Agency, UK

33 ¹⁵ Department of Mathematics, University of Manchester, Manchester, UK

34

35

36

37 Word count: 5,273

38

39 Contact information: Kelly Charniga, kelly.chniga@pasteur.fr

40

41

42

43 Keywords: epidemiological delays, natural history parameters, mathematical modeling, best
44 practices, checklist

45 **Abstract**

46

47 Epidemiological delays are key quantities that inform public health policy and clinical practice.

48 They are used as inputs for mathematical and statistical models, which in turn can guide control

49 strategies. In recent work, we found that censoring, right truncation, and dynamical bias were

50 rarely addressed correctly when estimating delays and that these biases were large enough to

51 have knock-on impacts across a large number of use cases. Here, we formulate a checklist of

52 best practices for estimating and reporting epidemiological delays. We also provide a flowchart

53 to guide practitioners based on their data. Our examples are focused on the incubation period

54 and serial interval due to their importance in outbreak response and modeling, but our

55 recommendations are applicable to other delays. The recommendations, which are based on

56 the literature and our experience estimating epidemiological delay distributions during outbreak

57 responses, can help improve the robustness and utility of reported estimates and provide

58 guidance for the evaluation of estimates for downstream use in transmission models or other

59 analyses.

60

61

62

63

64

65

66

67 **Introduction**

68

69 In recent years, an increasing number of real-time data streams have become available
70 for a variety of computational biology applications. For example, NEON [1] is a large-scale
71 observation platform for environmental data in the United States, and EarthRanger [2] is used
72 for wildlife conservation in several countries. These new data streams may improve the ability to
73 perform real-time analyses, such as monitoring biodiversity [3,4] and insect populations [5].
74 However, delays inherent in these systems (e.g., larval development times for mosquitoes,
75 ecological lags [6]) may pose challenges. The estimation of epidemiological delay distributions,
76 which we discuss in this paper, is a special case of using real-time data to understand the time
77 course of biological processes. Methods for estimating these delays could be broadly applicable
78 to other areas of computational biology.

79 Epidemiological delay distributions are key quantities for public health policy [7], clinical
80 practice [8], and infectious disease modeling [9–13]. An epidemiological delay is the time
81 between two epidemiological events, a primary event and a secondary event. There are
82 numerous examples of delays in infectious disease epidemiology. Three of the most important
83 delays include the incubation period (the time between infection and symptom onset), the serial
84 interval (the time between symptom onset in a given infected person and someone they infect),
85 and the generation interval (the time between infection in a given person and someone they
86 infect). Other commonly used epidemiological delays include hospital lengths of stay and delays
87 from symptom onset to hospitalization, hospitalization to death, and symptom onset to specimen
88 date, among others. Delays vary in length (e.g., between infected individuals) due to biological
89 and epidemiological factors [14–16], such that samples of delays can be characterized by a
90 distribution. These distributions are typically described by one or more parameters. Pair-
91 dependent delays are expected to be more heterogeneous than single-individual delays
92 because they involve more than one person.

93 Here, we focus on the incubation period and serial interval because they are key inputs
94 in mathematical and statistical models, such as those used for nowcasting/forecasting [9,17] or
95 for scenario-based modeling [18,19], which can guide control strategies. While the incubation
96 period must be positive, the serial interval can be positive or negative [20]. The right tail of the
97 incubation period distribution informs the length of quarantine, while the left tail indicates the
98 earliest time symptoms might develop after infection [8]. By comparing the incubation period
99 and serial interval, we can learn about a pathogen's tendency for pre-symptomatic versus
100 symptomatic transmission [21], which can inform the controllability of an epidemic [22]. At the
101 individual level, pre-symptomatic transmission occurs precisely when the generation interval is
102 shorter than the incubation period of the *infector* (or, equivalently, when the serial interval is
103 shorter than the incubation period of the *infectee*). The serial interval is often used as a proxy for
104 the generation interval.

105 Methods for estimating epidemiological delays have been improving, especially during
106 the COVID-19 pandemic, and recent research has highlighted the importance of appropriately
107 adjusting for three statistical issues inherent in the data collection process: censoring, right
108 truncation, and dynamical bias [23–27] (Table 1, Figures 1-2).

109 Not adjusting for bias can lead to incorrect estimates of delays, which can have direct
110 implications for public health practice. For example, Overton et al. found that the mean
111 incubation period for COVID-19 in early 2020 (corresponding to the ancestral strain of SARS-
112 CoV-2) was 3.49 days without adjusting for right truncation compared to 4.69 days when
113 adjusted [25]. Overton et al. further showed that the unadjusted estimate would suggest a
114 quarantine of 10 days would capture 99% of cases, compared to a 14-day quarantine for the
115 adjusted estimate [25]. If the unadjusted estimate from this study had been used to inform the
116 length of quarantine, more still-infectious individuals would have gone on to infect others.
117 Similarly, Park et al. found that ignoring right truncation for a fast-growing epidemic with

118 relatively long delays could result in underestimation of the mean delay distribution by up to
119 50% [23].

120 Not adjusting for bias can also lead to incorrect estimates of other parameters of interest
121 that rely on accurate estimates of delays, such as the reproduction number. The time-varying
122 reproduction number, R_t , is the average number of secondary cases caused by a single infected
123 individual in a population. Gostic et al. showed how mis-specifying the mean, variance, or form
124 of the generation interval led to biased estimates of R_t when using three empirical methods on
125 synthetic data [28]. The bias was greatest early in the epidemic. Underestimation of R_t during
126 this phase could lead to an insufficient public health response. Similarly, if changes in the
127 generation interval (or serial interval) over the course of the epidemic are not accounted for, R_t
128 may not be accurate [29].

129 In the aftermath of the COVID-19 pandemic, there is considerable need for and interest
130 in the estimation and reporting of epidemiological parameters, including delays, with at least two
131 R packages [30,31], a World Health Organization working group [32], and meta-analyses of
132 priority pathogens [33–35] working to collect and make these data accessible. In this
133 perspective, we present best practices for the estimation and reporting of epidemiological
134 delays, illustrated by examples for the incubation period and serial interval of directly transmitted
135 (person-to-person) infectious diseases. Our recommendations are based on our experience
136 estimating and using delays across multiple outbreaks and recent methodological work [23].

137 To make these best practices easier to follow, we developed two checklists (Tables 2-3)
138 and a flowchart (Figure 3), which can be used to understand which biases need to be adjusted
139 for based on available data. To provide context for our recommendations, we additionally
140 provide details about the data needed to estimate delays and how they should be prepared for
141 analysis. Then, we discuss biases that can affect the estimation of epidemiological delay
142 distributions, followed by strategies that can be used to reduce the impact of these biases.
143 Technical details about these biases and how to adjust for them can be found in Park et al. [23]

144 along with extensive simulation and case studies. Our paper goes beyond the work of Park et
145 al. by offering practical guidelines, a suggested workflow, and checklists, for a broader technical
146 audience, such as modelers at public health agencies.

147 **Data to estimate delays**

148 Primary and secondary events can be observed (e.g., symptom onset time) or
149 unobserved (e.g., usually infection time). Data to estimate the incubation period include the
150 times of probable exposure (from which infection time can be inferred) and perceived symptom
151 onset for each case. Data to estimate the serial interval are those in which symptom onset has
152 been observed for primary and secondary cases. Examples of study designs or public health
153 activities that generate such data include contact tracing [24,36], prospective cohort studies
154 [37–39], household studies [40], or other types of intensive cohort monitoring [41]. Data from
155 passive surveillance, which involves healthcare providers reporting cases to public health
156 agencies, can also be used to estimate delay distributions; however, key information about
157 exposures (e.g., dates, settings, and types of contact) may be missing, incomplete, or
158 abstracted from other variables [42,43].

159 **Biases in delay data**

160 Three main biases can affect the estimation of epidemiological delay distributions, 1.
161 censoring, 2. right truncation bias, and 3. dynamical (or epidemic-phase) bias (Table 1). All of
162 them affect both single-individual delays, such as the incubation period, and pair-dependent
163 delays, such as the serial interval.

164 Censoring is knowing that an event occurred but not precisely when. Data can be right
165 censored (the event is known to have occurred after a certain time), left censored (the event is
166 known to have occurred before a certain time), or interval censored (the event is known to have
167 occurred within a certain time interval). In epidemiological delay data, censoring can affect
168 either primary or secondary events (single interval censoring) or both (double interval censoring)
169 [44]. Epidemiological data are almost always doubly interval-censored due to discretization of

170 the end points of the intervals being measured. For example, when reporting occurs daily with a
171 cutoff at midnight, a patient could experience the event of interest (e.g., symptom onset) at any
172 time between 12:00 am and 11:59 pm on a particular day. Double interval censoring is shown in
173 Figure 1 by the brackets around each event (circles and squares). Events occur anywhere on
174 the x axis, which represents continuous time, but the reporting of events only occurs at discrete
175 time points, or observation times. Some events are prone to longer censoring intervals than
176 others (e.g., exposure intervals may be longer than one day for cases with multiple possible
177 exposures). Not or incorrectly accounting for censoring of event intervals can lead to biased
178 estimates of a delay [23].

179 Right truncation is defined as the inability to observe intervals (e.g. incubation periods)
180 greater than a threshold (e.g. greater than the number of days elapsed since infection). It
181 typically applies to real-time settings, when events with longer intervals may not have occurred
182 yet, leading to an overrepresentation of shorter intervals when estimating delays. Right
183 truncation is common in data where case ascertainment depends on the secondary event, e.g.
184 we rarely observe an individual's incubation period until after symptoms develop. Not
185 accounting for right truncation can lead to underestimating the mean delay [23]. Although right
186 truncation is mainly a problem for real-time analyses, retrospective data can be right-truncated if
187 surveillance ended prematurely.

188 Right truncation should not be confused with right censoring. The latter occurs when we
189 observe the primary event of a case or future case but cannot observe it long enough to witness
190 its secondary event [45], which could be due to, for instance, a study ending prematurely. As a
191 result, we only know that the secondary event did not occur during the observation period and
192 therefore have a right-censored interval for a data point. Right censoring is shown in Figure 1A.
193 The intervals on the top and bottom of the panel are included in the analysis, but we do not
194 know when the secondary events will occur because they happen after the observation time. In
195 contrast, right truncation means that certain delays are completely missing from our data as

196 observing primary events depends on identifying secondary events first. Right truncation is
197 shown in Figure 1B. Here, the intervals at the top and bottom of the panel are not included in
198 the analysis; we are unaware of these data points because their secondary events occur after
199 the observation time.

200 Dynamical bias is another type of common sampling bias which is analogous to right
201 truncation. During the increasing phase of an epidemic, patients with short delays are
202 overrepresented in the recent data, leading to underestimation of delay intervals. Conversely,
203 when the epidemic is decreasing, patients with long delays are overrepresented in the recent
204 data, leading to the overestimation of delay intervals. Dynamical bias is especially problematic
205 during periods of exponential growth and decay of cases when cases are exponentially more
206 and less likely, respectively, to be infected recently rather than further back in time.

207 **Measuring epidemiological delays**

208 We aim to estimate the true underlying distribution for each epidemiological delay which
209 characterizes the time between the primary and secondary event. In general, we assume that
210 this distribution does not change over the course of an epidemic (although this may not always
211 be the case [29]). Cases can enter a dataset due to observation of either their primary or
212 secondary event (shown by the arrows in Figure 2). Regardless of how data were collected, we
213 can organize our data into cohorts using either a forward or backward approach. For the forward
214 approach, we start from primary events that occurred during the same period and prospectively
215 determine when the secondary events occurred—the resulting distribution of the delays is the
216 forward distribution (Figure 2A and 2C). In contrast, for the backward approach, we start with
217 secondary events that occurred during the same period and retrospectively determine when the
218 primary events occurred—the resulting distribution of the delays is the backward distribution
219 (Figure 2B) [23].

220 Data observed in real-time can be subject to either right truncation or right censoring.
221 Right truncation causes the observed forward distribution to be shorter than the true underlying

222 distribution and has the largest effect when the epidemic is growing because relatively many
223 recently infected individuals with long delays are not observed. In a declining epidemic, right
224 truncation will have a smaller impact on the forward distribution because the proportion of
225 recently infected individuals with unobserved secondary events is lower [23]. Excluding right-
226 censored data from the analysis is equivalent to right truncating the data and leads to
227 underestimation of the delay. Backward distributions are not susceptible to right truncation but
228 can have a delay distribution that is shorter or longer than the forward distribution and the true
229 underlying distribution depending on the phase of the epidemic (i.e. dynamical bias). Both right
230 truncation and dynamical biases are minimal if data from the entire epidemic are available and
231 included in a delay estimate [23].

232 Figure 1 shows the impact of different biases on the forward and backward distributions.
233 We recommend always analyzing delay distributions as forward distributions and accounting for
234 potential biases (i.e., censoring and right truncation) as this approach does not require
235 additional information on incidence trajectories, which are used to adjust for dynamical bias in
236 the backward distribution [23].

237 **Adjusting for common biases**

238 Figure 3 shows a decision tree for assessing which biases need to be addressed
239 depending on the approaches taken for data collection and processing. For each of three
240 possible scenarios, it includes an explanation about the potential impact of the biases as well as
241 the methods needed to adjust for them.

242 In general, adjusting for double interval censoring involves estimating the conditional
243 probabilities of the primary and secondary events occurring between their observed lower and
244 upper bounds. Interval censoring should always be adjusted for, and the adjustment method is
245 the same irrespective of the epidemic phase [23]. Right censoring can be handled using
246 methods for survival analysis, such as the Kaplan-Meier approach [45]. Adjusting for right
247 truncation involves normalizing the probability of observing a given delay from the untruncated

248 forward distribution by the probability of observing any delay before the final observation time,
249 and adjusting for dynamical bias involves incorporating the epidemic trajectory (e.g., the growth
250 or decay rate of the epidemic) into the analysis [23,26,46,47].

251 There are several available methods and tools for estimating epidemiological delay
252 distributions; however, most of these approaches have not been validated or do not correct for
253 all potential biases [23]. One example is the *coarseDataTools* R package developed by Reich et
254 al. [48]. This tool has been validated and can correct for double interval censoring, but it does
255 not adjust for right truncation or dynamical bias. In contrast, *epidist* [49], an R package
256 developed by some of the authors of this study, contains methods which can adjust for all three
257 potential biases. In Park et al. [23], a simulation study was used to evaluate multiple methods
258 and found that the approximate latent variable censoring and truncation method emerged as the
259 best performer for both real-time and retrospective analyses for most real-world use cases. This
260 method corresponds to the double interval censoring and right truncation adjusted model
261 developed by Ward et al. [24] which we recommend.

262 Ward et al.'s approach estimates the probability of observing a secondary event
263 conditional on observing the primary event by a given final observation time [24]. Estimated
264 event times for each case are included in the model as unobserved, or latent, variables, and
265 uniform prior distributions are used for both the primary and secondary event times, which can
266 accommodate censoring intervals of arbitrary length. However, when censoring intervals are
267 long, the event time distribution within the censoring interval will deviate from the uniform
268 approximation (as its shape depends on underlying epidemic dynamics) and should be taken
269 into account [24]. This model has important limitations. For example, Park et al. [23] found that
270 this method was not able to estimate the mean or standard deviation as well in epidemic
271 simulations characterized by very rapid exponential growth and long delays because primary
272 event times are poorly approximated by a uniform distribution.

273 Certain practices should be avoided if using an alternative approach to adjust for biases
274 in delays. For example, we suggest avoiding approaches that adjust for right truncation and
275 dynamical biases simultaneously because they lead to overestimation of the mean delay by
276 overcompensating for intervals that have not yet been observed [23]. We also do not
277 recommend using a midpoint imputation rule (i.e., use the midpoint of the interval as the
278 “observed” value and construct downstream inferences based on that imputation) on interval-
279 censored data as such imputation approaches may introduce bias [23,50].

280 **Additional modeling recommendations**

281 Beyond correctly adjusting for biases, there are several common issues with reported
282 epidemiological delays that may lead to biased conclusions when used in practice or impact
283 their ability to be used at all. Historically, the incubation period and serial interval were often
284 reported using only the mean (and sometimes the range) [51–54]. However, models can be
285 fitted to delay data to adjust for some of the biases we have described and better characterize
286 the tail of the distribution. Assuming a modeling approach is taken, we summarize our
287 recommendations for estimating and reporting for epidemiological delay distributions in Table 2
288 and give more details in the sections “Reporting epidemiological delay distributions” and
289 “Reporting the incubation period and serial interval.” Table 2 contains details about each
290 recommendation, examples of diseases for which it has been implemented, and possible
291 solutions. It is divided into two sections, estimation and reporting. Whenever estimating delays,
292 we recommend going through the table to make sure all the steps are taken.

293 *Fit multiple probability distributions.* We recommend fitting multiple probability distributions to
294 summarize the empirical delay distribution [55]. This approach has greater utility for users of the
295 estimates compared to non-parametric approaches. Use appropriate model comparison criteria
296 (e.g., widely applicable information criterion [WAIC] or leave-one-out information criterion
297 [LOOIC] for Bayesian models) to suggest the best-fitting model alongside visual checks.

298 Common distributions for epidemiological delays in the literature include the gamma, lognormal,

299 and Weibull distributions [56]. For delays that can have negative values, distributions that
300 accept negative values, such as the skew-normal or skew-logistic distributions [57], may be
301 used, or less ideally, the delay data may be shifted to allow for fitting of distributions that only
302 allow positive numbers [58]. Mixture distributions may be appropriate for some delays, such as
303 those with bimodal distributions [59–62]. Fitting parametric distributions may not be appropriate
304 for all delay distributions. Semi- or non-parametric approaches, such as hazard models, may be
305 considered [63]. Non-parametric estimates can also be used to assess the relative fit of
306 parametric models [27].

307 *Visualize the distributions.* It is also important to visualize the fitted distributions to check that
308 they fit the data [36]. When doing so, we recommend visualizing the estimated distribution in
309 conjunction with the modeled observation process (e.g., double interval censoring and right
310 truncation). In other words, estimate the latent (continuous) distribution. Then, from the latent
311 distribution, simulate elements of the observation process, such as double interval censoring
312 (e.g. for date-level censoring this will transform continuous delay times into an integer number of
313 days elapsed) and right truncation (this will change the shape of the observed distribution,
314 relative to the latent distribution) [49]. Not accounting for the observation process after
315 estimating the latent distribution makes visual assessment of the fit difficult, because the
316 observation data and the latent distribution may differ in shape and data type. For example,
317 Sender et al. illustrate how good visualizations can help make intuitive comparisons across
318 different distributions [64].

319 *Correctly convert parameters.* Care should be taken when converting the parameters of fitted
320 probability distributions to the summary statistics of interest. For example, the gamma
321 distribution may use either a scale or a rate parameter, in addition to its shape parameter [65],
322 while the standard lognormal parameters, log mean and log sd do *not* correspond to the log of
323 the mean and the log of the standard deviation of the lognormal [66]. Some R packages contain
324 functions to perform parameter conversion, such as EpiNow2 (lognormal) [67], epitrix (gamma)

325 [68], mixR (gamma, lognormal, and Weibull) [69], and epiparameter (gamma, lognormal,
326 Weibull, negative binomial, and geometric) [30].

327 *Add subgroups or stratify estimates.* If sample size allows, we recommend stratifying delay
328 estimates whenever there are hypothesized differences across groups as delays, such as the
329 incubation period and the serial interval, may vary by route of exposure [15], viral species [70] or
330 clade, disease severity, sex [71], age [71], vaccination status, or other factors. Ideally, this
331 stratification should be done jointly in a statistically robust framework [49,63,72]. Wider
332 application of joint modeling approaches could be achieved with more availability of easy-to-use
333 tools [49].

334 *Check model diagnostics.* When Bayesian methods are used, visualize the posterior distribution
335 against data and check model diagnostics, such as trace plots, the potential scale reduction
336 statistic (R-hat), divergent transitions, and effective sample sizes, and report them [73–75].
337 Convergence issues may indicate that the model is mis-specified, making the results unreliable.
338 For more advice about using Bayesian methods, see [73,74,76].

339 **Reporting epidemiological delay distributions**

340 We recommend reporting an estimate of variability (e.g., standard deviation or
341 dispersion) along with central tendency (e.g., mean or median) for all estimated delay
342 distributions (sometimes more than one distribution fits the data similarly well as in [36])
343 alongside the quantiles and underlying parameters of the fitted distributions. These quantities
344 are often used as inputs in infectious disease models and can inform both clinical practice and
345 public health policy [8]. If possible, the probability density function should be specified to avoid
346 ambiguity about the parameters. For Bayesian analyses, samples of the posterior distribution
347 should be made available as summarizing estimates may obfuscate valuable information about
348 the correlations between parameters and the shape of the posterior distributions.

349 All summary statistics should always be accompanied by credible intervals or confidence
350 intervals for Bayesian and frequentist analyses, respectively (usually 90% or 95% with the width

351 of the reported interval also being reported). High uncertainty in parameter estimates can have
352 substantial impacts on downstream modeling [77,78] and can indicate that more data need to
353 be collected.

354 Estimates of delays should be accompanied by contextual information to aid in
355 interpretation. For example, we recommend reporting the study sample size; the epidemic
356 curve; which, if any, control measures are in place; and summary statistics on age, sex,
357 geographic location, vaccination status, and possible exposure route(s). Control measures and
358 summary statistics can be used to assess generalizability of the estimates (see earlier advice on
359 stratification of estimates). The epidemic curve can indicate at which stage of the epidemic the
360 analysis took place and whether the outbreak is now over (i.e., whether certain biases need to
361 be adjusted for).

362 Code and data should be uploaded to repositories, such as GitHub (<https://github.com/>)
363 or Zenodo (<https://zenodo.org/>), to ensure reproducibility of the analysis and facilitate re-use of
364 the code. Apart from allowing others to reproduce, validate and potentially improve analyses,
365 providing data along with estimates of delay distributions also ensures that the estimates can be
366 integrated in future pooled estimation efforts as methods continue to be improved. These data
367 should ideally be provided in linelist format with all necessary information required for estimation
368 (e.g., the left and right boundaries of the possible infection and symptom onset times for the
369 incubation period [44]). Importantly, the data should be anonymized/de-identified to protect
370 patient privacy according to local health data laws and regulations. If data cannot be shared, we
371 recommend at minimum providing samples of the posterior distribution in a permanent online
372 repository to facilitate future re-analyses (as in [24]).

373 **Reporting the incubation period and serial interval**

374 In addition to the checklist for reporting epidemiological delay distributions, we
375 recommend additional considerations specific to the incubation period and serial interval in
376 Table 3. Like Table 2, this checklist contains details about each recommendation, examples

377 from real outbreaks, and possible solutions. It should be used each time an incubation period or
378 serial interval is estimated.

379 For the incubation period, a case may have had multiple possible exposures prior to
380 symptom onset, especially when community transmission of a pathogen is high. If a case
381 reports multiple exposures, we recommend defining an exposure window that includes all
382 possible exposure dates [7], using disjointed exposure windows where appropriate. Other
383 methods that take this uncertainty into account could be used (such as [79] who used a
384 Bayesian framework to infer the incubation period of the infector). We caution against restricting
385 the analysis to cases with a high degree of certainty about their exposure periods as this can
386 introduce biases [80].

387 For the serial interval, we only use case pairs where we are fairly confident transmission
388 has occurred [58] (usually based on exposure information collected from patient interviews, see
389 [36,81]). Although this approach could bias the serial interval towards specific lengths of
390 intervals, it is usually preferable than to using mis-specified case pairs. Bias from including mis-
391 specified case pairs in the analysis is likely larger than that from removing pairs between which
392 transmission likely did not occur; however, future research should formally assess this
393 convention with simulations. In terms of the direction of transmission, a number of approaches
394 can be taken to order the case pairs. Where there is strong evidence that pre-symptomatic
395 transmission does not occur for the disease of interest, reported negative intervals are
396 presumed to be erroneous, and can be removed [36] or reversed [24]. Some studies assume
397 the direction of transmission between epidemiologically linked cases based on the date order of
398 symptom onset [24]; where negative serial intervals are possible, this should be avoided. It is
399 also possible to use genomic data to order the pairs [82,83]. The ideal approach where negative
400 intervals are possible is to use information about pair ordering and fit a distribution that allows
401 for negative values (see “Additional modeling recommendations”). When no such information is
402 available, a method that does not rely on knowing the order of case pairs would be ideal as this

403 would enable the use of more available data and avoid biases from certain types or durations of
404 exposure being easier or more difficult to link epidemiologically. Although such methods have
405 been developed, they assume a fully sampled population [84,85]. It is important to be
406 transparent about the approach taken. Uncertainty in the source of infection (such as the
407 potential for multiple possible infectors) should also be considered.

408 **Other considerations**

409 *How to best use new data.* Delays should be (re-)estimated when possible, especially if current
410 estimates are poor or lacking. New estimates can be compared to those from previous
411 epidemics or from a different phase of the same epidemic. One could consider any new dataset
412 on its own. Alternatively, using mixed effects models to partially pool information across different
413 outbreaks is likely a better use of available data. When choosing an approach, it is important to
414 consider whether the primary modes of transmission or pathogen properties are different than in
415 the past [56,86].

416 *Choice of prior distributions.* When using Bayesian methods to estimate epidemiological delay
417 distributions, it is important to think about the choice of prior distributions. Some R packages
418 that estimate epidemiological delays have default prior distributions [48,49]. Users of these tools
419 should carefully consider the appropriateness of default prior distributions for their analyses and
420 should explore the impact of different prior distributions on their results. We recommend against
421 using uninformative prior distributions, especially uniform prior distributions [87], for the
422 parameters of epidemiological delay distributions. When the delay distribution is already well
423 reported in the literature, this knowledge can be used to inform prior distributions; however, the
424 methods should be clearly communicated and accompanied by estimates generated from
425 weakly informative prior distributions as sensitivity analyses.

426 *Meta-analyses.* To reduce uncertainty from small sample sizes, some researchers have
427 combined estimates of epidemiological delays from different studies through meta-analysis
428 [34,35] or pooled analysis (re-analyzing published individual-level data) [8,88]. The latter

429 method is preferred but may not always be possible. If performing a meta-analysis, we
430 recommend performing sensitivity analyses when some estimates from the literature have not
431 been corrected for bias [89]. Published estimates can also be adjusted for bias post-hoc by
432 using the relationship between the backward and forward distributions, as in Park et al. [89].
433 Some authors have designed custom quality assessment scales to assess bias in studies
434 included in meta-analyses of epidemiological parameters [33,34].
435 *Time-varying delays.* While some delay distributions are expected to remain stationary during
436 an epidemic wave, others can change over time in response to interventions [29] and changes
437 in reporting, among other factors. For example, Ali et al. found that the serial interval of COVID-
438 19 in mainland China shortened after non-pharmaceutical interventions were implemented in
439 early 2020 [29]. We can study time-varying delays by analyzing changes in forward delay
440 distributions across cohorts. When doing so, we still need to account for both censoring and
441 right truncation. We do not recommend inferring time-varying delays from backward delay
442 distributions because dynamical bias also causes changes in this distribution even when the
443 forward distribution remains stationary.

444 **Discussion**

445 Epidemiological delay distributions are key parameters for preparedness and response
446 to epidemics and pandemics. Their importance has been highlighted during the COVID-19
447 pandemic [29], the global mpox (formerly known as monkeypox) outbreak in 2022 [7], and other
448 epidemics over the last two decades [13]. We have focused on the incubation period and serial
449 interval, for which estimates are often made at the beginning of an epidemic when contact
450 tracing data are available to support early characterization of the pathogen. While the estimates
451 are most useful for real-time response during this time, they are also the most susceptible to
452 bias. Indeed, adjusting for bias when estimating delay distributions is one of the most important
453 recommendations we highlight. as not doing so may lead to the propagation of bias into
454 downstream modeling [23] and therefore an incorrect understanding of an epidemic (e.g., over-

455 or under-estimating the risk to the host population) [25]. In addition to adjusting for bias,
456 estimates need to be clearly and fully reported to maximize utility and make the most of data
457 that are both costly and difficult to collect [90,91]. Our recommendations can assist with this.

458 A limitation of current methods for correcting common biases is that they do not fully
459 account for time-varying changes in delay distributions [23]. Future work on delay distributions
460 or nowcasting (which demonstrates how to model time-varying delays using time-to-event
461 modeling) [92–94] should extend current methods or develop new methods to account for these
462 changes.

463 Many of the best practices outlined in this paper also apply to other epidemiological
464 delays. However, there are issues which we did not cover. For example, we did not focus on
465 methods for estimating delay distributions for vector-borne diseases, such as dengue and
466 Yellow fever, as these require additional considerations (e.g., accounting for vector biology)
467 [13,95]. Also, we did not aim to provide a systematic review; rather, we provided insights based
468 on our experiences. The examples presented in this work were selected to illustrate specific
469 points.

470 In conclusion, we have provided recommendations in the form of two checklists for
471 generating and evaluating epidemiological delay distributions. We gave examples of good
472 practice for the incubation period and serial interval from various infectious disease outbreaks
473 over the last few decades; though few examples in the literature incorporate all the best
474 practices outlined in this paper. We hope that our recommendations will provide clarity and
475 structured guidance about what should be reported and how to adjust for biases in delay data.
476 We also hope our flowchart and checklists will be adopted by the infectious disease modeling
477 community to understand the limitations of existing estimates and improve future estimates.

478
479 **Funding statement**
480

481 AC is funded by the National Institute for Health and Care Research (NIHR) Health Protection
482 Research Unit in Modelling and Health Economics, a partnership between the UK Health
483 Security Agency, Imperial College London and LSHTM (grant code NIHR200908); and
484 acknowledges funding from the MRC Centre for Global Infectious Disease Analysis (reference
485 MR/X020258/1), funded by the UK Medical Research Council (MRC). This UK funded award is
486 carried out in the frame of the Global Health EDCTP3 Joint Undertaking. SC acknowledges
487 funding support from the Laboratoire d'Excellence Integrative Biology of Emerging Infectious
488 Diseases program (grant ANR-10-LABX-62-IBEID) and the INCEPTION project (PIA/ANR16-
489 CONV-0005). The funders had no role in study design, data collection and analysis, decision to
490 publish, or preparation of the manuscript.

491

492 **Disclaimer**

493

494 The findings and conclusions in this report are those of the authors and do not necessarily
495 represent the official position of the CDC, U.S. Department of Health and Human Services,
496 NIHR, UK Health Security Agency, or the UK Department of Health and Social Care.

497

498 **Author contributions**

499

500 Conceptualization: KC, SWP, SA

501 Data curation: NA

502 Formal analysis: NA

503 Funding acquisition: NA

504 Investigation: KC, SWP, SA

505 Methodology: NA

506 Project administration: NA

507 Resources: NA

508 Software: NA

509 Supervision: SA

510 Validation: NA

511 Visualization: KC, SWP, SA, AC

512 Writing - original draft preparation: KC, SWP

513 Writing - review & editing: All authors

514

515 **Data availability statement**

516

517 No data were used in this study.

Tables

Table 1. Adjusting for biases in epidemiological delay distributions. Examples of diseases for each bias were selected based on convenience (either papers written by this study's authors or those encountered during the course of their work). Note that not all methods in this table are discussed in the text, and more details can be found in Park et al. [23].

Bias	Interval censoring	Right truncation	Dynamical bias
Details	The exact timing of the primary event or secondary event (single interval censoring) or both events (double interval censoring) is unknown (e.g., except for experimental studies, exposure is usually reported as a date or range of dates, rather than a time of day).	Right truncation is a type of sampling bias. It arises because only cases whose secondary event has occurred can be observed. In an ongoing epidemic, right truncation biases the incubation period and serial interval toward shorter intervals because individuals with longer incubation periods may not have developed symptoms or have been reported yet.	Dynamical bias is another type of sampling bias that can be present when case ascertainment is based on the secondary event. It is related to epidemic dynamics: during a growth phase, cases that developed symptoms recently are overrepresented in the observed data, while during a declining phase, these cases are underrepresented. This means that the backward distribution is not representative of the forward distribution during these periods.
Impact	Not accounting for interval censoring can lead to biased estimates of a delay's standard deviation. Incorrectly accounting for it can also bias the mean.	Not accounting for right truncation can lead to underestimation of the mean delay [23].	Not adjusting for dynamical bias when estimating the forward distribution from the backward distribution can lead to under- or over-estimation of delay intervals depending on whether the epidemic is in a growth or declining phase, respectively.
Diseases for which this bias has been considered in analyses	Incubation period: mpox [36]; Zika [96]; COVID-19 [97]; 6 vector-borne diseases [88] Serial interval: mpox [36]	Incubation period: COVID-19 [25,97] Serial interval: mpox [24]	Incubation period: COVID-19 [89] Serial interval: mpox [24]

Possible solutions	Use methods, such as Reich et al. 2009 [44], that adjust for double interval censoring; however, this method does not adjust for right truncation. Alternatively, use Ward et al.'s double interval censoring and right truncation adjusted model, which adjusts for interval censoring and right truncation simultaneously [23,24].	Use an approximate latent variable method, such as Ward et al's double interval censoring and right truncation adjusted model [24] or similar alternatives [23].	If both primary and secondary event dates are known and the incidence of primary events is changing exponentially at a constant rate, it is possible to use the approach of Verity et al., Britton et al., and Park et al. [26,46,47] to adjust for dynamical bias; however, uncertainty in both growth rate estimates and observed delays need to be taken into account carefully with this approach and the assumption of constant growth rates may not be met in practice. Park et al. present a version of this method that allows for a time-varying growth rate, but it requires untruncated incidence data or assumptions to be made about the recent growth rate [23]. For most settings, considering the forward distribution and accounting for right truncation is recommended.
Practices to avoid	Do not adjust for single censoring if the data are doubly interval-censored as this will result in a biased mean [23]. Note that even if both primary and secondary events are reported to have occurred on a single date, the data should be considered doubly interval-censored. Do not use a midpoint imputation rule on interval-censored data as this may introduce bias [50].	Do not adjust for right truncation and dynamical bias at the same time early in an outbreak as doing this can lead to overadjustment of the downward bias and therefore to overestimation of the delay [23].	Avoid adjusting for right truncation and dynamical bias at the same time [23]. When analyzing the forward distribution, adjust for right truncation; when analyzing the forward distribution via the backward distribution, adjust for dynamical bias. We recommend the former when possible.

Table 2. Checklist for reporting/reported epidemiological delay distributions. Examples of diseases for each checklist item were selected based on convenience (either papers written by this study’s authors or those encountered during the course of their work).

Checklist item	Details	Diseases for which this item has been implemented	Possible solutions
Estimation			
Adjust for biases	See Table 1	See Table 1	Adjust for censoring (always), right truncation (when needed), and dynamical bias (when needed). Clearly state that all these adjustments have been made, and report both right-truncation-adjusted and right-truncation-unadjusted estimates. Consider using the approach of Reich et al. [44] or Ward et al’s double interval censoring corrected model [23,24] to obtain estimates that are not adjusted for right truncation.
Compare multiple probability distributions	Estimated delays may depend on the fitted probability distribution, so it is important to use the distribution that best represents the data.	Incubation period: mpox [56]; COVID-19 [97,98]; dengue [95]; HIV [71]; malaria [60] Serial interval: MERS [79]	Fit more than one probability distribution to the data [55] and use appropriate model selection criteria to compare them. Visualize the fit of distributions to the data.
Correctly convert parameters of probability distributions to summary statistics	Incorrectly converting parameters leads to wrong estimates of delays.	Incubation period: COVID-19 [97]	If writing equations, double check them or use a software package with built-in functions for parameter conversion, such as EpiNow2 (lognormal) [67], epitrix (gamma) [68], mixR (gamma, lognormal, and Weibull) [69], and epiparameter (gamma, lognormal, Weibull, negative binomial, and geometric) [30].

Add subgroups to your model or stratify the estimates whenever appropriate	Estimated delays may vary based on a variety of factors.	Incubation period: Ebola (virus species) [70]; West Nile virus disease (transplant/transfusion versus not) [88]; dengue (serotype) [95]; HIV (sex and age category) [71]	If sample size allows and a difference across groups is hypothesized, add subgroups to your model or stratify the estimates by exposure type, genetic variant/clade, or other factors (e.g., sex, age, vaccination status, etc.). The former approach, also known as joint modeling, is preferred.
Check model diagnostics	If model requirements are not met, the estimated delay distributions may not be reliable.	Incubation period: dengue [95]	If using Bayesian methods, make sure all models converge. Visually inspect trace plots. Check potential scale reduction statistic (R-hat), divergent transitions, effective sample sizes, and any other diagnostics and report when appropriate. Examples of R packages to perform these checks include coda [99], rstan [75], and bayesplot [100].
Reporting			
Report measures of central tendency and variability	These may be used as inputs for infectious disease models and can inform both clinical practice and public health policy [8].	Incubation period: mpox [56], COVID-19 [97,98]; 6 vector-borne diseases [88]; Zika [96]; 9 respiratory diseases [8] Serial interval: MERS [79]; mpox [36]	Report multiple summary statistics and clearly state which is which. Report at least the mean or median as well as standard deviation, variance, or dispersion.
Report quantiles of the probability distribution	The left tail of the incubation period distribution indicates the earliest time symptoms could develop following	Incubation period: mpox [56]; COVID-19 [97,98]; 6 vector-borne diseases [88]; Zika	Report key quantiles (e.g., 2.5, 5, 25, 50, 75, 95, 97.5, 99) of the distribution in a table.

	infection, while the right tail is often of interest for control strategies, such as monitoring people who have been exposed [25,98]. The right tail of the serial interval can also inform the length of quarantine.	[96]; 9 respiratory diseases [8] Serial interval: mpox [24]	
Report the parameters for all fitted probability distributions	These inputs may be used for mathematical modeling and are key for correctly defining the distribution.	Serial interval: mpox [36,101]	E.g. for the gamma distribution, report shape and scale; for the lognormal distribution, report logmean and log standard deviation. If possible, the probability density function should be specified to avoid ambiguity about the parameters.
Report uncertainty in the estimates	Communicating uncertainty is a key aspect of outbreak analysis and modeling [78]. Also, high uncertainty in parameter estimates can affect downstream modeling [77].	Incubation period: mpox [7,56,101]; COVID-19 [97,98]; 6 vector-borne diseases [88]; dengue [95]; Zika [8,13]; MERS [79,102]; 9 respiratory diseases [8] Serial interval: mpox [24,36]; COVID-19 [20]	Report 90% or 95% credible intervals or confidence intervals for all estimates (central tendency, variability, quantiles, and parameters for the fitted probability distributions). Ideally, provide joint posterior samples for Bayesian analyses as these are important for characterizing covariance in the posterior distribution. Make sure to report how intervals were defined (i.e. report that they are 95% credible intervals).
Report characteristics about study sample	Characteristics about the study sample can provide epidemiological context for the estimates which can help with interpretation.	Incubation period: MERS [79]; 9 respiratory diseases [8]; mpox [56] Serial interval: MERS [79]	Report the sample size of the study, demographic characteristics of patients (e.g., age, sex, geographic location, vaccination status), and route of exposure(s) (if known).

<p>Report the epidemic curve and which, if any, control measures are in place</p>	<p>The epidemic curve can provide context about the epidemic phase (increasing, decreasing, or stable) and whether right truncation bias or dynamical bias needs to be considered.</p>	<p>Incubation period: mpx [101] Serial interval: COVID-19 [46]; MERS [102]</p>	<p>Include a figure of the epidemic curve or provide a reference to the curve on a permanent website (doi). Ideally, the underlying data for the curve would be made available to download. Alternatively, provide an estimate of the growth rate for the study period.</p>
<p>Provide anonymized/de-identified data and documented code</p>	<p>This step improves the reproducibility of the study, and the code can be reused by other teams during future epidemics. It can also facilitate meta-analyses and joint analyses of multiple datasets.</p>	<p>Incubation period: COVID-19 [97,98] Serial interval: mpx [81]</p>	<p>Anonymized/de-identified linelist-level data should be provided with relevant stratifying variables. For small epidemics, some authors have reported data relative to an unspecified reference date to protect patient identities [56]. However, the data should still ideally be linked to the epidemic trajectory to address dynamical bias issues. An alternative approach could be to widen the censoring intervals.</p>

Table 3. Additional checklist items for reporting the incubation period and serial interval.

Delay	Checklist item	Details	Diseases for which this item has been implemented	Possible solutions
Incubation period	Investigate the potential for multiple possible exposures	Sometimes, a case may have had multiple opportunities to be exposed before symptom onset.	Mpox [103]; MERS [79]	For these cases, use an exposure window that includes all possible exposure dates, such as travel to a high-risk area. Use a disjointed exposure window where appropriate (however, this approach may cause some issues for the sampler with Ward et al.'s double interval censoring and right truncation adjusted model [24]). The method developed by Cowling et al. 2015 [79] for MERS could be used as well.
Serial interval	Check for negative serial intervals	Negative serial intervals can occur when symptom onset in the infectee occurs before symptom onset in the infector [20].	Mpox [24,36]	Assuming the data are correct, including negative serial intervals, keep the ordering of the pairs and fit a distribution that allows for negative values (such as normal). If there are negative serial intervals but there is strong evidence that pre-symptomatic transmission does not occur for the disease of interest, consider removing [36] or reversing the order [24] of those case pairs. There are also methods that do not depend on knowing the order of case pairs [84,85].
	Investigate the potential for multiple possible infectors	Sometimes, a case could have been exposed to more than one infected person prior to symptom onset.	Mpox [36,81]; MERS [79]	Restrict the analysis to only cases with a high degree of certainty that the secondary case was infected by the primary case. Then, do a sensitivity analysis with all cases and compare the results.

Figures

Figure 1. Common biases involved in the estimation of epidemiological delay distributions. The y axis in each panel represents distinct observations of delays. The circles represent primary events, while the squares represent secondary events. The horizontal lines represent the delay between events, while the vertical lines represent the time at which the data are observed. The arrows represent how cases enter the dataset: arrows pointing toward the right indicate that the primary event was observed first, while arrows pointing toward the left indicate that the secondary event was observed first. The brackets “[]” represent interval censoring of the primary and secondary events. Note how primary and secondary events occur in continuous time, while the reporting of events (brackets) always occur at discrete intervals. Delays and events in gray are unobserved. (A) and (B) demonstrate the same scenario, but the direction of the arrows is different. In (A), observation of the delay is based on the primary event, and there is right censoring, while in (B), observation is based on the secondary event, and there is right truncation. (C) demonstrates an example of a backward distribution in a growing epidemic, when the majority of delays that make up the distribution will be short; hence, the backward distribution will be shorter than the equivalent forward distribution. (D) demonstrates the reverse in a declining epidemic. Both (C) and (D) show the impact of dynamical bias.

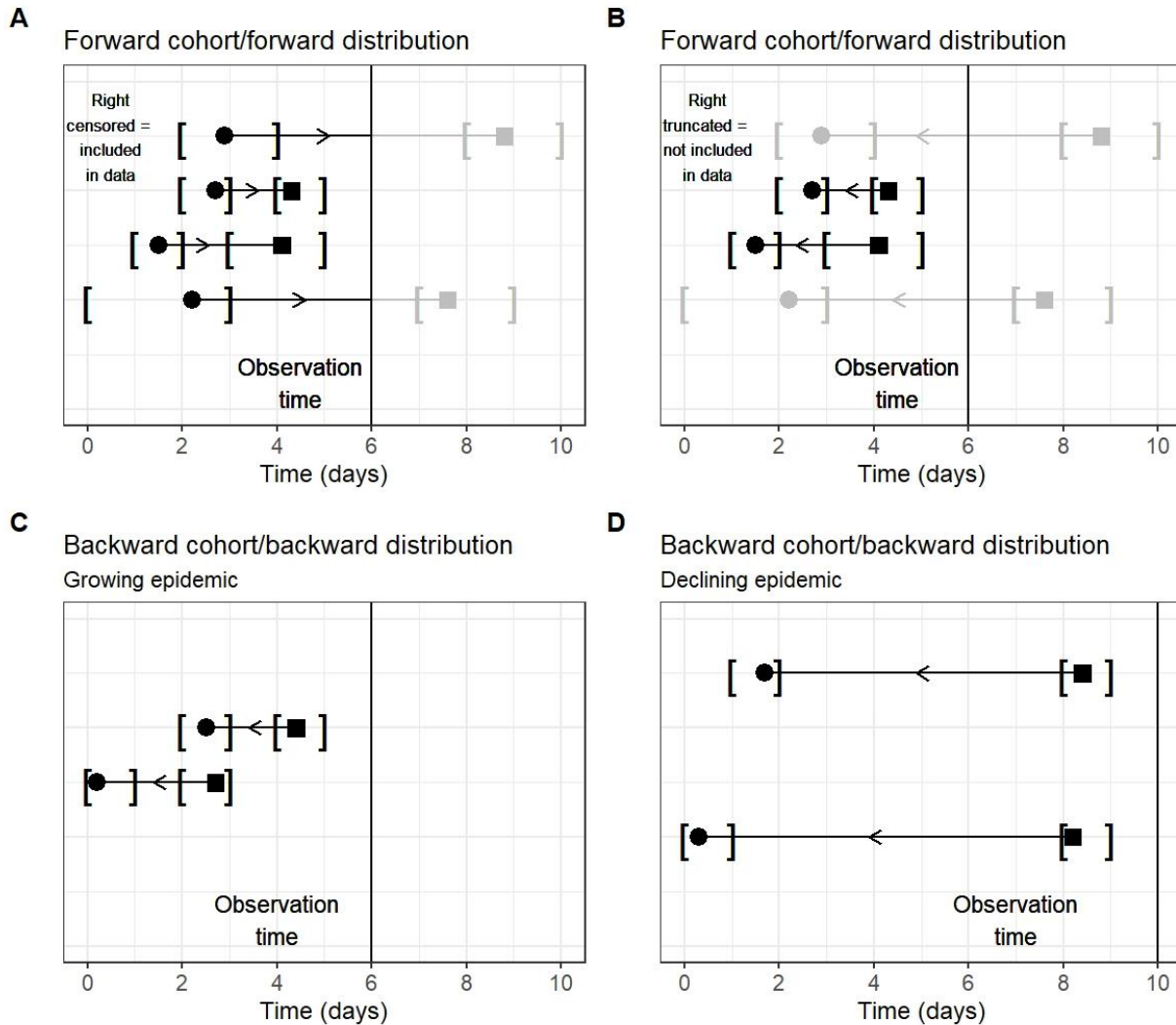
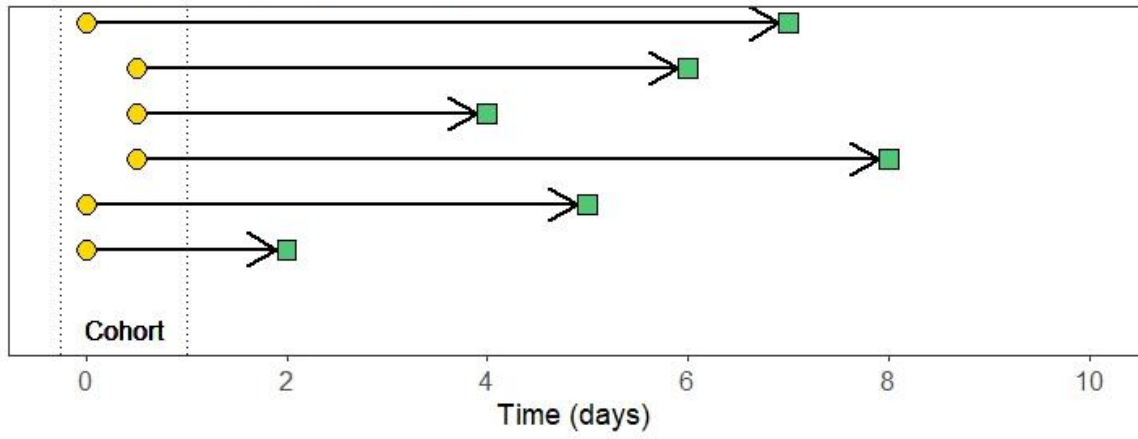


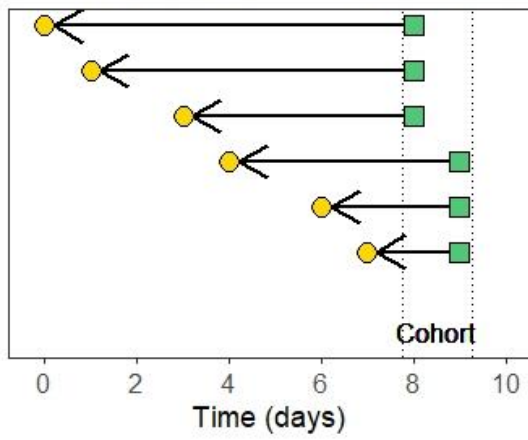
Figure 2. Forward and backward approaches for organizing events into cohorts and analyzing data to estimate epidemiological delay distributions. The y-axis represents distinct observations of delays. The yellow circles represent the primary events, while the green squares represent the secondary events. The black horizontal lines represent the delay between primary and secondary events, and the vertical dotted lines show the cohorts. The arrows represent how cases enter the dataset: arrows pointing toward the right indicate that the primary event was observed first, while arrows pointing toward the left indicate that the secondary event was observed first. Note that the case ascertainment method does not impact the direction we can cohort the data. For forward cohorts (A and C), all primary events that occurred during the same period are selected and prospectively followed until the secondary event occurs. For backward cohorts (B), all secondary events that occurred during the same period are selected; the timing of the primary events is identified retrospectively. For simplicity, interval censoring has been omitted from this figure; events appear to occur at discrete intervals when in reality, they occur in continuous time.

A

Forward cohort/forward distribution

**B**

Backward cohort/backward distribution

**C**

Forward cohort/forward distribution

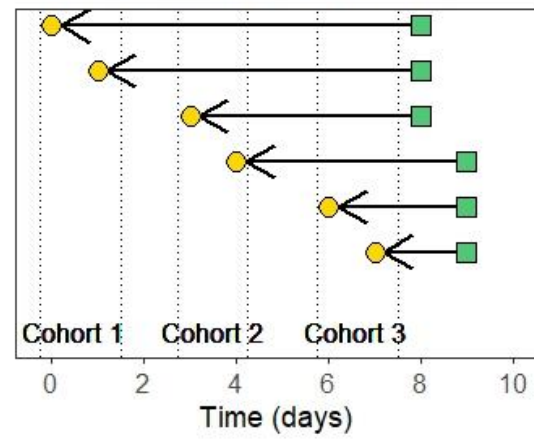
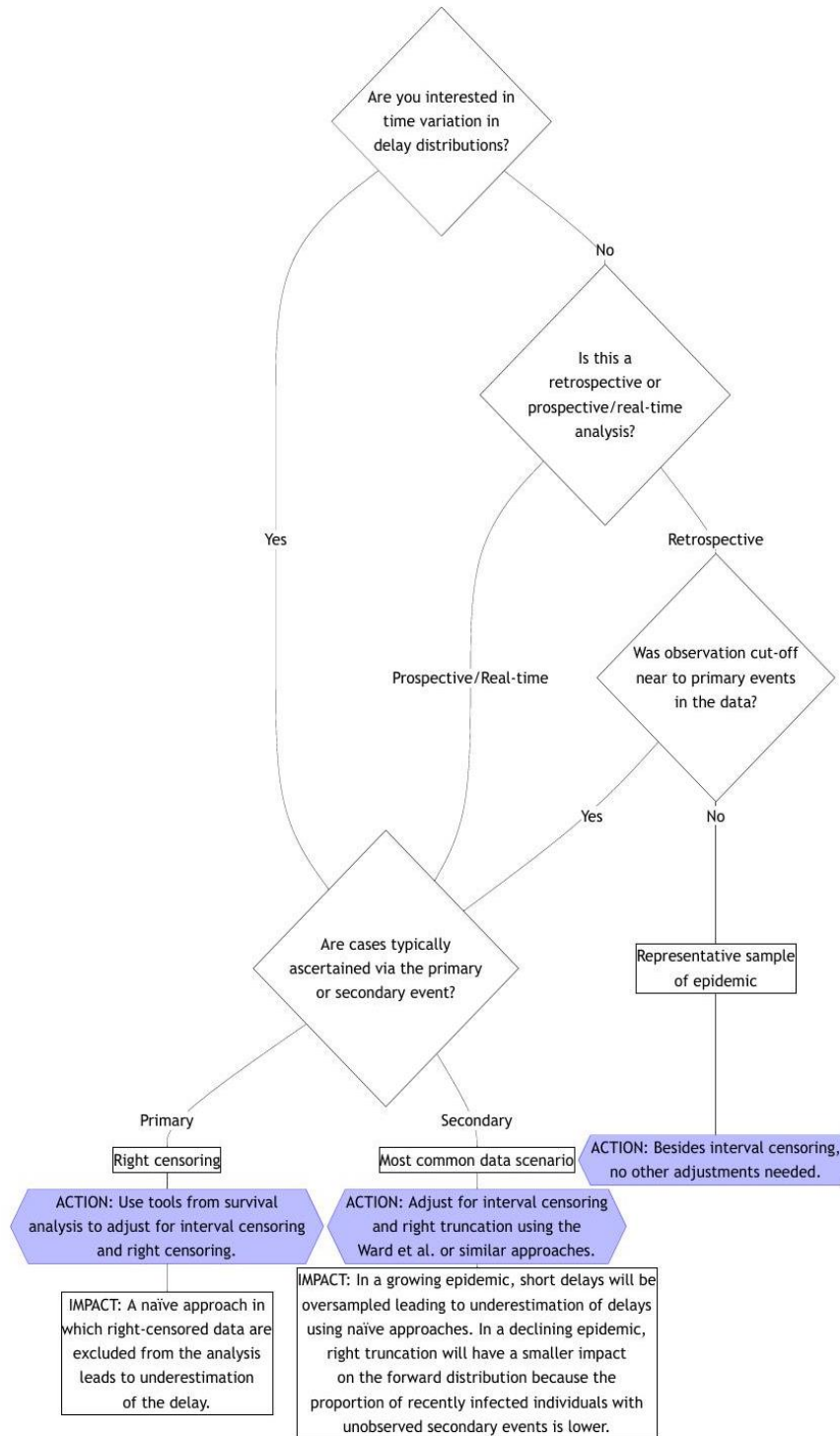


Figure 3. Flowchart for bias adjustment when estimating delay distributions, assuming that double interval censoring is always adjusted for and that the forward distribution is being modeled directly (i.e., not via the backwards distribution and dynamical bias correction) as we recommend regardless of data collection approach. If you have an estimate of the backwards distribution from the literature, see the section on “Other considerations” for advice.



References

1. Home | NSF NEON | Open Data to Understand our Ecosystems. [cited 23 Aug 2024]. Available: <https://www.neonscience.org/>
2. EarthRanger: Protecting Wildlife With Real-Time Data. [cited 23 Aug 2024]. Available: <https://www.earthranger.com/>
3. Paz A, Silva TS, Carnaval AC. A framework for near-real time monitoring of diversity patterns based on indirect remote sensing, with an application in the Brazilian Atlantic rainforest. *PeerJ*. 2022;10: e13534. doi:10.7717/peerj.13534
4. Panigrahi S, Maski P, Thondiyath A. Real-time biodiversity analysis using deep-learning algorithms on mobile robotic platforms. *PeerJ Comput Sci*. 2023;9: e1502. doi:10.7717/peerj-cs.1502
5. Venverloo T, Duarte F. Towards real-time monitoring of insect species populations. *Sci Rep*. 2024;14: 18727. doi:10.1038/s41598-024-68502-8
6. Cornford R, Spooner F, McRae L, Purvis A, Freeman R. Ongoing over-exploitation and delayed responses to environmental change highlight the urgency for action to promote vertebrate recoveries by 2030. *Proc R Soc B Biol Sci*. 2023;290: 20230464. doi:10.1098/rspb.2023.0464
7. Charniga K, Masters NB, Slayton RB, Gosdin L, Minhaj FS, Philpott D, et al. Estimating the incubation period of monkeypox virus during the 2022 multi-national outbreak. *medRxiv*. 2022; 2022.06.22.22276713. doi:10.1101/2022.06.22.22276713
8. Lessler J, Reich NG, Brookmeyer R, Perl TM, Nelson KE, Cummings DA. Incubation periods of acute respiratory viral infections: a systematic review. *Lancet Infect Dis*. 2009;9: 291–300. doi:10.1016/S1473-3099(09)70069-6
9. Overton CE, Abbott S, Christie R, Cumming F, Day J, Jones O, et al. Nowcasting the 2022 mpox outbreak in England. Lau EHY, editor. *PLOS Comput Biol*. 2023;19: e1011463. doi:10.1371/journal.pcbi.1011463
10. Lavezzo E, Franchin E, Ciavarella C, Cuomo-Dannenburg G, Barzon L, Del Vecchio C, et al. Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo'. *Nature*. 2020;584: 425–429. doi:10.1038/s41586-020-2488-1
11. WHO Ebola Response Team. Ebola Virus Disease in West Africa — The First 9 Months of the Epidemic and Forward Projections. *N Engl J Med*. 2014;371: 1481–1495. doi:10.1056/NEJMoa1411100
12. Fraser C, Donnelly CA, Cauchemez S, Hanage WP, Van Kerkhove MD, Hollingsworth TD, et al. Pandemic Potential of a Strain of Influenza A (H1N1): Early Findings. *Science*. 2009;324: 1557–1561. doi:10.1126/science.1176062

13. Ferguson NM, Cucunubá ZM, Dorigatti I, Nedjati-Gilani GL, Donnelly CA, Basáñez M-G, et al. Countering the Zika epidemic in Latin America. *Science*. 2016;353: 353–354. doi:10.1126/science.aag0219
14. Ward T, Glaser A, Overton CE, Carpenter B, Gent N, Seale AC. Replacement dynamics and the pathogenesis of the Alpha, Delta and Omicron variants of SARS-CoV-2. *Epidemiol Infect*. 2023;151: e32. doi:10.1017/S0950268822001935
15. Reynolds MG, Yorita KL, Kuehnert MJ, Davidson WB, Huhn GD, Holman RC, et al. Clinical Manifestations of Human Monkeypox Influenced by Route of Infection. *J Infect Dis*. 2006;194: 773–780. doi:10.1086/505880
16. Galmiche S, Cortier T, Charmet T, Schaeffer L, Chény O, Von Platen C, et al. SARS-CoV-2 incubation period across variants of concern, individual factors, and circumstances of infection in France: a case series analysis from the ComCor study. *Lancet Microbe*. 2023;4: e409–e417. doi:10.1016/S2666-5247(23)00005-8
17. Cramer E, Ray E, Lopez V, Bracher J, Brennen A, Rivadeneira A, et al. Evaluation of individual and ensemble probabilistic forecasts of COVID-19 mortality in the United States. *PNAS*. 2022;119: e2113561119.
18. Borchering RK, Healy JM, Cadwell BL, Johansson MA, Slayton RB, Wallace M, et al. Public health impact of the U.S. Scenario Modeling Hub. *Epidemics*. 2023;44: 100705. doi:10.1016/j.epidem.2023.100705
19. Clay PA, Pollock ED, Saldarriaga EM, Pathela P, Macaraig M, Zucker JR, et al. Modeling the impact of prioritizing first or second vaccine doses during the 2022 mpox outbreak. *Infectious Diseases (except HIV/AIDS)*; 2023 Oct. doi:10.1101/2023.10.13.23297005
20. Nishiura H, Linton NM, Akhmetzhanov AR. Serial interval of novel coronavirus (COVID-19) infections. *Int J Infect Dis*. 2020;93: 284–286. doi:10.1016/j.ijid.2020.02.060
21. Liu Y, Centre for Mathematical Modelling of Infectious Diseases nCoV Working Group, Funk S, Flasche S. The contribution of pre-symptomatic infection to the transmission dynamics of COVID-2019 [version 1; peer review: 3 approved]. *Wellcome Open Res*. 2020;5: 58. doi:10.12688/wellcomeopenres.15788.1
22. Fraser C, Riley S, Anderson RM, Ferguson NM. Factors that make an infectious disease outbreak controllable. *Proc Natl Acad Sci*. 2004;101: 6146–6151. doi:10.1073/pnas.0307506101
23. Park SW, Akhmetzhanov AR, Charniga K, Cori A, Davies NG, Dushoff J, et al. Estimating epidemiological delay distributions for infectious diseases. *medRxiv*; 2024. doi:10.1101/2024.01.12.24301247
24. Ward T, Christie R, Paton RS, Cumming F, Overton CE. Transmission dynamics of monkeypox in the United Kingdom: contact tracing study. *BMJ*. 2022; e073153. doi:10.1136/bmj-2022-073153

25. Overton CE, Stage HB, Ahmad S, Curran-Sebastian J, Dark P, Das R, et al. Using statistics and mathematical modelling to understand infectious disease outbreaks: COVID-19 as an example. *Infect Dis Model.* 2020;5: 409–441. doi:10.1016/j.idm.2020.06.008
26. Britton T, Scalia Tomba G. Estimation in emerging epidemics: biases and remedies. *J R Soc Interface.* 2019;16: 20180670. doi:10.1098/rsif.2018.0670
27. Seaman SR, Presanis A, Jackson C. Estimating a time-to-event distribution from right-truncated data in an epidemic: A review of methods. *Stat Methods Med Res.* 2022;31: 1641–1655. doi:10.1177/09622802211023955
28. Gostic KM, McGough L, Baskerville EB, Abbott S, Joshi K, Tedijanto C, et al. Practical considerations for measuring the effective reproductive number, R_t . Pitzer VE, editor. *PLOS Comput Biol.* 2020;16: e1008409. doi:10.1371/journal.pcbi.1008409
29. Ali ST, Wang L, Lau EHY, Xu X-K, Du Z, Wu Y, et al. Serial interval of SARS-CoV-2 was shortened over time by nonpharmaceutical interventions. *Science.* 2020;369: 1106–1109. doi:10.1126/science.abc9004
30. Lambert JW, Kucharski A. epiparameter. 2023. Available: <https://github.com/epiverse-trace/epiparameter>
31. Nash R, Morgenstern C, Cuomo-Dannenburg G, McCabe R, McCain K. epireview. [cited 28 Dec 2023]. Available: <https://mrc-ide.github.io/epireview/>
32. WorldHealthOrganization/collaboratory-epiparameter-community. World Health Organization; 2023. Available: <https://github.com/WorldHealthOrganization/collaboratory-epiparameter-community>
33. McAloon C, Collins Á, Hunt K, Barber A, Byrne AW, Butler F, et al. Incubation period of COVID-19: a rapid systematic review and meta-analysis of observational research. *BMJ Open.* 2020;10: e039652. doi:10.1136/bmjopen-2020-039652
34. Cuomo-Dannenburg G, McCain K, McCabe R, Unwin HJT, Doohan P, Nash RK, et al. Marburg virus disease outbreaks, mathematical models, and disease parameters: a systematic review. *Lancet Infect Dis.* 2023; S1473309923005157. doi:10.1016/S1473-3099(23)00515-7
35. Doohan P, Jorgensen D, Naidoo TM, McCain K, Hicks JT, McCabe R, et al. Lassa fever outbreaks, mathematical models, and disease parameters: a systematic review and meta-analysis. *medRxiv*; 2024. p. 2024.03.23.24304596. doi:10.1101/2024.03.23.24304596
36. Madewell ZJ, Charniga K, Masters NB, Asher J, Fahrenwald L, Still W, et al. Serial Interval and Incubation Period Estimates of Monkeypox Virus Infection in 12 Jurisdictions, United States, May–August 2022. *Emerg Infect Dis.* 2023;29: 818–821. doi:10.3201/eid2904.221622
37. Humphrey JH, Marinda E, Mutasa K, Moulton LH, Iliff PJ, Ntozini R, et al. Mother to child transmission of HIV among Zimbabwean women who seroconverted postnatally: prospective cohort study. *BMJ.* 2010;341: c6580–c6580. doi:10.1136/bmj.c6580

38. Antonucci G, Girardi E, Raviglione M, Ippolito G, Almi P, Angarano G, et al. Risk Factors for Tuberculosis in HIV-Infected Persons: A Prospective Cohort Study. *JAMA*. 1995;274: 143–148. doi:10.1001/jama.1995.03530020061033
39. Munywoki PK, Koech DC, Agoti CN, Lewa C, Cane PA, Medley GF, et al. The Source of Respiratory Syncytial Virus Infection In Infants: A Household Cohort Study In Rural Kenya. *J Infect Dis*. 2014;209: 1685–1692. doi:10.1093/infdis/jit828
40. Levy JW, Cowling BJ, Simmerman JM, Olsen SJ, Fang VJ, Suntarattiwong P, et al. The Serial Intervals of Seasonal and Pandemic Influenza Viruses in Households in Bangkok, Thailand. *Am J Epidemiol*. 2013;177: 1443–1451. doi:10.1093/aje/kws402
41. Russell TW, Hellewell J, Jarvis CI, Van Zandvoort K, Abbott S, Ratnayake R, et al. Estimating the infection and case fatality ratio for coronavirus disease (COVID-19) using age-adjusted data from the outbreak on the Diamond Princess cruise ship, February 2020. *Eurosurveillance*. 2020;25. doi:10.2807/1560-7917.ES.2020.25.12.2000256
42. Centers for Disease Control and Prevention. Introduction to Public Health. In: *Public Health 101 Series*. 2014 [cited 9 Jan 2024]. Available: <https://www.cdc.gov/training/publichealth101/surveillance.html>
43. Sharpe JD, Charniga K, Byrd KM, Stefanos R, Lewis L, Watson J, et al. Possible Exposures Among Mpox Patients Without Reported Male-to-Male Sexual Contact — Six U.S. Jurisdictions, November 1–December 14, 2022. *MMWR Morb Mortal Wkly Rep*. 2023;72: 944–948. doi:10.15585/mmwr.mm7235a2
44. Reich NG, Lessler J, Cummings DAT, Brookmeyer R. Estimating incubation period distributions with coarse data. *Stat Med*. 2009;28: 2769–2784. doi:10.1002/sim.3659
45. Dey T, Lipsitz SR, Cooper Z, Trinh Q-D, Krzywinski M, Altman N. Survival analysis—time-to-event data and censoring. *Nat Methods*. 2022;19: 906–908. doi:10.1038/s41592-022-01563-7
46. Park SW, Sun K, Champredon D, Li M, Bolker BM, Earn DJD, et al. Forward-looking serial intervals correctly link epidemic growth to reproduction numbers. *Proc Natl Acad Sci*. 2021;118: e2011548118. doi:10.1073/pnas.2011548118
47. Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infect Dis*. 2020;20: 669–677. doi:10.1016/S1473-3099(20)30243-7
48. Reich NG, Lessler J, Azman A, Kamvar ZN. coarseDataTools: Analysis of Coarsely Observed Data. 2021. Available: <https://cran.r-project.org/web/packages/coarseDataTools/index.html>
49. Park SW, Abbott S. Estimating delay distributions accounting for common biases. 2023. Available: <https://github.com/parksw3/epidist-paper>
50. Leung K-M, Elashoff RM, Afifi AA. CENSORING ISSUES IN SURVIVAL ANALYSIS. *Annu Rev Public Health*. 1997;18: 83–104. doi:10.1146/annurev.publhealth.18.1.83

51. Nolen LD, Osadebe L, Katomba J, Likofata J, Mukadi D, Monroe B, et al. Extended Human-to-Human Transmission during a Monkeypox Outbreak in the Democratic Republic of the Congo. *Emerg Infect Dis.* 2016;22: 1014–1021. doi:10.3201/eid2206.150579
52. Mertens PE, Patton R, Baum JJ, Monath TP. Clinical presentation of lassa fever cases during the hospital epidemic at Zorzor, Liberia, March-April 1972. *Am J Trop Med Hyg.* 1973;22: 780–784. doi:10.4269/ajtmh.1973.22.780
53. Martini G. Marburg virus disease. *Postgrad Med J.* 1973;49: 542–546.
54. WHO International Study Team. Ebola haemorrhagic fever in Sudan, 1976. *Bull World Health Organ.* 1978;56: 247–270.
55. Nishiura H. Early efforts in modeling the incubation period of infectious diseases with an acute course of illness. *Emerg Themes Epidemiol.* 2007;4: 2. doi:10.1186/1742-7622-4-2
56. Miura F, Van Ewijk CE, Backer JA, Xiridou M, Franz E, Op De Coul E, et al. Estimated incubation period for monkeypox cases confirmed in the Netherlands, May 2022. *Eurosurveillance.* 2022;27. doi:10.2807/1560-7917.ES.2022.27.24.2200448
57. Ferretti L, Ledda A, Wymant C, Zhao L, Ledda V, Abeler-Dörner L, et al. The timing of COVID-19 transmission. *Epidemiology;* 2020 Sep. doi:10.1101/2020.09.04.20188516
58. McAloon CG, Wall P, Griffin J, Casey M, Barber A, Codd M, et al. Estimation of the serial interval and proportion of pre-symptomatic transmission events of COVID- 19 in Ireland using contact tracing data. *BMC Public Health.* 2021;21: 805. doi:10.1186/s12889-021-10868-9
59. Akhmetzhanov AR, Cheng H-Y, Dushoff J. Incubation-period estimates of Omicron (BA.1) variant from Taiwan, December 2021–January 2022, and its comparison to other SARS-CoV-2 variants: a statistical modeling, systematic search and meta-analysis. *Infectious Diseases (except HIV/AIDS);* 2023 Jul. doi:10.1101/2023.07.20.23292983
60. Lover AA, Zhao X, Gao Z, Coker RJ, Cook AR. The distribution of incubation and relapse times in experimental human infections with the malaria parasite *Plasmodium vivax*. *BMC Infect Dis.* 2014;14: 539. doi:10.1186/1471-2334-14-539
61. Klinkenberg D, Nishiura H. The correlation between infectivity and incubation period of measles, estimated from households with two cases. *J Theor Biol.* 2011;284: 52–60. doi:10.1016/j.jtbi.2011.06.015
62. Hoffmann T, Alsing J. Faecal shedding models for SARS-CoV-2 RNA among hospitalised patients and implications for wastewater-based epidemiology. *J R Stat Soc Ser C Appl Stat.* 2023;72: 330–345. doi:10.1093/jrsssc/qlad011
63. Abbott S, Lison A, Funk S, Choi-Hannah, Gruson H, Pearson CAB, et al. *epinowcast/epinowcast: Epinowcast 0.2.2.* Zenodo; 2023. doi:10.5281/ZENODO.5637165
64. Sender R, Bar-On Y, Park SW, Noor E, Dushoff J, Milo R. The unmitigated profile of COVID-19 infectiousness. *eLife.* 2022;11: e79134. doi:10.7554/eLife.79134

65. Gamma distribution. Wikipedia. 2024. Available: https://en.wikipedia.org/w/index.php?title=Gamma_distribution&oldid=1210997995
66. Log-normal distribution. Wikipedia. 2024. Available: https://en.wikipedia.org/w/index.php?title=Log-normal_distribution&oldid=1210518171
67. Abbott S, Hellewell J, Sherratt K, Gostic K, Hickson J, Badr HS, et al. EpiNow2: Estimate Real-Time Case Counts and Time-Varying Epidemiological Parameters. 2023. Available: <https://cran.r-project.org/web/packages/EpiNow2/index.html>
68. Jombart T, Cori A, Kamvar ZN, Schumacher D, Finger F, Whittaker C. epitrix: Small Helpers and Tricks for Epidemics Analysis. 2023. Available: <https://cran.r-project.org/web/packages/epitrix/index.html>
69. Yu Y. mixR: Finite Mixture Modeling for Raw and Binned Data. 2021. Available: <https://cran.r-project.org/web/packages/mixR/index.html>
70. Van Kerkhove MD, Bento AI, Mills HL, Ferguson NM, Donnelly CA. A review of epidemiological parameters from Ebola outbreaks to inform early public health decision-making. *Sci Data*. 2015;2: 150019. doi:10.1038/sdata.2015.19
71. Medley GF, Anderson RM, Cox DR, Billard L. Incubation period of AIDS in patients infected via blood transfusion. *Nature*. 1987;328: 719–721. doi:10.1038/328719a0
72. Lison A, Abbott S, Huisman J, Stadler T. Generative Bayesian modeling to nowcast the effective reproduction number from line list data with missing symptom onset dates. 2023 [cited 21 Feb 2024]. doi:10.48550/ARXIV.2308.13262
73. Hamra G, MacLehose R, Richardson D. Markov Chain Monte Carlo: an introduction for epidemiologists. *Int J Epidemiol*. 2013;42: 627–634. doi:10.1093/ije/dyt043
74. Lambert B. A student's guide to Bayesian statistics. Los Angeles London New Delhi Singapore Washington DC Melbourne: SAGE; 2018.
75. RStan: the R interface to Stan. [cited 10 Apr 2024]. Available: <https://mc-stan.org/rstan/articles/rstan.html>
76. Stan Development Team. Stan Modeling Language Users Guide and Reference Manual Version 2.34. 2022. Available: <https://mc-stan.org>
77. Elderd BD, Dukic VM, Dwyer G. Uncertainty in predictions of disease spread and public health responses to bioterrorism and emerging diseases. *Proc Natl Acad Sci*. 2006;103: 15693–15697. doi:10.1073/pnas.0600816103
78. Pollett S, Johansson MA, Reich NG, Brett-Major D, Del Valle SY, Venkatramanan S, et al. Recommended reporting items for epidemic forecasting and prediction research: The EPIFORGE 2020 guidelines. *PLOS Med*. 2021;18: e1003793. doi:10.1371/journal.pmed.1003793

79. Cowling BJ, Park M, Fang VJ, Wu P, Leung GM, Wu JT. Preliminary epidemiological assessment of MERS-CoV outbreak in South Korea, May to June 2015. *Eurosurveillance*. 2015;20. doi:10.2807/1560-7917.ES2015.20.25.21163
80. Cheng H-Y, Akhmetzhanov AR, Dushoff J. SARS-CoV-2 Incubation Period during Omicron BA.5–Dominant Period, Japan. *Emerg Infect Dis*. 2024;30. doi:10.3201/eid3001.230208
81. Miura F, Backer JA, Van Rijckevorsel G, Bavalia R, Raven S, Petrignani M, et al. Time Scales of Human Mpox Transmission in The Netherlands. *J Infect Dis*. 2023; jiad091. doi:10.1093/infdis/jiad091
82. Campbell F, Didelot X, Fitzjohn R, Ferguson N, Cori A, Jombart T. outbreaker2: a modular platform for outbreak reconstruction. *BMC Bioinformatics*. 2018;19: 363. doi:10.1186/s12859-018-2330-z
83. Hall M, Woolhouse M, Rambaut A. Epidemic Reconstruction in a Phylogenetics Framework: Transmission Trees as Partitions of the Node Set. *PLOS Comput Biol*. 2015;11: e1004613. doi:10.1371/journal.pcbi.1004613
84. Forsberg White L, Pagano M. A likelihood-based method for real-time estimation of the serial interval and reproductive number of an epidemic. *Stat Med*. 2008;27: 2999–3016. doi:10.1002/sim.3136
85. Wu KM, Riley S. Estimation of the Basic Reproductive Number and Mean Serial Interval of a Novel Pathogen in a Small, Well-Observed Discrete Population. *PLOS ONE*. 2016;11: e0148061. doi:10.1371/journal.pone.0148061
86. Madewell ZJ, Yang Y, Longini IM, Halloran ME, Vespignani A, Dean NE. Rapid review and meta-analysis of serial intervals for SARS-CoV-2 Delta and Omicron variants. *BMC Infect Dis*. 2023;23: 429. doi:10.1186/s12879-023-08407-5
87. Stan Development Team. Prior Choice Recommendations. In: GitHub [Internet]. [cited 16 Apr 2024]. Available: <https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations>
88. Rudolph KE, Lessler J, Moloney RM, Kmush B, Cummings DAT. Incubation Periods of Mosquito-Borne Viral Infections: A Systematic Review. *Am J Trop Med Hyg*. 2014;90: 882–891. doi:10.4269/ajtmh.13-0403
89. Park SW, Sun K, Abbott S, Sender R, Bar-on YM, Weitz JS, et al. Inferring the differences in incubation-period and generation-interval distributions of the Delta and Omicron variants of SARS-CoV-2. *Proc Natl Acad Sci*. 2023;120: e2221887120. doi:10.1073/pnas.2221887120
90. Fields VL, Kracalik IT, Carthel C, Lopez A, Schwartz A, Lewis NM, et al. Coronavirus Disease Contact Tracing Outcomes and Cost, Salt Lake County, Utah, USA, March–May 2020. *Emerg Infect Dis*. 2021;27: 2999–3008. doi:10.3201/eid2712.210505
91. Wu Zeng, Hadia Samaha, Michel Yao, Steve Ahuka-Mundeke, Thomas Wilkinson, Thibaut Jombart, et al. The cost of public health interventions to respond to the 10th Ebola outbreak in the Democratic Republic of the Congo. *BMJ Glob Health*. 2023;8: e012660. doi:10.1136/bmjgh-2023-012660

92. Höhle M, An Der Heiden M. Bayesian nowcasting during the STEC O104:H4 outbreak in Germany, 2011. *Biometrics*. 2014;70: 993–1002. doi:10.1111/biom.12194
93. Van De Kasstelee J, Eilers PHC, Wallinga J. Nowcasting the Number of New Symptomatic Cases During Infectious Disease Outbreaks Using Constrained P-spline Smoothing. *Epidemiology*. 2019;30: 737–745. doi:10.1097/EDE.0000000000001050
94. Bastos LS, Economou T, Gomes MFC, Villela DAM, Coelho FC, Cruz OG, et al. A modelling approach for correcting reporting delays in disease surveillance data. *Stat Med*. 2019;38: 4363–4377. doi:10.1002/sim.8303
95. Chan M, Johansson MA. The Incubation Periods of Dengue Viruses. Vasilakis N, editor. *PLoS ONE*. 2012;7: e50972. doi:10.1371/journal.pone.0050972
96. Lessler J, Ott CT, Carcelen AC, Konikoff JM, Williamson J, Bi Q, et al. Times to key events in Zika virus infection and implications for blood donation: a systematic review. *Bull World Health Organ*. 2016;94: 841–849. doi:10.2471/BLT.16.174540
97. Linton N, Kobayashi T, Yang Y, Hayashi K, Akhmetzhanov A, Jung S, et al. Incubation Period and Other Epidemiological Characteristics of 2019 Novel Coronavirus Infections with Right Truncation: A Statistical Analysis of Publicly Available Case Data. *J Clin Med*. 2020;9: 538. doi:10.3390/jcm9020538
98. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Ann Intern Med*. 2020;172: 577–582. doi:10.7326/M20-0504
99. Plummer M, Best N, Cowles K, Vines K. CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News*. 6: 7–11.
100. Gabry J, Mahr T. bayesplot: Plotting for Bayesian Models. Available: bayesplot: Plotting for Bayesian Models
101. Guzzetta G, Mammone A, Ferraro F, Caraglia A, Rapiti A, Marziano V, et al. Early Estimates of Monkeypox Incubation Period, Generation Time, and Reproduction Number, Italy, May–June 2022. *Emerg Infect Dis*. 2022;28: 2078–2081. doi:10.3201/eid2810.221126
102. Assiri A, McGeer A, Perl TM, Price CS, Al Rabeeah AA, Cummings DAT, et al. Hospital Outbreak of Middle East Respiratory Syndrome Coronavirus. *N Engl J Med*. 2013;369: 407–416. doi:10.1056/NEJMoa1306742
103. McFarland SE, Marcus U, Hemmers L, Miura F, Iñigo Martínez J, Martínez FM, et al. Estimated incubation period distributions of mpox using cases from two international European festivals and outbreaks in a club in Berlin, May to June 2022. *Eurosurveillance*. 2023;28. doi:10.2807/1560-7917.ES.2023.28.27.2200806