

1 Estimating epidemiological delay distributions for 2 infectious diseases

3 Sang Woo Park¹, Andrei R. Akhmetzhanov², Kelly Charniga³, Anne
4 Cori⁴, Nicholas G. Davies^{5, 6}, Jonathan Dushoff^{7, 8, 9}, Sebastian
5 Funk^{5,6}, Katie Gostic¹⁰, Bryan Grenfell¹, Natalie M. Linton¹¹, Marc
6 Lipsitch^{10,12}, Adrian Lison¹³, Christopher E. Overton^{14,15,16}, Thomas
7 Ward¹⁵, and Sam Abbott^{6,7}

8 ¹*Department of Ecology and Evolutionary Biology, Princeton, NJ, USA*

9 ²*College of Public Health, National Taiwan University, 17 Xu-Zhou Road, Taipei 10055, Taiwan*

10 ³*Division of High-Consequence Pathogens and Pathology, National Center for Emerging & Zoonotic
11 Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, GA, USA*

12 ⁴*MRC Centre for Global Infectious Disease Analysis, School of Public Health, Imperial College London,
13 UK*

14 ⁵*Department of Infectious Disease Epidemiology, London School of Hygiene & Tropical Medicine, London,
15 UK*

16 ⁶*Centre for Mathematical Modelling of Infectious Diseases, London School of Hygiene & Tropical
17 Medicine, London, UK*

18 ⁷*Departments of Mathematics & Statistics and Biology, McMaster University, Hamilton, Ontario, Canada*

19 ⁸*Department of Biology, McMaster University, Hamilton, L8S 4L8 ON, Canada*

20 ⁹*M. G. DeGroot Institute for Infectious Disease Research, McMaster University, Hamilton, L8S 4L8 ON,
21 Canada*

22 ¹⁰*Center for Forecasting and Outbreak Analytics, US Centers for Disease Control and Prevention (CDC),
23 Atlanta, GA, USA*

24 ¹¹*Graduate School of Medicine, Hokkaido University, Kita 15 Jo Nishi 7 Chome, Kita-ku, Sapporo-shi,
25 Hokkaido 060-8638, Japan*

26 ¹²*Center for Communicable Disease Dynamics, Department of Epidemiology, Harvard T.H. Chan School
27 of Public Health, Boston, MA, USA*

28 ¹³*ETH Zurich, Department of Biosystems Science and Engineering, Zurich, Switzerland*

29 ¹⁴*Department of Mathematical Sciences, University of Liverpool, UK*

30 ¹⁵*All Hazards Intelligence, Data Analytics and Surveillance, UK Health Security Agency, UK*

31 ¹⁶*Department of Mathematics, University of Manchester, UK*

32 January 13, 2024

33 Disclaimer: The findings and conclusions in this report are those of the authors and
34 do not necessarily represent the official position of the CDC, U.S. Department of
35 Health and Human Services.

Abstract

36
37 Understanding and accurately estimating epidemiological delay distribu-
38 tions is important for public health policy. These estimates directly influence
39 epidemic situational awareness, control strategies, and resource allocation. In
40 this study, we explore challenges in estimating these distributions, including
41 truncation, interval censoring, and dynamical biases. Despite their impor-
42 tance, these issues are frequently overlooked in the current literature, often
43 resulting in biased conclusions. This study aims to shed light on these chal-
44 lenges, providing valuable insights for epidemiologists and infectious disease
45 modellers.

46 Our work motivates comprehensive approaches for accounting for these is-
47 sues based on the underlying theoretical concepts. We also discuss simpler
48 methods that are widely used, which do not fully account for known biases.
49 We evaluate the statistical performance of these methods using simulated ex-
50ponential growth and epidemic scenarios informed by data from the 2014-2016
51 Sierra Leone Ebola virus disease epidemic.

52 Our findings highlight that using simpler methods can lead to biased es-
53 timates of vital epidemiological parameters. An approximate-latent-variable
54 method emerges as the best overall performer, while an efficient, widely im-
55 plemented interval-reduced-censoring-and-truncation method was only slightly
56 worse. Other methods, such as a joint-primary-incidence-and-delay method
57 and a dynamic-correction method, demonstrated good performance under cer-
58 tain conditions, although they have inherent limitations and may not be the
59 best choice for more complex problems.

60 Despite presenting a range of methods that performed well in the contexts
61 we evaluated, residual biases persisted, predominantly due to the simplifying
62 assumption that the distribution of event time within the censoring inter-
63 val follows a uniform distribution; instead, this distribution should depend
64 on epidemic dynamics. However, in realistic scenarios with daily censoring,
65 these biases appeared minimal. This study underscores the need for caution
66 when estimating epidemiological delay distributions in real-time, provides an
67 overview of the theory that practitioners need to keep in mind when doing so
68 with useful tools to avoid common methodological errors, and points towards
69 areas for future research.

70	Contents	
71	Summary	4
72	1 Introduction	8
73	2 Methods	11
74	2.1 General theory for measuring epidemiological delay distributions . . .	11
75	2.2 Biases in estimating forward distributions	16
76	2.3 Methods for estimating forward delay distributions from observed data	20
77	2.3.1 Simplifying assumptions for the following methods	20
78	2.3.2 Exact censoring and truncation method	21
79	2.3.3 Approximate latent variable censoring and truncation method	22
80	2.3.4 Interval-reduced censoring and truncation methods	22
81	2.3.5 Interval-reduced censoring method	24
82	2.3.6 Truncation method	24
83	2.3.7 Filtering method	25
84	2.3.8 Naive method	25
85	2.3.9 Discrete time methods for accounting for censoring and right	
86	truncation	26
87	2.3.10 Overview table	29
88	2.4 Simulation study	33
89	2.5 Case study: 2014-2016 Sierra Leone Ebola virus disease epidemic . .	34
90	2.6 Evaluation	35
91	2.7 Implementation	36
92	3 Results	37
93	3.1 Simulation study	37
94	3.1.1 Exponential growth simulation	37
95	3.1.2 Epidemic wave simulation	41
96	3.2 Case study: 2014-2016 Sierra Leone Ebola virus disease epidemic . .	44
97	3.2.1 Empirical observations	44
98	3.2.2 Model results	47
99	3.2.3 Implementation considerations	49
100	4 Discussion	51
101	4.1 Findings	51
102	4.2 Limitations	52
103	4.3 Generalisability	53
104	4.4 Conclusions	55

105 Summary

106 What was known prior to this paper

- 107 • **Importance of accurate estimates:** Estimating epidemiological delay dis-
108 tributions accurately is critical for model development, epidemic forecasts, and
109 analytic decision support.

- 110 • **Right truncation:** Right truncation describes the incomplete observation of
111 delays, for which the primary event already occurred but the secondary event
112 has not been observed (e.g. infections that have not yet become symptomatic
113 and therefore not been observed). Failing to account for the right truncation
114 can lead to underestimation of the mean delay during real-time data analysis.

- 115 • **Interval censoring:** Interval censoring arises when epidemiological events
116 occurring in continuous time are binned into time intervals (e.g., days or weeks).
117 Double censoring of both primary and secondary events needs to be considered
118 when estimating delay distributions from epidemiological data. Accounting for
119 censoring in only one event can lead to additional biases.

- 120 • **Dynamical bias:** Dynamical biases describe the effects of an epidemic's cur-
121 rent growth or decay rate on the observed delay distributions. Consider an anal-
122 ogy from demography: a growing population will contain an excess of young
123 people, while a shrinking population will contain an excess of older people,
124 compared to what would be expected from mortality profiles alone. Dynami-
125 cal biases have been identified as significant issues in real-time epidemiological
126 studies.

- 127 • **Existing methods:** Methods and software to adjust for censoring, truncation,
128 and dynamic biases exist. However, many of these methods have not been
129 systematically compared, validated, or tested outside the context in which they
130 were originally developed. Furthermore, some of these methods do not adjust
131 for the full range of biases.

132 What this paper adds

- 133 • **Theory overview:** An overview of the theory required to estimate distribu-
134 tions is provided, helping practitioners understand the underlying principles
135 of the methods and the connections between right truncation, dynamical bias,
136 and interval censoring.

- 137 • **Review of methods:** This paper presents a review of methods accounting
138 for truncation, interval censoring, and dynamical biases in estimating epidemi-
139 ological delay distributions in the context of the underlying theory.

- 140 • **Evaluation of methods:** Methods were evaluated using simulations as well
141 as data from the 2014-2016 Sierra Leone Ebola virus disease epidemic.
- 142 • **Cautionary guidance:** This work underscores the need for caution when
143 estimating epidemiological delay distributions, provides clear signposting for
144 which methods to use when, and points out areas for future research.
- 145 • **Practical guidance:** Guidance is also provided for those making use of delay
146 distributions in routine practice.

147 Key findings

- 148 • **Impact of neglecting biases:** Neglecting truncation and censoring biases can
149 lead to flawed estimates of important epidemiological parameters, especially in
150 real-time epidemic settings.
- 151 • **Equivalence of dynamical bias and right truncation:** In the context
152 of a growing epidemic, right truncation has an essentially equivalent effect as
153 dynamical bias. Typically, we recommend correcting for one or the other, but
154 not both.
- 155 • **Bias in common censoring adjustment:** Taking the common approach to
156 censoring adjustment of naively discretising observed delay into daily intervals
157 and fitting continuous-time distributions can result in biased estimates.
- 158 • **Performance of methods:** We identified an approximate-latent-variable
159 method as the best overall performer, while an interval-reduced-censoring-and-
160 truncation method was resource-efficient, widely implemented, and performed
161 only slightly worse.
- 162 • **Inherent limitations of some methods:** Other methods, such as jointly
163 estimating primary incidence and the forward delay, and dynamic bias correc-
164 tion, demonstrated good performance under certain conditions, but they also
165 had inherent limitations depending on the setting.
- 166 • **Persistence of residual biases:** Residual biases persisted across all methods
167 we investigated, largely due to the simplifying assumption that the distribution
168 of event time within the primary censoring interval follows a uniform distribu-
169 tion rather than one influenced by the growth rate. These are minimal if the
170 censoring interval is small compared to other relevant time scales, as is the case
171 for daily censoring with most human diseases.

172 Key limitations

- 173 • **Differences between right censoring and truncation:** We primarily
174 focus on right truncation, which is most relevant when the secondary events
175 are easier to observe than primary events (e.g., symptom onset vs. infection)—
176 in this case, we can't observe the delay until the secondary event has occurred.
177 In other cases, we can directly observe the primary event and wait for the
178 secondary event to occur (e.g., eventual recovery or death of a hospitalized
179 individual)—in this case, it would be more appropriate to use right censoring
180 to model the unresolved delays. For simplicity, we did not cover the right
181 censoring in this paper.
- 182 • **Daily censoring process:** Our work considered only a daily interval censoring
183 process for primary and secondary events. To mitigate this, we investigated
184 scenarios with short delays and high growth rates, mimicking longer censoring
185 intervals with extended delays and slower growth rates.
- 186 • **Deviation from uniform distribution assumption:** We show that the
187 empirical distribution of event times within the primary censoring interval de-
188 viated from the common assumption of a uniform distribution due to epidemic
189 dynamics. This discrepancy introduced a small absolute bias based on the
190 length of the primary censoring window to all methods and was a particular
191 issue when delay distributions were short relative to the censoring window's
192 length. In practice, other biological factors, such as circadian rhythms, are
193 likely to have a stronger effect than the growth rate at a daily resolution.
194 Nonetheless, our work lays out a theoretical ground for linking epidemic dy-
195 namics to a censoring process. Further work is needed to develop robust meth-
196 ods for wider censoring intervals.
- 197 • **Temporal changes in delay distributions:** The Ebola case study show-
198 cased considerable variation in reporting delays across the epidemic timeline,
199 far greater than any bias due to censoring or truncation. Further work is needed
200 to extend our methods to address such issues.
- 201 • **Lack of other bias consideration:** The idealized simulated scenarios we
202 used did not account for observation error for either primary or secondary
203 events, possibly favouring methods that do not account for real-world sources
204 of biases.
- 205 • **Limited distributions and methods considered:** We only considered log-
206 normal distributions in this study, though our findings are generalizable to
207 other distributions. Mixture distributions and non-parametric or hazard-based
208 methods were not included in our assessment.
- 209 • **Exclusion of fitting discrete-time distributions:** We focused on fitting
210 continuous-time distributions throughout the paper. However, fitting discrete-
211 time distributions can be a viable option in practice, especially at a daily

212 resolution. More work is needed to compare inferences based on discrete-time
213 distributions vs continuous-time distributions with daily censoring.

- 214 • **Exclusion of transmission interval distributions:** Our work primarily
215 focused on inferring distributions of non-transmission intervals, leaving out
216 potential complications related to dependent events. Additional considerations
217 such as shared source cases, identifying intermediate hosts, and the possibility
218 of multiple source cases for a single infectee were not factored into our analysis.

219 1 Introduction

220 Characterizing the distribution of time between two epidemiological events is essen-
221 tial to understanding the course of an infectious disease epidemic and making clinical
222 and public health decisions. For example, some intervals, such as the incubation pe-
223 riod (i.e., the time between infection and symptom onset), provide useful means of
224 summarizing aspects of the course of infection of each infected person (Lauer et al.,
225 2020; Linton et al., 2020; Verity et al., 2020a). Other intervals, such as the generation
226 interval (i.e., the time between infection and transmission) and serial interval (i.e.,
227 the time between symptom onsets in a transmission pair), describe the transmission
228 pattern across multiple individuals (Madewell et al., 2023) and can provide informa-
229 tion about how infectiousness varies over the course of an infection (Sender et al.,
230 2022). Combining these intervals can further help determine the controllability of
231 epidemics (e.g., potential for pre-symptomatic or asymptomatic transmission (Fraser
232 et al., 2004)) and inform guidelines for intervention measures, such as isolation of
233 cases (Hellewell et al., 2020), and travel screening (Gostic et al., 2020). Other in-
234 tervals, such as reporting delays (e.g., the time from symptom onset until a case is
235 reported), are needed for real-time data interpretation, analysis, and for epidemic
236 forecasting (Marinović et al., 2015; Overton et al., 2022; Abbott et al., 2020; Beesley
237 et al., 2022). Biases in delay distribution estimates can therefore translate to biases
238 in the chain of evidence used to inform decisions (Lipsitch et al., 2020).

239 There are multiple sources of bias that can affect the estimation of epidemiological
240 delay distributions. Some of these pertain to data reliability issues, such as recall
241 bias, whereas others are intrinsically linked to the structure of the data collection
242 process. In this paper, we primarily focus on the latter type of bias.

243 First, event observations are very often *censored*, meaning that we don't know
244 when the event happened exactly but we do know it occurred. Instead, epidemiolog-
245 ical events are typically reported using an interval (e.g. a date, a week, or a range
246 of dates) rather than the exact time at which an individual experienced an event
247 (Lindsey and Ryan, 1998). This is known as *interval censoring* as we only know
248 the interval in which the event occurred. Only on rare occasions, the time of the
249 event known is known more precisely (i.e. to the hour, minute, or very rarely to the
250 second), such as the time of death recorded on a death certificate, but even in these
251 cases uncertainty often remains.

252 Interval censoring can be particularly problematic if the reporting interval is
253 relatively wide compared to the typical length of a delay. This is a common problem
254 for short delays: for example, influenza has short generation intervals (2–3 days) and
255 incubation periods (1–2 days) (Fraser et al., 2009), meaning that even daily censoring
256 is expected to be problematic. Even when the delays are extremely long (e.g., the
257 incubation periods for TB or HIV/AIDS), interval censoring can be problematic
258 because the widths of censoring intervals are also just as wide.

259 Even when interval censoring is taken into account, many studies only adjust
260 for the censoring of a single event, rather than both primary and secondary events.

261 Examples of this include accounting for censoring of the date of exposure but not the
262 date of symptom onset when estimating the incubation period (Backer et al., 2020,
263 2022). In addition to interval censoring, data can be either left- or right-censored,
264 where only the upper or lower bounds, respectively, of the event times are known.
265 In the case of left censoring, we can always put a realistic lower bound, such as the
266 beginning of an epidemic, meaning that standard methods for interval censoring can
267 be readily applied. On the other hand, right censoring typically corresponds to the
268 case where we have already observed the primary event but failed to observe the
269 secondary event, typically due to a dropout of a patient from a study and failure
270 to follow up. Multiple methods have been developed for modelling right-censoring
271 (Ghani et al., 2005).

272 Second, epidemic data often suffer from *right truncation*, meaning that we only
273 observe events that have already happened and been reported (Brookmeyer and
274 Damiano, 1989; Kalbfleisch and Lawless, 1989; Gelman et al., 2013). In contrast
275 to right censoring, for which we have a partial observation of the lower bound of
276 the secondary event, we do not have any information in the case of truncation. For
277 example, incubation periods are often truncated because we are unable to observe
278 the primary event (infection) directly until the secondary event (symptom onset) is
279 reported. Failure to correct for right truncation can bias the data toward observation
280 of shorter intervals (e.g. if only individuals with symptoms that resolved before the
281 end of a study are included in the analysis). Moreover, only a limited number of
282 studies consider the interaction between censoring and truncation (Linton et al.,
283 2020; Ward and Johnsen, 2021).

284 Finally, recent studies have highlighted the role of dynamical biases: during the
285 growth phase of an epidemic, we are more likely to observe shorter delays because
286 a disproportionately large number of individuals have been infected more recently;
287 this effect is reversed during epidemic decay (Britton and Scalia Tomba, 2019; Park
288 et al., 2022). The effects of dynamical biases on the observed delay distributions are
289 quantitatively equivalent to truncation biases during the exponential growth phase,
290 but their equivalence has not always been clear, leading to attempts to address both
291 biases simultaneously (Linton et al., 2020; Guo et al., 2023b; Verity et al., 2020c).
292 These approaches further highlight that clearer guidance, and more robust methods
293 that take account of this guidance, are needed to handle biases found in different
294 epidemiological contexts, which can depend on the data-collection method, type
295 of data (e.g., single-individual delays such as incubation periods vs pair-dependent
296 delays such as generation or serial interval), and underlying epidemic dynamics.

297 Several methods and software packages accounting for censoring and truncation
298 corrections in epidemiological data already exist. The recent COVID-19 and mpox
299 epidemics have seen a marked increase in the development of both methods and
300 software implementations (Backer et al., 2020; Linton et al., 2020; Tindale et al.,
301 2020; Verity et al., 2020a; Hart et al., 2021). However, many of these methods were
302 developed for a particular context, have not been validated sufficiently, or do not
303 cover the full range of potential biases. Developed in 2009, `CoarseDataTools` is

304 widely used (Thompson et al., 2019; Madewell et al., 2023). It provides methods
305 that can account for double censoring (censoring of both primary and secondary
306 events) (Reich et al., 2009, 2010). However, it does not account for truncation bias,
307 and was not implemented to be readily extensible. Part of its popularity is driven by
308 its implementation in the popular effective-reproduction-number-estimation package
309 **EpiEstim** for serial-interval estimation (Cori et al., 2013; Thompson et al., 2019).
310 Unfortunately, naively relying on this implementation may lead to biased effective
311 reproduction number estimates.

312 Other widely used examples include the method of Backer et al. (2020), which
313 allows for uniform censoring in the primary event using latent variables but does not
314 account for right truncation. Similarly, the method of Linton et al. (2020) allows
315 for double censoring as well as right truncation adjustment. The method of Ward
316 et al. (2022) also accounts for double censoring and right truncation. However, so
317 far, these approaches have not been validated against simulations.

318 Tools exist for analyzing non-domain-specific interval-censored data (see Pan
319 et al. (2020) for a detailed review and comparison), but only a few of them ac-
320 count for right truncation and these methods are rarely used within the infectious
321 disease modelling community. The lack of ready-to-use software implementations, or
322 standardised methods, that can adjust for both censoring and right truncation means
323 that many of the estimated delay distributions present in the literature are likely bi-
324 ased. An additional issue is the predominance of methods developed specifically for
325 the application in which they were used (Backer et al., 2020; Linton et al., 2020;
326 Guo et al., 2023b), which means that even when known biases appear to have been
327 accounted for this can often be hard to verify due to the lack of robust evaluation.
328 Estimates from early in epidemics are likely particularly biased due to unaccounted-
329 for right-truncation, while retrospective estimates, though likely less biased, may still
330 be problematic when censoring is not properly taken care of.

331 In this work, we aim to provide clear methodological and practical guidance
332 for researchers tasked with estimating epidemiological delay distributions. We also
333 aim to provide robust and flexible tools and methods, both rederivations of those
334 presented elsewhere and novel, for them to apply this guidance in practice. We do
335 this by first introducing some general theories for characterizing epidemiological delay
336 distributions justified by grouping individuals into cohorts based on their observation
337 time. We then introduce in detail the biases that are common when estimating
338 epidemiological delay distributions. Based on this understanding, we then introduce
339 an exact method for accounting for common biases. We also introduce a range
340 of other commonly used and novel methods that represent simplifications of this
341 approach. We then evaluate these methods, both on simulated scenarios and on
342 data from the 2014-2016 Sierra Leone Ebola Virus disease epidemic. Finally, we
343 consider areas that require further development. Whilst the details of each section
344 are important, we also provide a summary for each that contains the main points in
345 order to aid understanding. In addition, we have summarised what was known prior
346 to this paper, what this paper adds, key findings, and key limitations.

347 2 Methods

348 2.1 General theory for measuring epidemiological delay dis- 349 tributions

350 Here we give a conceptual, visual, and mathematical overview of the general theory
351 for measuring epidemiological delay distributions and the specific theory relating to
352 dynamic bias. This theory is then used in later sections to relate to other forms of
353 bias and to justify and develop inference methods. We first give a summary of the
354 key points.

355 Summary

- 356 • The *intrinsic* distribution is the theoretical distribution that characterises the
357 underlying epidemiological process of interest. It describes the probability of
358 waiting a certain amount of time between a *primary* and a *secondary* event
359 (e.g. between infection and symptom onset) under constant conditions in the
360 population.
- 361 • Realised epidemiological delays can be measured both forward, starting from
362 the primary event toward the secondary event, or backward, starting from
363 the secondary event back to the primary event. For any given individual, the
364 direction of measurement does not affect the length of the delay. Intrinsic and
365 realized transmission intervals (e.g., generation interval and serial interval) can
366 differ systematically due to changes in transmission conditions (e.g., susceptible
367 pool in the population).
- 368 • The *forward distribution* is measured from a cohort of individuals who experi-
369 enced the primary event at the same time and is expected to give a good
370 estimate of the intrinsic distribution when conditions remain constant. For
371 modelling purposes, the forward distribution is often preferred over the back-
372 ward distribution as it better approximates the intrinsic distribution.
- 373 • The *backward distribution* is measured from a cohort of individuals who experi-
374 enced the secondary event at the same time. For a given intrinsic distribution,
375 the backward distribution can systematically vary over time and differ from the
376 forward distribution. This is due to the interaction between the observation
377 process and the temporal change in the incidence of the primary event.

378 **Conceptual overview** For any epidemiological process, there is an underlying *in-*
379 *trinsic* distribution that describes the time difference between events in the process.
380 In the context of epidemiological delays, the intrinsic distribution is a theoretical
381 distribution that describes the probability of waiting a certain amount of time be-
382 tween a *primary* and a *secondary* event (e.g. between infection and symptom onset)

383 under constant conditions in the population (Champredon and Dushoff, 2015). This
384 distribution can be calculated by averaging across individuals: for example, the in-
385 trinsic generation-interval distribution only depends on the average infectiousness of
386 infected individuals at a given time and does not depend on other factors, such as
387 intervention measures or the proportion susceptible (Park et al., 2021). As the in-
388 trinsic distribution characterises the underlying infection characteristic, it is used for
389 modeling, and is therefore often the distribution that we want to estimate. However,
390 the intrinsic distribution is generally not directly observable, as it may differ from
391 the realised distributions that are measured from the actual primary and secondary
392 events observed during an epidemic. In particular, realised distributions can change
393 over the course of an epidemic: for example, changes in realised generation intervals
394 can reflect changes in transmission dynamics, including susceptible depletion. For
395 practitioners, characterizing both the intrinsic and realised distribution is important.

396 In most cases, the primary event always occurs before the secondary event (e.g.,
397 infection followed by symptom onset). However, there are exceptions: for example,
398 an infectee may develop symptoms before their infector (Svensson, 2007). Negative
399 delays can also happen for single-individual events: an infected individual may test
400 positive before or after symptom onset (Singanayagam et al., 2020). Here we focus on
401 non-negative, single-individual delays, i.e. where the secondary event always happens
402 after the primary event. We also focus on the case where forward distributions do not
403 change over the course of an epidemic—later in the Ebola virus disease example, we
404 show that this is not always the case. Many of the lessons in this paper will however
405 carry over to more complicated cases involving negative delays and/or delays between
406 epidemiological events from two individuals.

407 We take the reporting delay—defined as the time from symptom onset (the pri-
408 mary event) to confirmation (the secondary event)—as an example. There are two
409 different approaches to measuring this interval. If we group individuals based on
410 their time of symptom onset and follow them until confirmation we are measuring
411 *forward* delay. If we instead group individuals based on when they tested positive
412 and ask when they developed symptoms, we are measuring *backward* delays. For
413 any individual realisation of a delay as a paired primary and secondary events, the
414 length of forward and backward delays are identical. However, in a *cohort* of indi-
415 viduals that experienced the primary (forward) or secondary (backward) events at
416 the same time, the resulting forward and backward distributions can systematically
417 differ when incidence is changing over time. Taking the cohort approach also allows
418 us to ask how the forward and backward distributions change over the course of an
419 epidemic.

420 As an example, consider a population where the incidence of infection is growing
421 exponentially (Fig. 1A). If we take a cohort of individuals who developed symptoms
422 on day 25, then we observe Fig. 1B, which corresponds to a *forward* distribution.
423 In this case, we see that the forward distribution in Fig. 1B matches the intrinsic
424 distribution that we used to simulate Fig. 1A. In general, we expect the forward
425 distribution to approximate the intrinsic distribution reasonably well, in particular

426 under relatively stable external conditions (Park et al., 2021). This means that
427 the forward distribution (rather than the backward distribution) is most useful for
428 downstream analysis, modeling, and decision support.

429 Some delay distributions are observed when the secondary event is reported, and
430 the timing of the primary event is then recalled or inferred (e.g. in a reporting delay,
431 typically the confirmation is first observed and the date of symptom onset recalled
432 by the patient). Since at this point, both primary and secondary events have already
433 happened, backward-looking cohorts, and hence backward distributions, typically do
434 not have truncation. For example, if we take a cohort of individuals who tested posi-
435 tive on day 25, we observe Fig. 1C which corresponds to a *backward* distribution.
436 In this case, we see that the backward distribution has a shorter mean than both
437 the intrinsic and forward distributions. This is because the backward-looking cohort
438 contains a mixture of individuals who developed symptoms recently (shorter report-
439 ing delays) and individuals who developed symptoms further in the past (longer
440 reporting delays). In the case of a growing epidemic, individuals who developed
441 symptoms recently are more abundant, and so the backward distribution is shifted
442 toward shorter intervals. Similarly, when the epidemic is declining, individuals who
443 developed symptoms recently are rare, and so the backward distribution is shifted
444 toward longer intervals (and has a longer mean than the intrinsic distribution) (Xin
445 et al., 2021). More generally, the backward distribution always depends on the time-
446 varying incidence of primary events. We refer to this dependence as dynamical bias
447 because it biases the backward distribution in comparison to the forward distribution
448 (Park et al., 2021).

449 Additionally, the realized forward and backward distributions are both susceptible
450 to a range of biases. Both can depend on events that are censored and both may
451 be truncated. In the case of forward-looking cohorts, the most common bias is right
452 truncation as the data set may be finalized for analysis before all individuals are
453 observed. Similarly, the backward distribution suffers from left truncation when
454 data on primary events are not available before some time point (Cain et al., 2011).

455 In particular, we note that, in the exponential growth case, there is a theoretical
456 equivalence between backward and truncated forward cohorts. This can be seen in
457 Fig. 1D, which can be calculated by taking all individuals who were infected and
458 developed symptoms before day 25 (therefore truncating all events that happened
459 after day 25). This results from an equivalence of the backward and truncated cohorts
460 in exponential growth settings. In the following section, we describe the effects of
461 truncation (including its similarities and differences with the dynamical bias) and
462 censoring in detail.

463 **Mathematical definition** To sharpen our discussion of forward and backward
464 distributions, we define them mathematically. As in the conceptual overview and
465 Fig. 1, we often define cohorts based on time intervals (e.g., a group of individu-
466 als who were infected on the same day, same week, or since the start of an epi-
467 demic). Mathematically, we can instead define incidence, and therefore cohorts, at

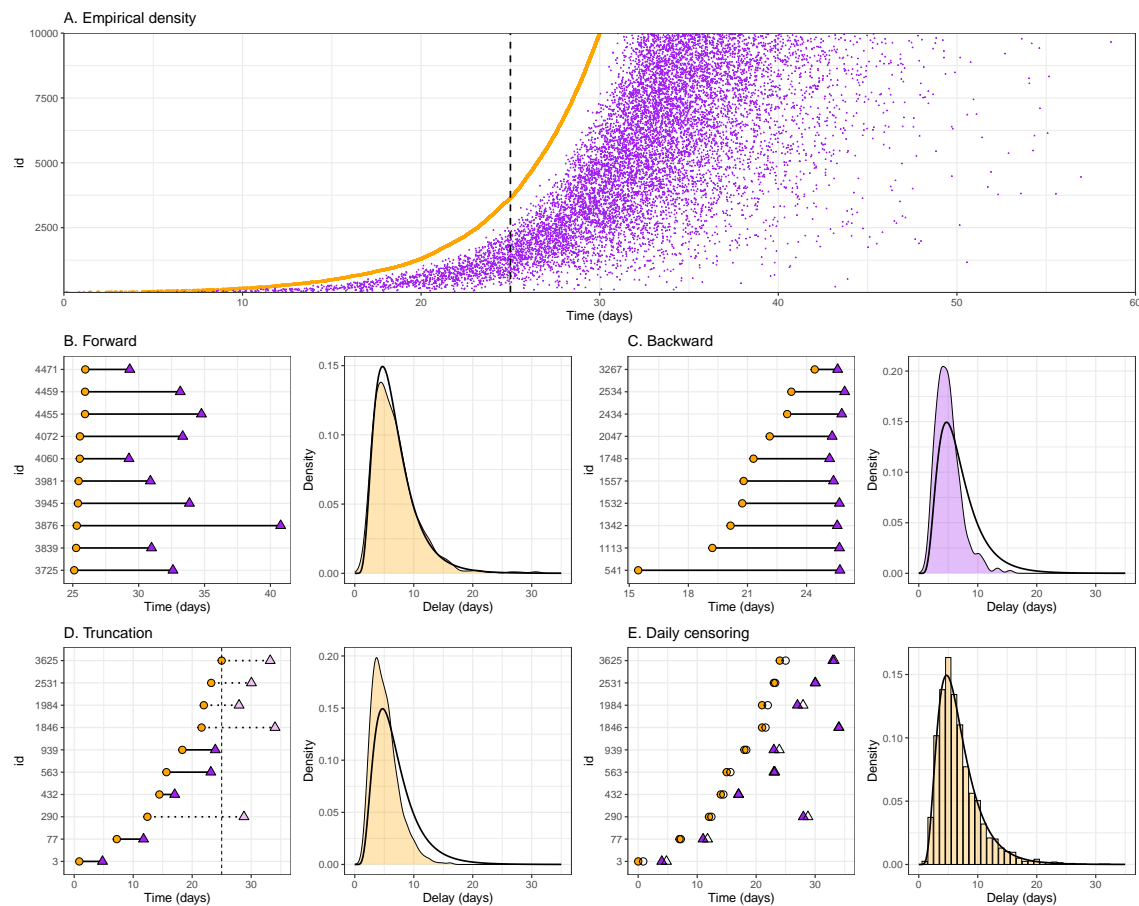


Figure 1: Schematic diagrams showing forward and backward distributions with the impact of truncation and censoring biases. (A) The density plot of exponentially increasing cumulative cases of primary events (orange) and the corresponding secondary events (purple). Case IDs (y-axis) are ordered by the primary event time. The dashed vertical line indicates day 25. Panels B–E show four different delay distributions using day 25 as a reference day. The left panels show a sample data set with circles representing the timing of primary events and triangles representing the timing of secondary events. The right panels show the corresponding distributions (filled areas) against the intrinsic distribution (transparent areas, solid lines). (B) The empirical forward distribution measured from a cohort of individuals who experienced the primary event on day 25. (C) The empirical backward distribution measured from a cohort of individuals who experienced the secondary event on day 25. (D) The truncated distribution measured from all individuals who experienced both the primary and secondary events before day 25 (i.e., those with solid lines connecting their events). (E) The observed distribution under daily reporting from all individuals who experienced the primary event before day 25 (without truncation). Empty circles and triangles represent the event dates under daily reporting.

468 any given, infinitely precise, *time point*: for example, in the standard continuous-
 469 time Susceptible-Infected-Recovered (SIR) model, the incidence of infection at time
 470 t corresponds to $\beta S(t)I(t)$, where β represents the transmission rate, S represents
 471 the number of susceptible individuals, and I represents the number of infected indi-
 472 viduals.

473 Here, we write $f_p(\tau)$ to represent the conditional probability density of observing
 474 a forward delay of length τ given that the primary event occurred at time p . Whilst
 475 the forward distribution can vary, for example over time or by risk or age groups,
 476 we primarily focus on cases where the forward distribution remains stable over the
 477 course of an epidemic (i.e., $f_p(\tau) = f(\tau)$ for all p in this paper), and is thus identical
 478 to the intrinsic distribution. We also write $\mathcal{P}(p)$ to denote the incidence of primary
 479 events (i.e., the rate at which individuals experience the primary event at time p).
 480 Likewise, we write $b_s(\tau)$ to represent the conditional probability density of observing
 481 a backward delay of length τ given that the secondary event occurred at time s .
 482 We also write $\mathcal{S}(s)$ to denote the incidence of secondary events (i.e., the rate at
 483 which individuals experience the secondary event at time s). Then, the total density
 484 of individuals $\mathcal{T}(p, s)$ who experienced the primary event at time p and secondary
 485 event at time s can be equivalently expressed in terms of the forward distribution:

$$\mathcal{T}(p, s) = \mathcal{P}(p)f_p(s - p), \quad (1)$$

486 and the backward distribution:

$$\mathcal{T}(p, s) = \mathcal{S}(s)b_s(s - p). \quad (2)$$

487 We can then write down a relationship between the forward and backward dis-
 488 tributions using the above relationship:

$$\mathcal{P}(p)f_p(s - p) = \mathcal{S}(s)b_s(s - p). \quad (3)$$

489 By substituting $p = s - \tau$ and rearranging, we see that the backward distribution
 490 is given by the incidence of primary events normalized by the size of the relevant
 491 cohort:

$$b_s(\tau) = \frac{\mathcal{P}(s - \tau)f_{s - \tau}(\tau)}{\mathcal{S}(s)} = \frac{\mathcal{P}(s - \tau)f_{s - \tau}(\tau)}{\int_0^\infty \mathcal{P}(s - x)f_{s - x}(x) dx}. \quad (4)$$

492 Here, the denominator $\mathcal{S}(s) = \int_0^\infty \mathcal{P}(s - x)f_{s - x}(x) dx$ represents the normalization
 493 constant such that the probability distribution integrates to 1. In particular, when
 494 the incidence of primary events is growing exponentially at rate r and the forward
 495 distribution is static (i.e. $f_p(\tau) = f(\tau)$), the backward distribution is also static, and
 496 given by:

$$b_{\text{exp}}(\tau) = \frac{\exp(-r\tau)f(\tau)}{\int_0^\infty \exp(-rx)f(x) dx}. \quad (5)$$

497 This simplified relationship has been used extensively in the literature (Britton and
 498 Scalia Tomba, 2019; Verity et al., 2020b; Park et al., 2022). When $r > 0$, the

499 backward distribution will have a shorter mean than the corresponding forward dis-
500 tribution because we are more likely to observe individuals who experienced the
501 primary event recently. This dependence on the incidence of the primary event leads
502 to *dynamical bias* when the backward distribution is used as an estimate of the in-
503 trinsic distribution (Park et al., 2022). As can be seen from Eq. (5), the forward and
504 backward distributions will be equivalent when incidence is stable ($r = 0$).

505 2.2 Biases in estimating forward distributions

506 In this section, we lay out mathematical foundations for understanding truncation
507 and censoring biases. The framework described in this section will provide a basis for
508 deriving likelihoods for inference in the following section. We first give a summary
509 of the key points.

510 Summary

- 511 • Right truncation occurs because we cannot observe event pairs whose secondary
512 event lies in the future. We do not know how much of the data is truncated
513 until all event pairs have been fully observed. This effect biases us toward
514 observing shorter delays when working in real time.
- 515 • The amount of right truncation increases as the epidemic grows faster because
516 the amount of recent primary events becomes relatively more common.
- 517 • When incidence is growing exponentially, the effect of right truncation on the
518 forward distribution is quantitatively equivalent to the effect of dynamical bias
519 on the backward distribution. When incidence is shrinking, this is not the
520 case: there will be minimal right truncation but the dynamical bias causes the
521 backward distribution to have a longer mean.
- 522 • Censoring occurs when the precise time of an event is unknown but we know
523 an event has occurred. Here, we focus on fitting continuous-time distribu-
524 tions. This means that the recording of the exact *date* of an event is implicitly
525 censored because the exact *time* of an event (within a given date) is still un-
526 known. Alternatively, it would be possible to model this using discrete-time
527 distributions—we do not explore this approach in this paper.
- 528 • When we estimate delay distributions, we have to account for censoring in both
529 primary and secondary events. Typically, many researchers rely on a single-day
530 censoring window, which can give biased inferences.
- 531 • Assumptions about event time distributions within a censoring interval can bias
532 the estimation of continuous-time distributions from the observed discrete-time
533 delays as well as the discretisation of continuous-time distributions. Modeling
534 choices for fitting continuous- vs discrete-time distributions need to be carefully
535 considered.

536 **Right truncation** Right truncation arises from the inability to observe secondary
537 events that happen after the observation time T (see Fig. 1D). For example, during
538 an ongoing epidemic, we tend to underestimate the infection fatality risk because
539 we cannot observe all of the outcomes for infected individuals in real time due to
540 the delay between infection and death (or recovery) and because we typically base
541 observation on individuals having died (or recovered) (Ghani et al., 2005; Lipsitch
542 et al., 2015). In general, we don't know how much of the data is truncated until
543 all outcomes have been fully observed (Gelman et al., 2013, p. 227). Since right
544 truncation limits our ability to observe long intervals (i.e. those in the right tail of a
545 distribution), it will generally bias us towards observing shorter delays (and therefore
546 to underestimate the mean).

To define this mathematically, we introduce random variables P and S representing primary and secondary events. Then, the conditional probability of observing a forward delay of length τ (i.e., $S = P + \tau$) given primary event time $P = p$ and observation time T (i.e., $S < T$) can be written as:

$$P(S = P + \tau | P = p, S < T) = \frac{P(P = p, S = P + \tau)}{P(P = p, S < T)} \quad (6)$$

$$= \frac{P(S = P + \tau | P = p)P(P = p)}{P(S = P + \tau < T | P = p)P(P = p)} = \frac{P(S = P + \tau | P = p)}{P(S = P + \tau < T | P = p)} \quad (7)$$

$$= \frac{f_p(\tau)}{\int_0^{T-p} f_p(x) dx} \quad (8)$$

$$= \frac{f_p(\tau)}{F_p(T - p)}, \quad (9)$$

547 where f_p is the forward distribution and F_p is the forward cumulative probability
548 distribution. Conceptually, Eq. (9) is saying that the probability of observing a
549 given delay from a truncated distribution is equal to the probability of observing
550 that delay from the untruncated forward distribution normalised by the probability
551 of observing any delay before the end of observation. From a cohort perspective,
552 when we measure forward delays from individuals who experienced primary events
553 at time $p < T$, we can only observe delays that are shorter than $T - p$. This means
554 that the amount of truncation is higher in a cohort of individuals who experienced
555 their primary events more recently.

556 The population-level effect of right truncation will be exaggerated when an epi-
557 demic is growing because there will be more individuals who experienced primary
558 events recently, meaning their secondary events may not have happened yet. When
559 growth is exponential, the effect of right truncation is quantitatively equivalent to
560 dynamical bias, as shown in Fig. 1C–D. In contrast, when the epidemic is declining
561 quickly, fewer individuals experienced primary events recently, and the effect of right
562 truncation can become negligible as few delays are unobserved. On the other hand,
563 dynamical bias will still be an issue because a cohort of individuals who experienced
564 secondary events during the decay phase will include a relatively large proportion of

565 those who had their primary events further in the past, causing the backward dis-
 566 tribution to have a longer mean than the forward distribution. As dynamical biases
 567 specifically refer to the dependence of backward distribution on epidemic dynamics,
 568 they need not be taken into account when analyzing the forward distribution as long
 569 as biases due to truncation are considered.

570 **Interval censoring** Censoring occurs when the exact event time is unknown. In
 571 contrast to truncation, we know that the focal event happened in the case of inter-
 572 val censoring; we just do not know precisely when. For example, we know that a
 573 symptomatic individual was infected sometime before symptom onset but we usually
 574 don't know when. Even if the exact date of the event is known, implicit censoring
 575 remains because the event could have theoretically occurred at any time within a
 576 24-hour time window. This implicit censoring discretises the observed delay distri-
 577 bution (Fig. 1E). To define censoring of both primary and secondary events, which is
 578 known as double censoring (Reich et al., 2009), we start by first defining the simple
 579 single censoring case. Here the secondary event is censored between time S_L and S_R ,
 580 and the primary event (p) is known. Then, we have the following relationship,

$$P(S_L < S < S_R | P = p) = \int_{S_L}^{S_R} f_p(y - p) dy. \quad (10)$$

581 Integrating this leaves

$$P(S_L < S < S_R | P = p) = F_P(S_R - P) - F_P(S_L - P). \quad (11)$$

582 This relationship is commonly used in practice to discretise delay distributions, par-
 583 ticularly into daily intervals (Flaxman et al., 2020; Abbott et al., 2020). However,
 584 as we have conditioned on the primary event being known, this approach is not ap-
 585 propriate for the common case where P is also censored, for example, if only a date
 586 is known. The degree of bias this will introduce depends on event probability within
 587 the primary censoring interval.

For the more common setting where the primary event is also censored between
 time P_L and P_R the conditional probability that the secondary event occurs between
 time S_L and S_R can be written as:

$$P(S_L < S < S_R | P_L < P < P_R) = \frac{P(P_L < P < P_R, S_L < S < S_R)}{P(P_L < P < P_R)} \quad (12)$$

$$= \frac{\int_{P_L}^{P_R} \int_{S_L}^{S_R} g_P(x) f_x(y - x) dy dx}{\int_{P_L}^{P_R} g_P(x) dx} \quad (13)$$

$$= \int_{P_L}^{P_R} \int_{S_L}^{S_R} g_P(x | P_L, P_R) f_x(y - x) dy dx \quad (14)$$

588 Here, $g_P(x)$ is the probability distribution of the primary event time, and $g_P(x | P_L, P_R)$
 589 is the conditional probability distribution of the primary event time given its lower

590 and upper bounds,

$$g_P(x|P_L, P_R) = \frac{g_P(x)}{\int_{P_L}^{P_R} g_P(z) dz} \quad (15)$$

591 As $g_P(x)$ depends on the incidence of primary events, and therefore epidemic dy-
592 namics, this can be rewritten as,

$$g_P(x|P_L, P_R) = \frac{\mathcal{P}(x)}{\int_{P_L}^{P_R} \mathcal{P}(z) dz}, \quad (16)$$

593 where $P(x)$ is a continuous-time incidence, as explained earlier; in practice, account-
594 ing for this factor will require estimating continuous-time incidence from discrete-
595 time incidence. Making the simplifying assumption of a fixed epidemic growth rate
596 r , the conditional probability can then be rewritten as,

$$P(S_L < S < S_R | P_L < P < P_R) = \frac{\int_{P_L}^{P_R} \int_{S_L}^{S_R} \exp(rx) f_x(y-x) dy dx}{\int_{P_L}^{P_R} \exp(rx) dx} \quad (17)$$

597 The shape of the conditional probability distribution $g_P(x|P_L, P_R)$ is shown in Fig. 2
598 for different epidemic growth rates.

Censoring and truncation Finally, to understand the joint effects of right trun-
cation and interval censoring on the forward distribution, we can quantify the condi-
tional probability of the secondary event occurring between time S_L and S_R given
the primary event occurring between P_L and P_R and the observation time T :

$$P(S_L < S < S_R | P_L < P < P_R, S < T) = \frac{P(P_L < P < P_R, S_L < S < S_R)}{P(P_L < P < P_R, S < T)} \quad (18)$$

$$= \frac{\int_{P_L}^{P_R} \int_{S_L}^{S_R} g_P(x) f_x(y-x) dy dx}{\int_{P_L}^{P_R} \int_x^T g_P(x) f_x(y-x) dy dx} \quad (19)$$

$$= \frac{\int_{P_L}^{P_R} \int_{S_L}^{S_R} g_P(x) f_x(y-x) dy dx}{\int_{P_L}^{P_R} g_P(x) F_x(T-x) dx} \quad (20)$$

By dividing both the numerator and denominator by $\int_{P_L}^{P_R} \mathcal{P}(z) dz$, we have:

$$\begin{aligned} & P(S_L < S < S_R | P_L < P < P_R, S < T) \\ &= \frac{\int_{P_L}^{P_R} \int_{S_L}^{S_R} g_P(x|P_L, P_R) f_x(y-x) dy dx}{\int_{P_L}^{P_R} g_P(x|P_L, P_R) F_x(T-x) dx} \end{aligned} \quad (21)$$

599 Here, expressing the probability distribution of primary events $g_P(x)$ in terms of the
600 conditional probability distribution $g_P(x|P_L, P_R)$ allows for a more flexible inferential

601 framework later on. We also see that the truncation problem also depends on the
602 censoring problem: uncertainties in the primary event time further affect the amount
603 of truncation. This framework generalises the work of Seaman et al. (2022) who
604 reviewed various methods for dealing with right truncation bias but did not account
605 for interval censoring.

606 **2.3 Methods for estimating forward delay distributions from** 607 **observed data**

608 In this section, we describe a method for estimating delay distributions from the
609 observed data based on the theory for censoring and truncation. We then introduce
610 a series of approximations in order to motivate commonly used methods from the
611 literature in order of least to most approximate (with a corresponding decrease in
612 their ability to account for right truncation and censoring biases).

613 **Summary**

- 614 • There is an exact solution for modelling delays that are double censored and
615 right truncated. In practice, this method is not generally practical or stable in
616 real-time settings, and for this reason, we discuss a series of approximations in
617 use.
- 618 • Many of the more commonly used approximations do not fully account for
619 censoring, truncation, or both forms of bias. There are also a range of trade-
620 offs that need consideration.
- 621 • Methods that use a latent variable censoring approach and adjust for the obser-
622 vation time are expected to be the most robust currently available for real-time
623 estimation.

624 **2.3.1 Simplifying assumptions for the following methods**

625 We only consider single-individual delays, where we have independent observations
626 of event times across different individuals. This means that in this work, we do not
627 consider methods that account for non-independence in generation or serial intervals.
628 These pair delays also introduce additional biases due to uncertainties about who
629 infected whom—we do not address these issues here. Nonetheless, the methods we
630 present here can still be applied to adjust for truncation, censoring, and dynamical
631 biases in estimating generation- and serial-interval distributions. Finally, we also as-
632 sume that the forward distribution stays constant over time (i.e., $f_p(\tau) = f_{\text{forward}}(\tau)$).
633 All of the methods we consider can be generalised to delays that vary in discrete
634 time using our implementations.

635 2.3.2 Exact censoring and truncation method

Under daily reporting, the exact timing of the primary and secondary event times are unknown. Instead, the corresponding observed event times are $P_L = \lfloor P \rfloor$ and $S_L = \lfloor S \rfloor$, respectively. Likewise, the corresponding upper bounds for the censoring interval are $P_R = P_L + 1$ and $S_R = S_L + 1$, respectively. Following Eq. (21), the joint likelihood can be written as:

$$\mathcal{L}_{\text{exact}}(\mathbf{Y}|\boldsymbol{\theta}) = \prod_i \left[\frac{\int_{p_{L,i}}^{p_{R,i}} \int_{s_{L,i}}^{s_{R,i}} g_P(x|p_{L,i}, p_{R,i}) f_{\text{forward}}(y-x) dy dx}{\int_{p_{L,i}}^{p_{R,i}} g_P(z|p_{L,i}, p_{R,i}) F_{\text{forward}}(T-z) dz} \right], \quad (22)$$

636 where $\boldsymbol{\theta}$ represents the parameter vector, and \mathbf{Y} represents the data vector.

From a practical standpoint, it is difficult to use this likelihood for inference, especially in real-time epidemic monitoring applications where estimation must be fast and robust. Solving double integrals analytically may be impossible, and calculating them numerically can be computationally costly and numerically unstable. Instead, we can implement this model using an equivalent Bayesian latent variable approach. Here we can treat the primary and secondary event times as latent variables (denoted x_i and y_i , respectively), conditional on the known upper and lower bound of these events for each individual, and then integrate across the uncertainty. Then, by taking away both integrals in the numerator, Eq. (22) can be re-written as:

$$x_i \sim g_P(x|p_{L,i}, p_{R,i}) \quad (23)$$

$$y_i \sim \text{Uniform}(s_{L,i}, s_{R,i}) \quad (24)$$

$$\mathcal{L}_{\text{exact}}(\mathbf{Y}|\boldsymbol{\theta}) = \prod_i \left[\frac{f_{\text{forward}}(y_i - x_i)}{\int_{p_{L,i}}^{p_{R,i}} g_P(z|p_{L,i}, p_{R,i}) F_{\text{forward}}(T-z) dz} \right] \quad (25)$$

637 Here, we use the uniform distribution for the secondary event time y_i , so that y_i
 638 does not affect the likelihood, which allows Eq. (25) to be equivalent to Eq. (22).
 639 Using any other distribution for y_i will cause Eq. (25) to deviate from Eq. (22) and
 640 therefore will be incorrect.

In practice, modeling the conditional distribution of primary event time $g_P(x|p_{L,i}, p_{R,i})$ is expected to be a difficult problem as it depends on the changes in the incidence of primary events (Eq. (16)). However, if we assume that the incidence of the primary event is changing at a fixed growth rate r within the censoring interval we can further simplify the problem:

$$g_P(x|p_{L,i}, p_{R,i}) = \frac{\exp(rx)}{\int_{p_{L,i}}^{p_{R,i}} \exp(rz) dz} \quad (26)$$

$$= \frac{r \exp(rx)}{\exp(rp_{R,i}) - \exp(rp_{L,i})} \quad (27)$$

641 Generalizing the problem beyond stable exponential growth within the interval re-
 642 quires a way of modeling realistic changes in incidence. Even with this simplifying

643 assumption in most real-world settings this would require joint estimation of the
 644 growth rate of the primary events in order to properly propagate uncertainty; how-
 645 ever, such a model is likely to be computationally more costly.

646 The exponential scenario provides insights for understanding the method intro-
 647 duced by Linton et al. (2020) during the beginning of the SARS-CoV-2 pandemic.
 648 Under Eq. (27), our likelihood is similar in form to the likelihood presented in
 649 Linton et al. (2020) with one major distinction (among several other minor dif-
 650 ferences). The conditional probability $g_P(x|p_{L,i}, p_{R,i})$ we present here converges to a
 651 uniform distribution as $r \rightarrow 0$ as expected when incidence is stable (Fig. 2A). On
 652 the other hand, the corresponding component in Linton et al. (2020) is modeled as
 653 $r \exp(-ru)/(1 - \exp(-ru))$, where u is a variable of integration; this term does not
 654 converge as $r \rightarrow 0$.

655 2.3.3 Approximate latent variable censoring and truncation method

When the primary event time is assumed to be uniformly distributed and censoring interval is narrow, the integral in the denominator of Eq. (25) ($\int_{p_{L,i}}^{p_{R,i}} g_P(z|p_{L,i}, p_{R,i}) F_{\text{forward}}(T - z) dz$) can be approximated by $F_{\text{forward}}(T - x_i)$ for some $p_{L,i} < x_i < p_{R,i}$. Under these conditions, the exact method can be rewritten as (Ward et al., 2022):

$$x_i \sim \text{Uniform}(p_{L,i}, p_{R,i}) \quad (28)$$

$$y_i \sim \text{Uniform}(s_{L,i}, s_{R,i}) \quad (29)$$

$$\mathcal{L}_{\text{latent}}(\mathbf{Y}_i|\boldsymbol{\theta}) = \frac{f_{\text{forward}}(y_i - x_i)}{F_{\text{forward}}(T - x_i)} \quad (30)$$

656 The likelihood for this method corresponds to the conditional probability under
 657 right truncation (Eq. (9)). This is a convenient approximation as it does not require
 658 additional integration. While it would be possible to extend this method to account
 659 for more complex prior distributions that capture epidemic growth or daily activities
 660 (e.g., circadian rhythms), the exact method (Eq. (25)) should be considered to avoid
 661 potential biases arising from the approximations.

662 2.3.4 Interval-reduced censoring and truncation methods

Without using latent variables, we can equivalently write down the approximate latent variable model as follows:

$$\mathcal{L}_{\text{latent}}(\mathbf{Y}_i|\boldsymbol{\theta}) = \int_{p_{L,i}}^{p_{R,i}} \int_{s_{L,i}}^{s_{R,i}} g_P(x_i|p_{L,i}, p_{R,i}) g_S(y_i|s_{L,i}, s_{R,i}) \frac{f_{\text{forward}}(y_i - x_i)}{F_{\text{forward}}(T - x_i)} dy_i dx_i, \quad (31)$$

where $g_S(y_i|s_{L,i}, s_{R,i})$ corresponds to the uniform distribution between $s_{L,i}$ and $s_{R,i}$. When we rewrite in terms of the delay, $d_i = y_i - x_i$, we know that d_i should range

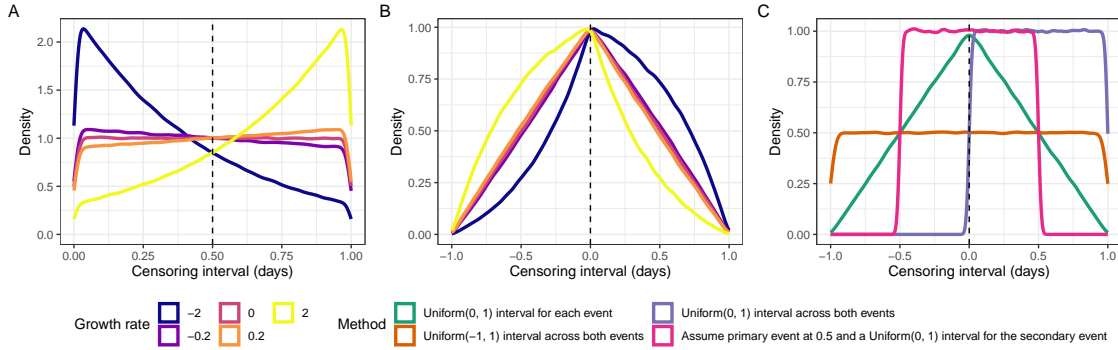


Figure 2: **Schematic diagrams showing different assumptions about interval censoring.** (A) The distribution of primary event times within a one-day censoring interval across different growth rates. Note that a sudden decrease in probability densities around 0 and 1 are numerical artifacts from plotting. (B) The corresponding distribution of weights for the interval-reduced censoring, which is a convolution between the distribution of event times of primary and secondary events within each censoring interval. (C) The distribution of weights for the interval-reduced censoring for different approximations.

between $s_{L,i} - p_{R,i}$ and $s_{R,i} - p_{L,i}$. Further assuming $F_{\text{forward}}(T - x_i) \approx F_{\text{forward}}(T - p_{R,i})$, we can write the above likelihood using a single integral:

$$\mathcal{L}_{\text{latent}}(\mathbf{Y}_i | \boldsymbol{\theta}) = \int_{s_{L,i} - p_{R,i}}^{s_{R,i} - p_{L,i}} w(d_i) \frac{f_{\text{forward}}(d_i)}{F_{\text{forward}}(T - p_{R,i})} dd_i. \quad (32)$$

663 Here, $w(d_i)$ is a convolution between $g_P(x_i | p_{L,i}, p_{R,i})$ and $g_S(y_i | s_{L,i}, s_{R,i})$:

$$w(d_i) = \int_{p_{L,i}}^{p_{R,i}} g_P(x_i | p_{L,i}, p_{R,i}) g_S(d_i - x_i | s_{L,i}, s_{R,i}) dx_i. \quad (33)$$

664 When both $g_P(x_i | p_{L,i}, p_{R,i})$ and $g_S(d_i - x_i | s_{L,i}, s_{R,i})$ are assumed to follow uniform
 665 distributions, $w(d_i)$ will have a trapezoid shape (Fig. 2C), as discussed in Reich et al.
 666 (2009).

667 In Fig. 2B–C, we show how the shape of the weight function $w(d_i)$ changes for
 668 different epidemic growth rates (B) and for different assumptions (C). In particular,
 669 assuming a uniform distribution for both events censoring intervals approximates the
 670 interval-reduced censoring window well when the growth rate is small compared to
 671 the width of the censoring window and less well when this is not the case. In Fig. 8,
 672 we present the consequences of different assumptions about $w(d_i)$ on the resulting
 673 continuous-time distributions.

674 We can further simplify the likelihood by assuming a uniform distribution for
 675 $w(d_i)$ (Fig. 2C):

$$\mathcal{L}_{\text{cens+trunc}}(\mathbf{Y} | \boldsymbol{\theta}) = \prod_i \left[\frac{F_{\text{forward}}(s_{R,i} - p_{L,i}) - F_{\text{forward}}(s_{L,i} - p_{R,i})}{F_{\text{forward}}(T - p_{R,i})} \right]. \quad (34)$$

676 This method is simple to implement as well as being computationally convenient and
677 supported by many software packages, such as the `brms` package (Bürkner, 2018) in
678 the R programming language (R Core Team, 2019).

679 However, it does not fully account for the censoring bias as the uniform assump-
680 tion across the combined interval is not a good match for the prior that results from
681 combining two uniforms or in more complex settings a growth-mediated prior and a
682 uniform (Fig. 2B–C) or the impact of censoring on the amount of truncation due to
683 the simplified denominator. This means that, depending on the amount of censoring
684 and the shape of the delay distribution, this method will have a residual bias in both
685 the mean and standard deviation of the estimated distribution (Fig. 8). In the spe-
686 cial case when the primary event time is exactly known (therefore, $p_{L,i} = p_i = p_{R,i}$),
687 this method would be accurate, and also equivalent to the latent variable method.

688 In the daily reporting setting, we consider in the rest of this paper, the likelihood
689 can be reformulated in terms of the observed, discrete-time delay ($d_i = S_{L,i} - P_{L,i}$):

$$\mathcal{L}_{\text{cens+trunc}}(\mathbf{Y}|\boldsymbol{\theta}) = \prod_i \left[\frac{F_{\text{forward}}(d_i + 1) - F_{\text{forward}}(d_i - 1)}{F_{\text{forward}}(T - p_{R,i})} \right]. \quad (35)$$

690 Rephrasing the likelihood in this way highlights that the correct censoring interval
691 for this approach is $[d_i - 1, d_i + 1]$. Instead, many researchers tend to rely on the
692 $[d_i, d_i + 1]$ interval (Flaxman et al., 2020; Abbott et al., 2020), which in effect only
693 accounts for the censoring of a single event and hence induces additional bias beyond
694 that introduced by assuming a uniform distribution. Other potential approximations
695 are $[d_i - 0.5, d_i + 0.5]$, which assumes the primary event is known to have taken place
696 at the midpoint of the interval, and that all primary and secondary event times are
697 known (He et al., 2020). The implications of these approximations are also explored
698 in Fig. 8.

699 2.3.5 Interval-reduced censoring method

If we ignore the presence of right truncation, then Eq. (35) reduces to:

$$\mathcal{L}_{\text{cens}}(\mathbf{Y}|\boldsymbol{\theta}) = \prod_i [F_{\text{forward}}(d_{O,i} + 1) - F_{\text{forward}}(d_{O,i} - 1)], \quad (36)$$

700 which, again, assumes a uniform distribution for $w(d_i)$. Reich et al. (2009) tested
701 the consequences of this approximation and showed that assuming $w(d_i)$ is uniform
702 results in low coverage probabilities (defined as the proportion of confidence inter-
703 vals that contain the true value across multiple simulations). However, the uniform
704 assumption remains in regular use in the applied literature.

705 2.3.6 Truncation method

706 If we ignore the presence of censoring, Eq. (31) reduces to

$$\mathcal{L}_{\text{trunc}}(\mathbf{Y}|\boldsymbol{\theta}) = \prod_i \frac{f_{\text{forward}}(s_{L,i} - p_{L,i})}{F_{\text{forward}}(T - p_{L,i})}, \quad (37)$$

707 which can also be derived directly from Eq. (9) (Sun, 1995). Here, we use the lower
708 bound for the primary event time to estimate the amount of truncation $T - p_{L,i}$ as it
709 represents the maximum degree of truncation. Other choices include using the upper
710 bound $T - p_{R,i}$ or the midpoint $T - (p_{L,i} + p_{R,i})/2$. The use of an upper bound may
711 be numerically less stable because the upper bound of primary event time can be
712 equal to the observation time, in which case the denominator becomes zero.

713 2.3.7 Filtering method

714 A simple solution to mitigate truncation bias is to drop (or filter out) the most recent
715 data (as these data will be most impacted by truncation) and then fit a probability
716 distribution on the remaining observations. Given a filtering time t_{filter} , the likelihood
717 can be written as:

$$\mathcal{L}_{\text{filter}}(\mathbf{Y}|\boldsymbol{\theta}) = \prod_{P_{L,i} < t_{\text{filter}}} f_{\text{filter}}(S_{L,i} - P_{L,i}|\boldsymbol{\theta}), \quad (38)$$

718 This approach can be readily extended to include double censoring as the filtering
719 and distribution fitting steps are independent.

720 Unfortunately, whilst the filtering method may be convenient, it is sensitive to
721 our choices of filtering time t_{filter} . If t_{filter} is too early, we might not have enough data
722 to estimate the focal delay distribution with certainty. If t_{filter} is too late, truncation
723 bias might be introduced. More broadly, dropping existing data is often not a good
724 practice and at the very least is likely to increase the uncertainty of any estimated
725 distribution. This can be a particular issue during epidemics when data can be
726 sparse and as accurate as possible estimates are needed to inform decision-making.
727 For simplicity, we assume $t_{\text{filter}} = 10$ days throughout the paper.

728 2.3.8 Naive method

729 A simple approach to estimating a delay distribution is to directly fit a probability
730 distribution f_{naive} to the observed discrete delays. In this case, the likelihood is given
731 by:

$$\mathcal{L}_{\text{naive}}(\mathbf{Y}|\boldsymbol{\theta}) = \prod_i f_{\text{naive}}(S_{L,i} - P_{L,i}|\boldsymbol{\theta}), \quad (39)$$

732 where $\boldsymbol{\theta}$ represents the parameter vector and \mathbf{Y} represents the data vector. This
733 method does not account for truncation or censoring biases but was historically used
734 in practice due to convenience and lack of awareness of its limitations. An even more
735 basic form of this approach is to simply report the empirically observed summary
736 statistics such as the mean, standard deviation, or range —this approach is common
737 in practice (Nolen et al., 2016; Gostic et al., 2020; Miura et al., 2022). The no interval
738 method in Fig. 2 highlights that this approach does not approximate the underlying
739 generative process well in the presence of censoring.

740 **2.3.9 Discrete time methods for accounting for censoring and right trun-**
 741 **cation**

742 So far, we have focused on fitting probability distributions to continuous forward
 743 delays directly by accounting for right truncation and censoring, presenting meth-
 744 ods in order from more exact to most approximate. In this section, we outline
 745 discrete-time approaches to estimating epidemiological delay distributions whilst ac-
 746 counting for right truncation and censoring. These methods make the further sim-
 747 plifying assumption that censoring is daily for both primary and secondary events.
 748 To counterbalance this they both have a range of advantages to other methods in
 749 certain settings. In Fig. 9, we also present how different assumptions about the prior
 750 distributions within censoring intervals affect the discretisation of continuous-time
 751 distributions.

752 **Dynamical correction method** In some cases, it may be favourable to exploit
 753 the relationship between backward and forward distributions to estimate the forward
 754 distribution. We refer to this method as the dynamical correction method because
 755 we explicitly take into account the biases in the backward distribution caused by
 756 epidemic dynamics. This approach may be particularly attractive when the backward
 757 epidemiological distribution has been reported without correction and the underlying
 758 data are not available or data on the time from primary event to observation time
 759 is not available. This is still common for incubation periods which are typically
 760 calculated by identifying symptomatic individuals and asking when they became
 761 infected.

Previously, we considered the distribution of the primary event time to account for the censoring problem in the forward distribution. Analogously, we now have to consider the distribution of the secondary event time to account for the censoring problem in the backward distribution. The probability that the primary event occurs between time P_L and P_R given that the secondary event occurred between time S_L and S_R can be then written as:

$$P(P_L < P < P_R | S_L < S < S_R) = \frac{P(P_L < P < P_R, S_L < S < S_R)}{P(S_L < S < S_R)} \quad (40)$$

$$= \int_{S_L}^{S_R} \int_{P_L}^{P_R} g_S(y | S_L, S_R) b_y(y - x) dx dy \quad (41)$$

762 where $g_S(y | S_L, S_R)$ is the conditional probability distribution of the secondary event
 763 time given its lower S_L and upper S_R bounds. By substituting Eq. (4), we have:

$$P(P_L < P < P_R | S_L < S < S_R) = \int_{S_L}^{S_R} \int_{P_L}^{P_R} \frac{g_S(y | S_L, S_R) \mathcal{P}(x) f(y - x)}{\int_0^\infty \mathcal{P}(y - \tau) f(\tau) d\tau} dx dy \quad (42)$$

764 A special case of this framework is when the event times are exactly known and the
 765 incidence of primary events is changing exponentially at rate r . This special case was

766 used by Verity et al. (2020a) to estimate the symptom-onset-to-death distribution
 767 of COVID-19 and has been previously introduced in several other contexts (Britton
 768 and Scalia Tomba, 2019; Park et al., 2021).

769 The general case presented in Eq. (42) depends on a continuous-time incidence
 770 pattern (Eq. (4)), but these incidence data usually only exist in discrete time. There-
 771 fore, given patterns of incidence of primary events reported on a daily basis $\mathcal{P}_d(t)$,
 772 we approximate the continuous-time incidence as follows:

$$\mathcal{P}(t) \approx \mathcal{P}_d(\lfloor t \rfloor), \quad (43)$$

773 meaning that the primary incidence follows a step function, changing every day.

This approximation is convenient because it allows us to write the denominator of Eq. (42) in terms of a sum of integrals, which can be further simplified. In particular, setting $y = S_L$ (i.e., the reported time of secondary event), we have:

$$\int_0^\infty \mathcal{P}(S_L - \tau) f(\tau) d\tau \approx \int_0^\infty \mathcal{P}_d(\lfloor S_L - \tau \rfloor) f(\tau) d\tau \quad (44)$$

$$\approx \sum_{k=1}^{n-1} \mathcal{P}_d(S_L - k) \int_{k-1}^k f(\tau) d\tau \quad (45)$$

$$= \sum_{k=1}^{n-1} \mathcal{P}_d(S_L - k) (F(k) - F(k-1)), \quad (46)$$

where n represents the length of the incidence time series. Likewise, the integral in the denominator can be further simplified:

$$\int_{S_L}^{S_R} \int_{P_L}^{P_R} g_S(y|S_L, S_R) \mathcal{P}(x) f(y-x) dx dy \quad (47)$$

$$\approx \mathcal{P}_d(P_L) \int_{S_L}^{S_R} \int_{P_L}^{P_R} g_S(y|S_L, S_R) f(y-x) dx dy \quad (48)$$

$$\approx \mathcal{P}_d(P_L) (F(S_R - P_L) - F(S_L - P_R)), \quad (49)$$

774 where the last line follows from the interval-reduced censoring approximation used
 775 previously (i.e., assuming a uniform censoring across the censoring window of the
 776 delay, rather than each event). Putting everything together, we have:

$$P(P_L < P < P_R | S_L < S < S_R) \approx \frac{\mathcal{P}_d(P_L) (F(S_R - P_L) - F(S_L - P_R))}{\sum_{k=1}^{n-1} \mathcal{P}_d(S_L - k) (F(k) - F(k-1))} \quad (50)$$

777 Finally, the likelihood for the model can be written as a product of the above prob-
 778 ability:

$$\mathcal{L}_{\text{dynamical+cens}}(\mathbf{Y}|\boldsymbol{\theta}) = \prod_{i=1}^{\infty} \left[\frac{\mathcal{P}_d(p_{i,L}) (F(s_{i,R} - p_{i,L}) - F(s_{i,L} - p_{i,R}))}{\sum_{k=1}^{n-1} \mathcal{P}_d(s_{i,L} - k) (F(k) - F(k-1))} \right] \quad (51)$$

779 We note that these derivations rely on approximations that are specific to daily
780 censoring and therefore will not necessarily hold for the general censoring case.

781 Unfortunately, this method requires a complete time series, or some approxima-
782 tion of it such as an estimate of the growth rate, in order to properly account for right
783 truncation. In the following sections where we evaluate this approach, we first use
784 a partially reported (i.e., truncated) time series of primary event incidence (directly
785 aggregated from individual-level data) to evaluate the worst-case performance of this
786 method for real-time usage. We also consider the best case where the complete inci-
787 dence time series is known at the time of estimation. Approximations, such as using
788 an estimate of the real-time growth rate or nowcast of primary incidence, would then
789 perform intermediately between these two approaches depending on the accuracy of
790 the approximation (i.e. the accuracy of the growth rate estimate or nowcast) and the
791 propagation of uncertainty. Line lists may also not accurately reflect the incidence
792 pattern when there are changes in surveillance and reporting and so the ability of this
793 method to use potentially independent time series of incidence may be attractive.

794 **Jointly modelling incidence and the forward distribution.** Another alter-
795 native approach is to jointly model the discrete-time incidence of primary events
796 and the forward delay from primary to secondary events as a count process. This
797 method is more flexible than the previous dynamical approach as it does not require
798 a complete time series to be available. It also allows for an error model to be applied
799 to the primary incidence and can be naturally extended to include hazard effects and
800 non-parametric delay distributions (Abbott et al., 2021) which may be desirable for
801 complex, real-world, settings. This approach is routinely used in the nowcasting lit-
802 erature to reconstruct right truncated primary incidence curves (Abbott et al., 2021;
803 Höhle and an der Heiden, 2014; Günther et al., 2021). A downside of this approach
804 is that on top of specifying the reporting delay model, we must also define a model
805 for the evolution of the expectation of primary incidence over time. Here we use a
806 linear model for the exponential growth rate, though the package provides a large
807 range of alternatives, to allow some flexibility in the primary event incidence time
808 series.

809 As this method derives from the nowcasting literature it is commonly framed
810 in terms of the "reporting triangle" ($p_{t,d}$) (Höhle and an der Heiden, 2014). The
811 reporting triangle defines the observed data as a matrix with the date of the primary
812 event as the rows (t), and the date of the linked secondary event as the columns
813 (d , relative to the time of the primary event). In contrast to previous methods, this
814 frames the data as counts rather than as individual linked events. Primary event
815 incidence $\mathcal{P}_d(\lfloor t \rfloor)$ can then be recovered from this reporting triangle using

$$\mathcal{P}_d(\lfloor t \rfloor) = \sum_{d=0}^D n_{t,d}, \quad (52)$$

where D represents the maximum delay between the primary event and the secondary event which in theory could be infinite. However, in practice, we set it to a finite

value in order to make the model identifiable and computationally feasible. For each t , events are assumed to be drawn from a multinomial distribution with $P_d([t])$ trials and a probability vector $(s_{t,d})$ of length D indicating the probability of a secondary event after a given delay. We model this by estimating the components of this probability vector jointly with the expected number of primary events ($\lambda_t = \mathbb{E}[\mathcal{P}_d([t])]$) at time t . Here, we model the expected number of primary events (λ_t) using an instantaneous daily growth rate model as follows,

$$\lambda_t \sim \text{LogNormal}(\mu_t^l, \sigma_t^l), \quad t_w \leq 1 \quad (53)$$

$$\lambda_t = \lambda_{t-1} \exp(r_0 + \beta t_w), \quad t_w > 1 \quad (54)$$

$$r_0 \sim \text{Normal}(0, 1) \quad (55)$$

$$\beta \sim \text{Normal}(0, 1) \quad (56)$$

816 where the instantaneous growth rate ($r_0 + \beta t_w$) is defined using a linear model with
 817 r_0 as the initial growth rate and β as the linear rate of change in the growth rate
 818 by week (t_w). Note that this is an arbitrary model and that the precise specification
 819 should be appropriate to the modelled setting to avoid bias towards T (Lison et al.,
 820 2023). We assume that the delay distribution follows a daily discretised lognormal

$$p_{t,d} = \frac{F^{\mu_d, \nu_d}(d+1) - F^{\mu_d, \nu_d}(d-1)}{F^{\mu_d, \nu_d}(D+1) + F^{\mu_d, \nu_d}(D)} \quad (57)$$

$$\mu_d \sim \text{Normal}(0, 1) \quad (58)$$

$$\sigma_d \sim \text{Half-Normal}(0, 1), \quad (59)$$

821 where F^{μ_d, ν_d} is the cumulative density function of the lognormal distribution, μ_d
 822 is the log mean, and ν_d is the log standard deviation. Note that this approach to
 823 calculating a probability mass function is equivalent to the interval-reduced method
 824 introduced above and discussed in Section 2.3.4 and Fig. 9.

825 Expected primary events ($n_{t,d}$) by the time of primary event (t) and the time
 826 of the secondary event (d , relative to the time of the primary event) can now be
 827 calculated by multiplying expected primary events for each t with the probability of
 828 the secondary event occurring at a given date ($s_{t,d}$). We assume a negative binomial
 829 observation model to account for potential overdispersion (with a standard half nor-
 830 mal prior on 1 over the square root of the overdispersion (Stan Development Team,
 831 2020)).

$$n_{t,d} \mid \lambda_t, p_{t,d} \sim \text{NB}(\lambda_t \times p_{t,d}, \phi), \quad t = 1, \dots, T. \quad (60)$$

$$\frac{1}{\sqrt{\phi}} \sim \text{Half-Normal}(0, 1) \quad (61)$$

832 2.3.10 Overview table

Method	Details	Truncation	Censoring	Assumptions	Risk of bias
Exact latent variable censoring and truncation correction	Bayesian approach with the exact primary and secondary event times modeled as latent variables (using priors on their distribution over the censoring interval), and with a truncation likelihood integrating over the censoring uncertainty.	Yes	Yes	Exponential growth at a fixed rate within censoring interval.	Potential bias if growth rate changes quickly within the censoring interval.
Approximate latent variable censoring and truncation correction	Approximation of exact latent variable approach that results in a simpler likelihood identical to the plain right truncation case (no integration over primary event times required). Censoring uncertainty is only expressed via priors on the latent even time variables.	Yes	Yes	Primary event time is uniformly distributed between the lower and upper bound, and censoring interval of primary events is sufficiently narrow for the cumulative delay distribution to be approximately constant within its window.	Biased if the primary event times are not uniformly distributed. Small bias from the constant approximation of cumulative delay distribution within the censoring interval.
Interval-reduced censoring and truncation correction	Reformulation of the censoring and truncation likelihood that integrates over the delay (which is a convolution of the primary and secondary event time conditional distributions). Additional assumptions lead to an approximation, which does not require any integration.	Yes	Yes	Delay between primary and secondary events is uniformly distributed within its lower and upper bounds, and censoring interval of primary events is sufficiently narrow for the cumulative delay distribution to be approximately constant within its window.	The assumption of uniformly distributed delays within their bounds is likely unrealistic and deviates from the delay distribution that would result from uniformly distributed event times. Small bias from the constant approximation of cumulative delay distribution within a censoring interval.
Interval-reduced censoring correction	Same likelihood as interval-reduced censoring and truncation, but without conditioning on observation up to present.	No	Yes	Same as interval-reduced censoring and truncation, but additionally assuming no right truncation, i.e. all relevant event pairs have been observed.	Right truncation bias (leading to underestimation of mean) in addition to the biases described above.

Method	Details	Truncation	Censoring	Assumptions	Risk of bias
Truncation correction	Likelihood for delay distribution that uses lower bounds of event times (or different point in censoring interval, e.g. midpoint), conditioned on observation up to present.	Yes	No	Assumes no censoring, i.e. that true event times match the assumed event times within censoring interval.	Censoring bias (leading to over-estimation of standard deviation)
Filtering	Exclusion from the likelihood of observations with primary event after some chosen filtering time, and likelihood for delay distribution that uses lower bounds of event times (or different point in censoring interval, e.g. midpoint).	Yes	No	Assumes no censoring, i.e. that true event times match the assumed event times within censoring interval.	Truncation bias if filtering time is too late, censoring bias.
Filtering + censoring correction	Same as filtering, but with a likelihood that accounts for interval-reduced censoring.	Yes	Yes	Assuming interval-reduced censoring (i.e., uniformly distributed delay within the upper and lower bounds).	Truncation bias if filtering time is too late and biases from interval-reduced censoring.
Naïve estimation	Likelihood for delay distribution that uses lower bounds of event times (or different point in censoring interval, e.g. midpoint), not conditioned on observation up to present.	No	No	Assumes no right truncation, i.e. all relevant event pairs have been observed, and no censoring, i.e. that true event times match the assumed event times within censoring interval.	Both right truncation bias and censoring bias.

Method	Details	Truncation	Censoring	Assumptions	Risk of bias
Dynamical censoring correction (discrete)	Likelihood motivated from modeling the backward delay distribution under epidemic dynamics, tailored to the case of a discrete incidence time series.	Partially (depending on assumed incidence time series)	Yes	Discretisation of continuous incidence to dates (step function), centered on the lower bounds of primary and secondary events, respectively, and delay between primary and secondary event is uniformly distributed.	The assumption of uniformly distributed delays is likely unrealistic and deviates from the delay distribution that would result from uniformly distributed censoring times. Biased if the estimate of incidence time series is biased. Can also have a small discretisation bias.
Joint incidence and forward delay estimation (discrete)	Latent-variable nowcasting model, using a count likelihood for case counts by primary event and delay, with an epidemiological prior on the case counts by primary event (e.g. exponential growth model), and interval-reduced censoring of the delays.	Yes	Yes	Assumes that case counts in the cells of the reporting triangle have a certain count distribution, that incidence follows a certain time series model, and that delay between primary and secondary event is uniformly distributed.	The assumption of uniformly distributed delays is likely unrealistic and deviates from the delay distribution that would result from uniformly distributed censoring times. Potential bias if the epidemiological model for incidence is inappropriate.

833 2.4 Simulation study

834 In order to understand the impact of different biases and the performance of the
835 various methods that we have outlined, we conducted a series of simulation studies
836 designed to replicate different real-world scenarios.

837 Summary

- 838 • We evaluated the methods discussed above across a range of simulated expo-
839 nential growth and epidemic wave scenarios.
- 840 • To simulate exponential growth settings, we used a continuous-time individual-
841 based exponential growth model and to simulate epidemic wave settings, we
842 used a stochastic Susceptible-Infected-Recovered model.
- 843 • As this study was initially inspired by the analysis of COVID-19 outbreaks,
844 we considered settings with a range of delay distributions to assess the impact
845 of common biases across a range of common scenarios for COVID-19. These
846 delays had the following means (and standard deviations) on the natural scale:
847 3.6 (1.5) days, 5.9 (3.9) days, and 8.3 (7.9) days. We refer to these in the text
848 as “short”, “medium”, and “long” delays.

849 **Simulating infections** We explored both fixed growth rate and epidemic scenar-
850 ios. For all scenarios, we explored 3 lognormal distributions representing “short” (log
851 mean of 1.2 and log standard deviation of 0.4), “medium” (log mean of 1.6 and log
852 standard deviation of 0.6), and “long” (log mean of 1.8 and log standard deviation of
853 0.8) time delays. These assumptions correspond to distributions with the following
854 means (and standard deviations) on the natural scale: 3.6 (1.5) days, 5.9 (3.9) days,
855 and 8.3 (7.9) days (Fig. 4B).

856 To simulate settings with a fixed growth rate, we used continuous-time stochas-
857 tic simulations of exponential growth with a range of daily growth rate assump-
858 tions (assuming 10,000 individuals): “fast decay” (-0.2), “decay” (-0.1), “stable” (0),
859 “growth” (0.1), and “fast growth” (0.2).

860 To simulate epidemic scenarios, we used a stochastic Susceptible-Infected-Recovered
861 (SIR) model implemented using a Gillespie algorithm assuming an early exponential
862 growth rate of 0.2 per day and a recovery rate of 1/7 per day (corresponding to a
863 mean of 7 days) with 50 initial cases and a total population of 10,000.

864 For all scenarios, we took the primary event time to be the time of infection,
865 and we simulated secondary events by drawing delays from the assumed continuous
866 lognormal distribution for that scenario for each individual and then adding this
867 delay to their primary event time. This simulation approach assumes no observation
868 error for the primary event or secondary events beyond daily censoring and thus
869 represents an idealised system.

870 **Scenarios investigated** For the simulations with fixed growth rates, we assumed
871 a sample size of 200 randomly drawn pairs of primary and secondary events with
872 observation cut-off 30 days after the start of the simulation. We repeated each
873 sampling step 20 times independently so that we ended up with 20 replicates of each
874 simulated growth rate scenario.

875 For the epidemic wave simulation, we explored a range of sample sizes (10, 100,
876 200, and 400 samples) and observation time scenarios (including all observations up
877 to 15 days, 30 days, 45 days, and 60 days from the start of the simulation).

878 **2.5 Case study: 2014-2016 Sierra Leone Ebola virus disease** 879 **epidemic**

880 In order to explore the performance of each method in a real-world setting, we used
881 publicly available data from the 2014-2016 Sierra Leone Ebola virus disease epidemic.

882 **Summary**

- 883 • We studied the evolution over time of the empirically observed backward and
884 forward delay distributions of the time from symptom onset to positive test, and
885 the proportion of positive tests that were unobserved for a cohort of symptom
886 onsets over a rolling 60-day observation period.
- 887 • We used multiple observation times and compared retrospective estimates to
888 real-time estimates for each approach.

889 **Data** We downloaded line list data from Fang et al. (2016) which contained age,
890 sex, symptom onset date, sample test date, the district of the case, and the Chiefdom
891 of the case for Ebola virus disease (EVD) cases in Sierra Leone from May 2014
892 through September 2015. We then processed these data to keep only the date of
893 symptom onset and the date of the sample test. We assumed that censoring for each
894 of these dates was daily with a day defined as being from 12:00 AM to 11:59 PM.

895 **Empirical context** We calculated various summary statistics using the available
896 line list data, which we used to provide a context for comparisons of different infer-
897 ence methods. These statistics included changes in the mean forward and backward
898 distributions, the empirical forward distribution at each observation cut-off date, as
899 outlined in the 'Scenarios Investigated' section, in both real-time and retrospective
900 settings; and, the proportion of secondary events that were unobserved for a cohort
901 of individuals whose primary events took place within a rolling 60-day window.

902 **Scenarios investigated** We estimated the delay from symptom onset to sample
903 test using each method across four different observation windows of 60 days each
904 (0–60 days, 60–120 days, 120–180 days, and 180–240 days after the first symptom

905 onset). In particular, for each window, we only considered individuals who developed
906 symptoms and also got tested within the time period to match a real-time analysis
907 setting. Note that in this case study we do not account for additional potential
908 delays from the sample test date to the reporting of these tests as these data were
909 not available in Fang et al. (2016). However, our methods could naturally account for
910 these additional delays if data were available. As a comparison, we then re-estimated
911 the delay using each method for each observation period by including all individuals
912 who developed symptoms within the observation period (regardless of when they
913 got tested) and estimated the delay distribution. These retrospective estimates were
914 then used to represent the "true" distribution when calculating the relative difference
915 between estimates using data available in real-time and data available retrospectively
916 for each method. Note that in this setting the "true" distribution may still be biased
917 relative to the underlying forward distribution (e.g. methods that do not account
918 for censoring may still be biased in both real-time and retrospective settings).

919 2.6 Evaluation

920 We evaluated the recovery of the mean and standard deviation of the lognormal
921 distribution in our simulated scenarios visually using the posterior density normalised
922 with known synthetic values, where these were available, and quantitatively using the
923 Continuous Ranked Probability Score (CRPS, Gneiting and Raftery, 2007), which
924 is a generalisation of absolute error. The CRPS measures the distance between a
925 distribution and data as follows

$$\text{CRPS}(F, y) = \int_{-\infty}^{\infty} (F(x) - 1(x \geq y))^2 dx, \quad (62)$$

926 where y is the true observed value and F is the cumulative distribution function
927 (CDF) of the predictive distribution, and $1(x \geq y)$ is the indicator function such
928 that for $x \geq y$ its value is 1 and otherwise it is 0. The CRPS is a strictly proper
929 scoring rule, meaning that only a probabilistic estimate that exactly reflects the true
930 distribution of y minimises the expected score.

931 To allow for comparisons across simulated scenarios, we normalised predictions
932 by the known true value, took the natural log, and then calculated the CRPS. In
933 effect, this transforms the CRPS into a relative, rather than an absolute, score whilst
934 maintaining its propriety (Bosse et al., 2023). We then averaged across all obser-
935 vations, and then averaged across observation stratified by growth rate, parameter,
936 and delay distribution.

937 As a comparison, we also compute relative bias, root mean squared error (RMSE),
938 and coverage probabilities. The relative bias and RMSE were calculated using the
939 log of relative errors (i.e., the ratio between the estimated and true values) in the
940 same way as the CRPS. The coverage probability was calculated from the proportion
941 of 90% credible intervals that contain the true value.

942 2.7 Implementation

All models were implemented using the `brms` (Bürkner, 2018) R package and the `stan` probabilistic programming language (Stan Development Team, 2021) (Gabry and Češnovar, 2021) in R version 4.2.2 (R Core Team, 2019). As such they are readily extensible to model formulations covered by the GAM framework. Where not otherwise specified, we use the following priors for all methods,

$$\mu_d \sim \text{Student's } t(3, 0, 2.5) \quad (63)$$

$$v_d \sim \text{Student's } t(3, 0, 2.5), \quad (64)$$

943 where μ_d is the log mean, and v_d is the log standard deviation of the lognormal
944 distribution. Here, the student's t distribution has the following three parameters:
945 degrees of freedom, mean, and standard deviations.

946 The No-U-Turn Sampler (NUTS), which is an adaptive variant of Hamiltonian
947 Monte Carlo (HMC) was used for model fitting via `cmdstanr`. We used four Markov
948 chain Monte Carlo (MCMC) chains with each having 1000 warm-up and 1000 sam-
949 pling steps (Gabry and Češnovar, 2021); we did not consider any other sampling
950 schemes or settings for simplicity. We set `adapt_delta`, which represents the tar-
951 get average acceptance probability, to 0.95 due to the expected complexity of the
952 posterior distribution (Betancourt, 2017).

953 We assessed convergence using the Rhat diagnostic, where an R-hat close to 1
954 indicates that the MCMC chains have converged (we used 1.05 as a threshold for
955 further investigation) (Gelman and Rubin, 1992), and recorded run-time, the number
956 of divergent transitions, the exceedance of the maximum tree depth (which was 10,
957 its default setting), and the effective sample size (Gabry and Češnovar, 2021; Stan
958 Development Team, 2021). We did not assess the sensitivity of our inference to these
959 settings. Divergent transitions are useful as a model diagnostic when using HMC
960 since they indicate issues exploring the posterior, and when present, can mean that
961 estimates are unreliable (Betancourt, 2017).

962 The `scoringutils` R package (1.1.0) (Bosse et al., 2022) was used to calculate
963 the CRPS, RMSE, and to assess bias. We also evaluated the sampling performance
964 details of each method, including the run-time of inference and the distribution of
965 divergent transitions.

966 All core functionalities and inference approaches were implemented in the `epidist`
967 R package to facilitate reuse and user extension. Our simulation, scenario, and model
968 fitting pipeline was implemented in a `targets` workflow (Landau, 2021) and we pro-
969 vide an archive of our results to aid reuse. Our post-processing pipeline was also
970 implemented as a `targets` workflow.

971 To enhance the reproducibility of this analysis, we managed dependencies using
972 the `renv` R package (Ushey, 2021). In addition, we provide a versioned Dockerfile
973 and a prebuilt archived image (Boettiger, 2015). The code for this analysis can
974 be found here: <https://github.com/parksw3/epidist-paper>. The code for the
975 `epidist` R package can be found here: <https://github.com/epinowcast/epidist>.

976 3 Results

977 3.1 Simulation study

978 Summary

- 979 • Not accounting for the truncation bias led to an underestimation of the mean.
980 The degree of bias increased with the underlying growth rate of an epidemic.
- 981 • Not accounting for the censoring bias generally led to over-estimation of the
982 standard deviation. For daily reporting, assuming uniform censoring for each
983 event gave reasonable answers (i.e., the interval-reduced approach).
- 984 • Among methods that explicitly account for the censoring and truncation bias in
985 the forward distribution, the approximate-latent-variable approach performed
986 the best overall, followed by the interval-reduced-censoring-and-truncation ap-
987 proach.
- 988 • The joint modeling and dynamical correction methods had nearly compara-
989 ble performance. However, the joint modeling approach had wider uncertainty
990 than comparable approaches whilst the dynamical correction method was un-
991 reliable during negative growth periods and required a retrospective time series
992 to perform well.

993 3.1.1 Exponential growth simulation

994 The joint modelling approach performed best overall when evaluated using relative
995 CRPS (Fig. 3C), followed by the approximate-latent-variable model, the interval-
996 reduced censoring method, and the retrospective dynamical correction approach.
997 All of these approaches successfully recovered the simulated parameters to a rea-
998 sonable degree both visually (Fig. 3C), based on both probabilistic and point scores
999 (Fig. 10A), and based on coverage of the 90% credible interval (Fig. 10C). These
1000 methods performed worse for long delays with both an increase in uncertainty and
1001 in underestimation of both the mean and standard deviation. These biases were
1002 a particular issue for higher growth rates. The approximate-latent-variable model
1003 appeared most robust to this bias visually, though this was not conclusively demon-
1004 strated by any of the evaluation metrics used. The dynamical correction method us-
1005 ing retrospective incidence data was the least robust of the methods that performed
1006 well when growth rates were negative with biased estimates of the standard deviation
1007 across all delay distributions explored. Unlike other well-performing methods, the
1008 joint modelling approach had less variance in its relative CRPRs scores without the
1009 very low scores for large growth rates and small delays scenarios but also without
1010 reduced performance for scenarios with longer delays for the mean (though not the
1011 standard deviation).

1012 The dynamical correction method with real-time incidence performed better than
1013 other methods that do not fully account for censoring and truncation. However,
1014 it performed worse than all other methods that account for these biases. In the
1015 real-time setting, the dynamical correction method had similar characteristics as
1016 when used with retrospective data, but showed more under- (when growth rates
1017 were positive) and over- (when growth rates were negative) estimation of the mean
1018 due to the use of truncated real-time incidence. Real-world performance for this
1019 approach would be somewhere between the real-time and retrospective approaches,
1020 depending on the availability of robust estimates for primary incidence that correct
1021 for truncation (or alternatively a robust estimate of the growth rate in stable growth
1022 settings).

1023 All other models perform considerably worse than these methods with methods
1024 that only accounted for either truncation or censoring performing the worst. The
1025 naive method that accounts for no biases outperformed these methods as censoring
1026 and truncation biases partially cancelled each other when growth was fast (as the di-
1027 rection of these mechanisms bias estimates in different directions). Out of the poorly
1028 performing methods, the filtering methods performed the best with the potential to
1029 perform even better if the filtering horizon were optimised to the length of the delay
1030 (though at the cost of increased uncertainty). All of these methods under-covered, in-
1031 dicating that their credible intervals were too narrow, and produced biased estimates
1032 (Fig. 10).

1033 Methods that did not adjust for right truncation produced increasingly under-
1034 estimated means and standard deviations as the exponential growth rate increased
1035 ($r > 0$) with the degree of underestimation increasing with the length of the delay
1036 distribution. Whilst methods that do explicitly account for the truncation bias gen-
1037 erally give unbiased estimates of the mean (Fig. 3C) when the growth rate is high
1038 and the delay is longer (i.e., when the degree of truncation is large), the truncation
1039 method without censoring overestimated the mean and standard deviation. Better
1040 performing methods also struggled in this extreme setting with larger credible inter-
1041 vals for the standard deviation though the estimates were visually still close to the
1042 true values.

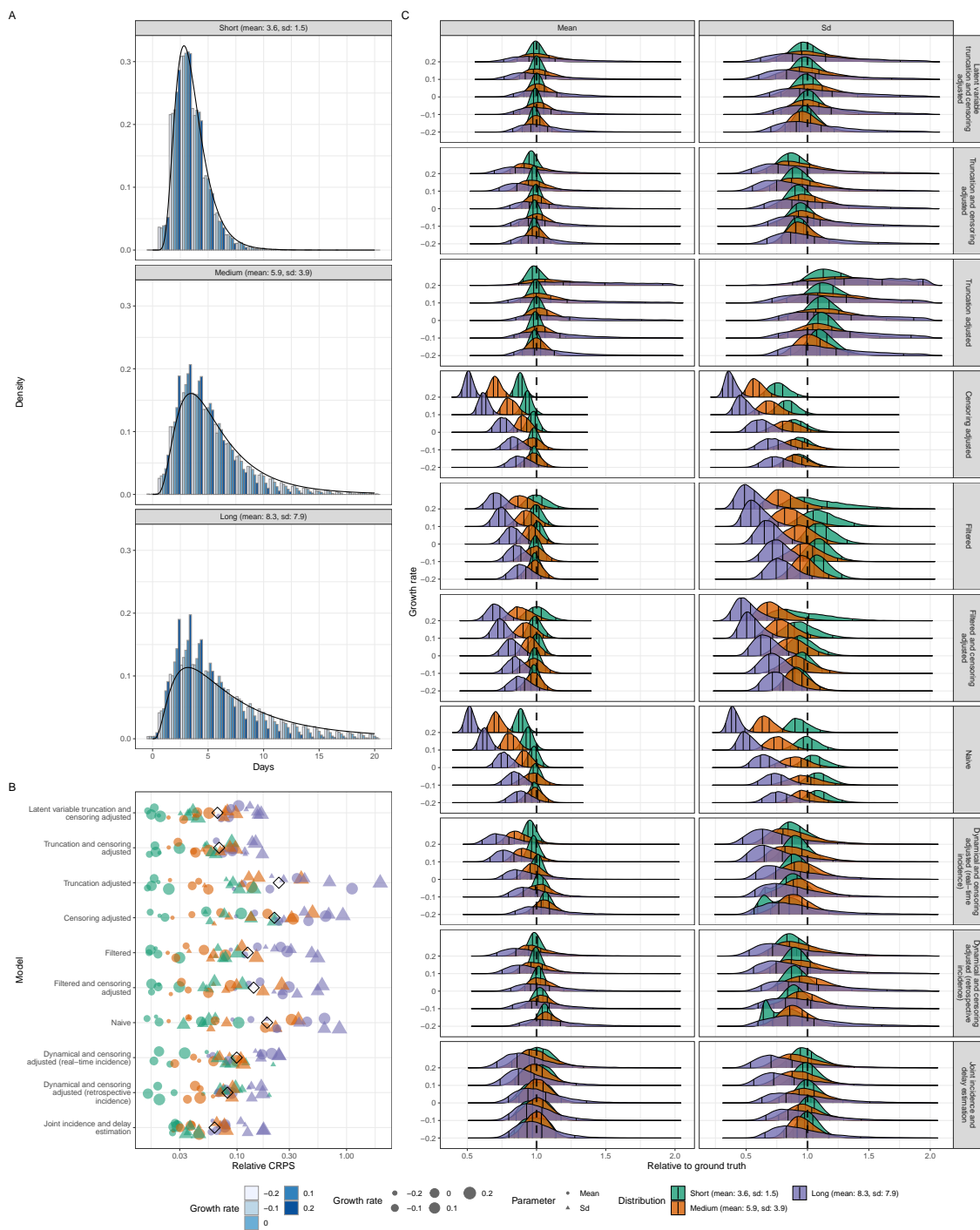


Figure 3

Figure 3: Simulation study with fixed growth rates. (A) Empirical distribution of the delay between primary and secondary events stratified by distribution. Shade indicates the exponential growth rate used in each simulated scenario. The black line indicates the true distribution used for simulation. (B) Relative CRPS of each method (smaller is better). Here black diamonds represent the global relative CRPS. Coloured points represent the mean CRPS for each distribution scenario (short, medium, and long) with the shape indicating if the score is for the mean (circle) or standard deviation (SD, triangle). (C) Posterior distributions of mean and standard deviation (Sd) relative to the values used when simulating. The vertical line at 1 indicates exact replication of the true value. Vertical lines represent the 5%, 35%, 65%, and 95% quantiles respectively. Models are ordered based on the order used in the methods section.

1043 3.1.2 Epidemic wave simulation

1044 For the epidemic wave simulation, the ordering of well-performing methods differed
1045 from that of the exponential growth simulations but the performance difference be-
1046 tween methods that did and did not account for both censoring and right truncation
1047 remained. Here the approximate-latent-variable approach performed the best overall
1048 (Fig. 4C) based on relative CRPS, followed by the dynamical approach using ret-
1049 rospective incidence, the joint modeling approach, and finally the interval-reduced-
1050 censoring-and-truncation approach.

1051 Early in the epidemic, all of these methods perform worse when combined with
1052 longer delays. This setting was particularly problematic for the approximate-latent-
1053 variable and the interval-reduced-censoring-and-truncation approaches (Fig. 4D),
1054 giving wider uncertainty intervals. Underestimation of the standard deviation was
1055 more common in the epidemic wave simulations across observation windows than it
1056 was in the exponential growth scenarios for these methods, with the approximate-
1057 latent-variable method being the least impacted.

1058 As in the exponential growth simulations, the dynamical correction approach with
1059 real-time incidence performed less well than other methods that sought to account
1060 for both right truncation and censoring due to the use of a truncated time series.
1061 Methods that did not account for both censoring and truncation also give biased
1062 estimates, especially at the beginning of the epidemic (i.e. when exponential growth
1063 was highest).

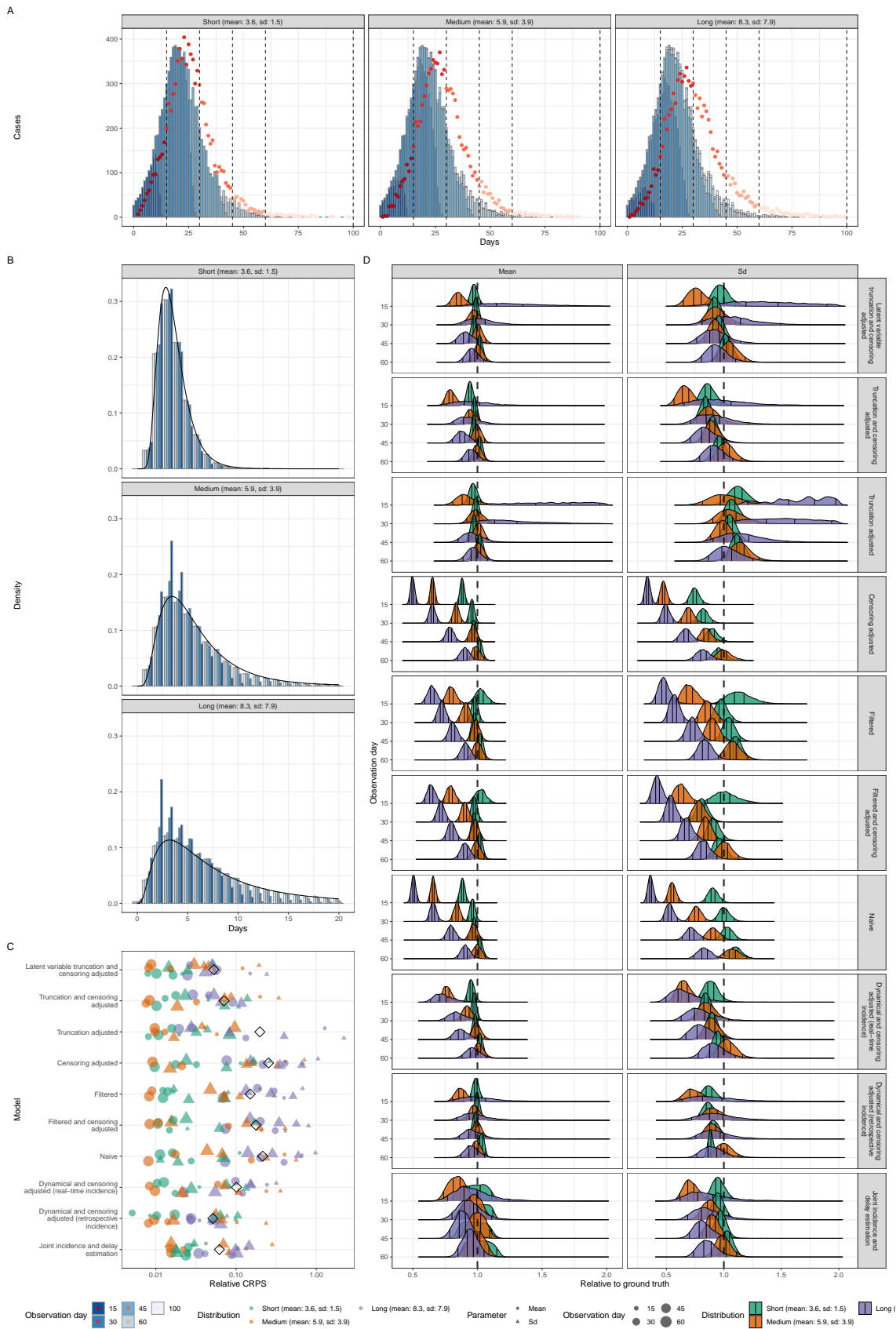


Figure 4

Figure 4: (A) Epidemic wave simulations across three distributions. Dashed bars represent four observation days. Blue bars represent the observed incidence of primary events on each observation day. Red points represent the observed incidence of secondary events on each observation day. On each observation day, we take 400 random samples from simulated primary and secondary event pairs and fit lognormal distributions while accounting for truncation and censoring biases. (B) Empirical distribution of the delay between primary and secondary events stratified by distribution. Shade indicates the day of observation. The black line indicates the true distribution used for simulation. (C) Relative CRPS of each method (smaller is better). Here black diamonds represent the global relative CRPS. Coloured points represent the mean CRPS for each distribution scenario with the shape indicating if the score is for the mean (circle) or standard deviation (Sd, triangle). (D) Posterior distributions of mean and standard deviation (Sd) relative to the values used when simulating. The vertical line at 1 indicates exact replication of the true value. Vertical lines represent the 5%, 35%, 65%, and 95% quantiles respectively. Models are ordered based on the order used in the methods section.

1064 **3.2 Case study: 2014–2016 Sierra Leone Ebola virus disease** 1065 **epidemic**

1066 **Summary**

- 1067 • The epidemic had four distinct phases: sporadic cases in the first phase, rapid
1068 growth in the second phase, plateau in the third phase, and decay in the fourth
1069 phase.
- 1070 • The delay in reporting cases (from symptom onset) increased as the number
1071 of cases grew, reaching a maximum mean delay of around 7 days in the third
1072 phase and then declining, potentially reflecting changes in testing capacity,
1073 reporting, or other mechanisms linked to incidence.
- 1074 • The forward and backward delay distributions exhibited similar trends, but
1075 with differences in mean delay during the growth and decay periods. Right
1076 truncation during the period of faster growth resulted in a difference of ap-
1077 proximately 0.5 days in the mean delay between real-time and retrospective
1078 observations. The degree of right truncation was relatively small due to the
1079 growth rate of this epidemic also being relatively small even during the peak
1080 growth period.
- 1081 • Methods that considered right truncation and censoring performed well, pro-
1082 ducing real-time estimates that closely matched retrospective estimates, similar
1083 to the findings in the simulation studies.
- 1084 • Differences between real-time and retrospective estimates were largest during
1085 the period of exponential growth. All well-performing methods slightly overes-
1086 timated the retrospective standard deviation in real-time during this period.
- 1087 • The joint modeling approach had larger credible intervals compared to other
1088 methods that accounted for right truncation and censoring and also slightly
1089 overestimated the retrospective mean in real-time, this was observed to some
1090 degree in other methods to a lesser extent, except for the dynamic correction
1091 method using retrospective incidence data.

1092 **3.2.1 Empirical observations**

1093 During the first phase of the epidemic, cases were reported sporadically without
1094 apparent growth (Fig. 5A). The epidemic then grew rapidly during the second phase,
1095 followed by a plateau during the third phase and decay during the fourth phase. For
1096 each phase, we used all available samples from the last 60 days which resulted in
1097 real-time (and retrospective) sample sizes of 834 (1032) at 60 days, 3149 (3532) at
1098 120 days, 2399 (2483) at 180 days, and 401 (428) at 240 days.

1099 The forward delay distribution (as observed retrospectively) changed over the
1100 course of the epidemic (Fig. 5B), potentially reflecting changes in the reporting pro-
1101 cess. During the initial phase, the delays were short with means around 5 days. As
1102 the number of cases increased, the mean delay also started to increase, potentially
1103 reflecting an overload in a testing capacity or another mechanism linked to incidence.
1104 During the third phase, the mean delay reached its maximum of around 7 days and
1105 started to decline along with incidence. Eventually, the mean delay decreased to
1106 5.9 days by day 240. The backward distribution exhibited a similar trend, though
1107 as explained earlier, it had a shorter mean than the forward distribution during the
1108 growth period and a longer mean during the decay period.

1109 The empirical distributions for each observation period, both as observed in real-
1110 time and retrospectively, are presented in Fig. 5C. Delay distributions observed in
1111 real-time (truncated) and retrospectively (untruncated) were broadly similar across
1112 all observation windows, except for days 60–120 during the period of rapid growth
1113 in incidence. During this period, the difference in the mean delay was around 0.5
1114 days with this difference reflecting large amounts of right truncation (Fig. 5D).

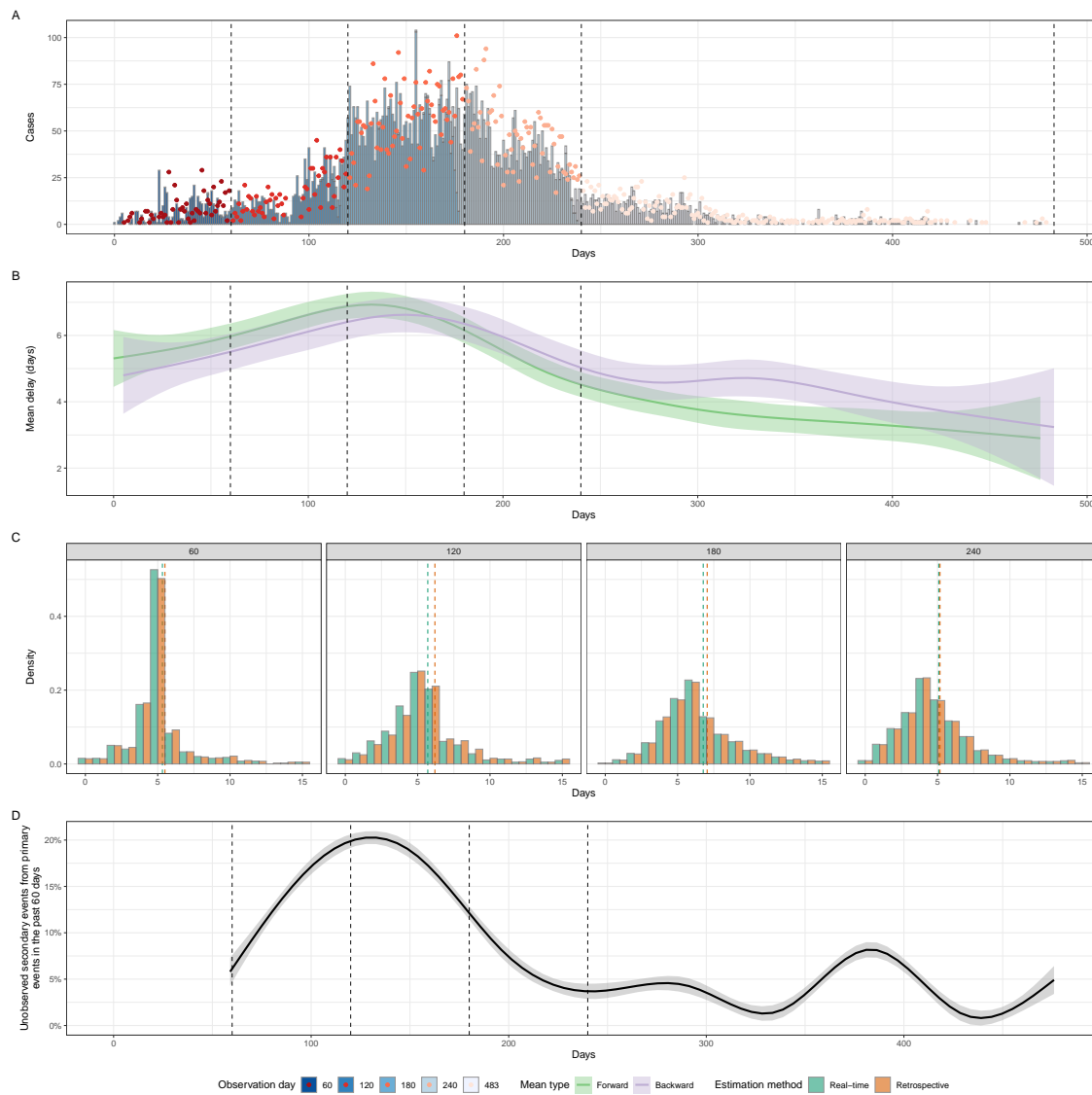


Figure 5: (A) Reported number of Ebola virus disease cases between May 18, 2014 and September 13, 2015. Dashed lines represent four observation days. Blue bars represent the daily incidence of primary events with different shades indicating the observed incidence in each observation period. Red points represent the daily incidence of secondary events with different shades indicating the observed incidence in each observation period. (B) Mean forward (measured across a cohort of individuals who developed symptoms on the same day) and backward (measured across a cohort of individuals who reported their infections on the same day) delays between symptom onset and case reports and the corresponding 95% confidence intervals. A Generalised Additive Model (GAM) with a default thin plate spline on the date of the event has been used to smooth daily estimates. (C) Empirical distribution of the delay between primary and secondary events both in real-time (green) and retrospectively (orange) stratified by observation. The dotted lines indicate the observed mean both in real-time (green) and retrospectively (orange). (D) Proportion of unobserved secondary events (and therefore, delays) from primary events in the last 60 days. A GAM with a default thin plate spline on the date of the primary event has been used to smooth daily estimates with the point estimate and its 95% confidence interval shown.

1115 3.2.2 Model results

1116 The joint modelling approach, the approximate-latent-variable model, the interval-
1117 reduced censoring method, and the retrospective dynamical correction approach
1118 again performed well in real-time, producing estimates that were comparable to those
1119 estimated using retrospective data (Fig. 6A, Fig. 12A). As expected, the largest dif-
1120 ferences between retrospective and real-time estimates are observed during the period
1121 of exponential growth (i.e., 60-120 days). During this period, all of these methods'
1122 real-time estimates overestimated the retrospective estimates of the standard devia-
1123 tion. The joint modeling approach had significantly wider credible intervals across all
1124 observation periods compared to other well-performing methods. The joint modeling
1125 approach also routinely overestimated the retrospective mean in real-time settings.
1126 We note that the same overestimation occurred to some degree for all other methods
1127 aside from the dynamical correction method using retrospective incidence data.

1128 These well-performing methods also generally captured the retrospective empir-
1129 ically observed mean well. However, the estimated standard deviations are con-
1130 siderably lower than the empirical values (Fig. 6A). While these differences likely
1131 reflect the bias in the empirical values caused by the censoring process, it is also
1132 possible that the lognormal distribution may not be the best choice of distribution
1133 for these data. The interval-reduced and dynamic correction methods particularly
1134 underestimated the empirically observed standard deviation.

1135 As in our simulation scenarios, not accounting for truncation gave real-time fits
1136 with lower estimates of the mean and standard deviation than the corresponding
1137 retrospective fit (Fig. 6A, Fig. 12A). As both retrospective and real-time fits are
1138 liable to censoring bias, the impact of not properly accounting for censoring is not
1139 highlighted by this case study. This also means that the estimates for the standard
1140 deviation from the truncation-only adjusted model should best reflect the empirical
1141 standard deviation, as both are biased upwards due to censoring when compared to
1142 the standard deviation of the continuous distributions.

1143 All methods were able to reasonably match the empirical mean of the data in
1144 each observation period despite assuming a constant mean and standard deviation
1145 within each period (Fig. 6B). Due to the relatively slow rate of exponential growth
1146 for all observation periods, the absolute differences between the estimated means for
1147 most methods were relatively small.

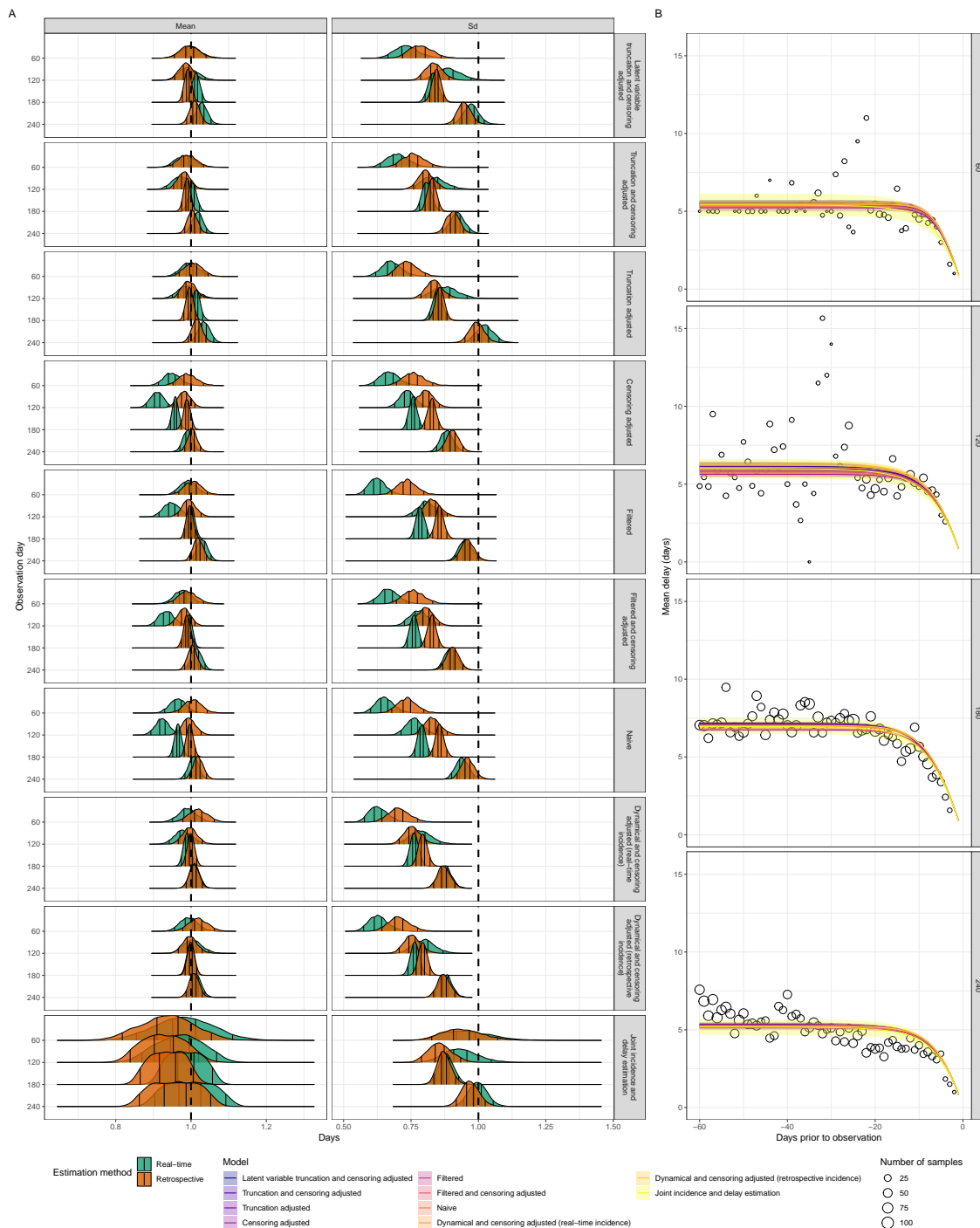


Figure 6: (A) Posterior distributions of mean and standard deviation (Sd) normalised by the retrospectively observed mean and standard deviation from the data. Vertical lines represent the 5%, 35%, 65%, and 95% quantiles respectively. Dashed vertical lines correspond to the unbiased estimate. Models are ordered based on the order used in the methods section. Recovery of the empirical standard deviation is not expected as the empirical values are biased due to the censoring process. (B) Posterior predictions of the truncated mean over time compared to the observed forward mean. Lines and shaded regions represent posterior median and 95% credible intervals. Points represent the observed mean values.

1148 3.2.3 Implementation considerations

1149 Summary

- 1150 • More complex methods for estimating epidemiological delay distributions have
1151 increased computational requirements relative to simpler methods. In most
1152 settings, these computational requirements are expected to still be feasible for
1153 routine usage.
- 1154 • The approximate-latent-variable method had exponential scaling in resource
1155 requirements when the sample size increased.
- 1156 • The interval-reduced-censoring-and-truncation method required the least com-
1157 putational resources by an order of magnitude of the methods that accounted
1158 for both censoring and truncation.
- 1159 • The dynamical adjustment method is unstable for short delays with larger sam-
1160 ple sizes both with and without a truncated time series, though the instability
1161 increased when a truncated time series was used (i.e., a real-time one).

1162 All the methods we considered are implementable with modest (i.e., laptop scale)
1163 computational hardware at the time of writing, and we are able to run our full anal-
1164 ysis pipeline of several thousand model fits within this resource budget at a practical
1165 time scale (i.e, within several hours). However, more complex methods required
1166 greater computational resources. For example, the approximate latent variable, dy-
1167 namic correction, and joint modeling approaches required an order of magnitude
1168 more resources for the same sample size than the interval-reduced-censoring-and-
1169 truncation method (Fig. 7A). The resource requirements for all models scaled with
1170 sample size, with the approximate latent variable model scaling the worst of the
1171 methods we explored. The dynamical correction approach had the highest variance
1172 in its computational requirements with some fits taking 10 to 100 times longer de-
1173 spite having the data used having the same sample size across fits. This was likely
1174 due to numerical instability from the integration step. It was a particular issue when
1175 a real-time incidence time series was used.

1176 In general, all the methods were numerically stable, excluding the dynamical cor-
1177 rection approach in some settings, and converged. Methods that better captured the
1178 data-generating process (i.e. that accounted for known biases) were the most stable
1179 and had fewer diagnostic warnings. A notable exception to this was the dynamical
1180 model which was the only model to cause sampling in `stan` to fail completely in six
1181 instances and more generally had the highest proportion of fits with divergent tran-
1182 sitions (Fig. 7B). These issues occurred most frequently in simulations with short
1183 delays and during periods of epidemic decline. In the case study, both retrospective
1184 and real-time fits failed at the 180-day observation point. Another driver of model
1185 issues was low sample sizes with the majority of instances of divergent transitions
1186 occurring in settings with a sample size of 10.

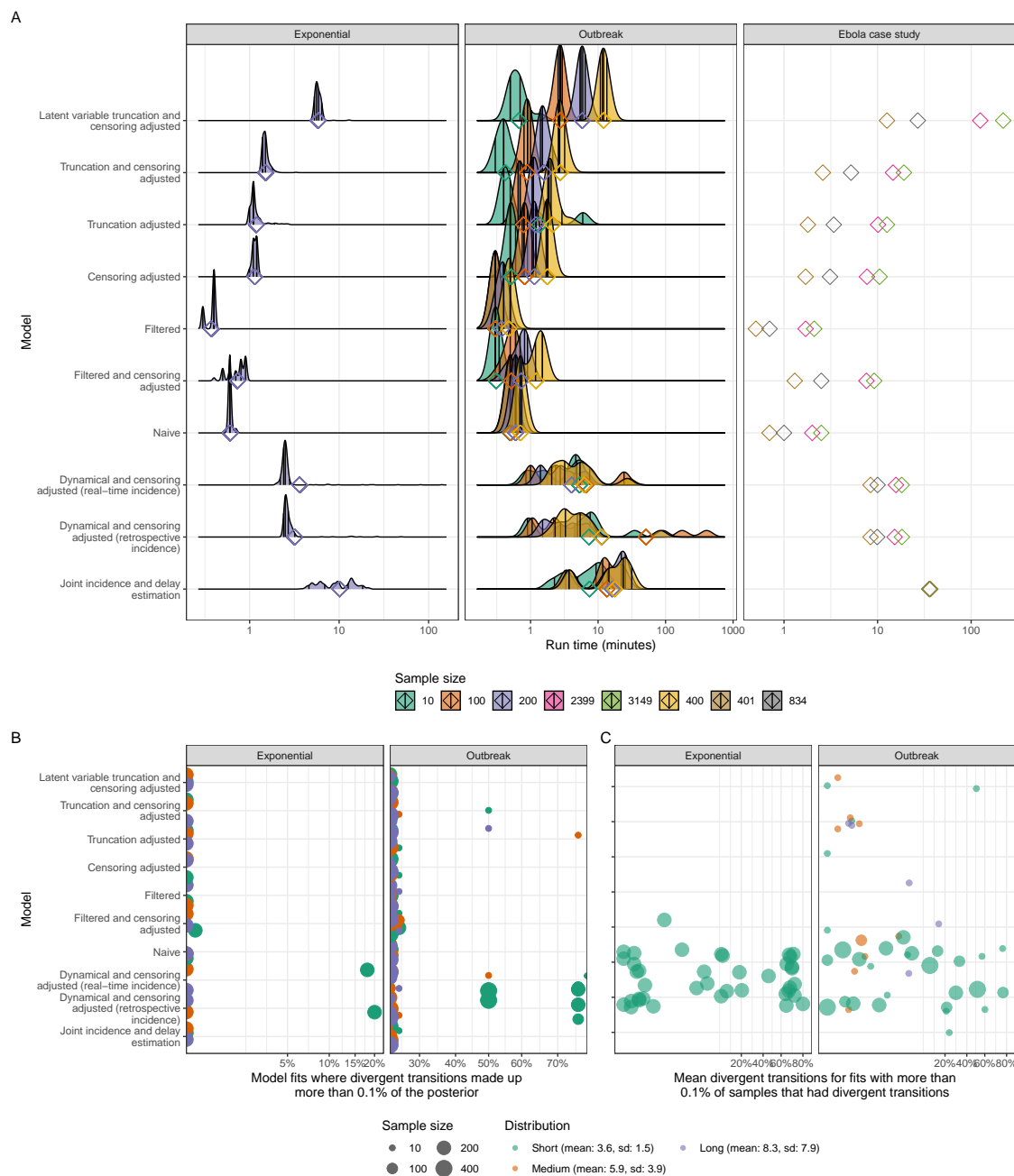


Figure 7: (A) Distributions of run-times for model fitting using a 2019 16-core AMD Threadripper on the log 10 scale. Diamonds represent the overall mean run times and vertical lines represent the 5%, 35%, 65%, and 95% quantiles respectively. Whilst run times are specific to the hardware used, the relative differences between models and scenarios should be more readily generalisable to other hardware. (B) Percentage of model fits with more than 0.1% divergent transitions on the logit scale. (C) Percentage of divergent transitions for fits with more than 0.1% of samples that had divergent transitions on the logit scale.

1187 4 Discussion

1188 4.1 Findings

1189 In this work, we provided methodological and practical guidance for researchers
1190 tasked with estimating epidemiological delay distributions. We first introduced the
1191 general theory of epidemiological delay distributions and the most common kinds of
1192 biases (namely censoring, truncation, and dynamic) that can impact their inference.
1193 Based on this theory, we derived an exact approach for accounting for these biases
1194 when estimating epidemiological delay distributions. As this approach lacks stability
1195 and practicality for real-time usage, we then presented a set of methods that approxi-
1196 mate the exact solution, and compared their performance. We made use of simulated
1197 scenarios and a case study using data from the 2014-2016 Sierra Leone Ebola virus
1198 disease epidemic to compare these methods in applied usage, evaluating not only
1199 their accuracy and calibration but also practical issues, such as their suitability for
1200 real-time usage, their computational requirements, and their numerical stability.

1201 We showed that naive methods that correct for none or only one form of bias
1202 can severely misestimate the distribution mean, e.g. by up to 50% in the most ex-
1203 treme cases we studied. Generally, we suggest using the approximate latent variable
1204 model Ward et al. (2022), which explicitly adjusts for both censoring and trunca-
1205 tion biases, and can produce much more accurate estimates both in real-time and
1206 retrospectively. However, this method was not foolproof and could not estimate
1207 the distribution mean or standard deviation with precision when the exponential
1208 growth rate was very high and the true delay distribution was long. If this method is
1209 computationally too costly or complex to implement, using the interval-reduced cen-
1210 soring and truncation model (which assumes a 2-day censoring window around each
1211 observed delay) gives only slightly more biased estimates. The joint modeling ap-
1212 proach and dynamic correction methods also performed well (in the latter case only
1213 when a retrospective time series was available) and may be sensible choices in some
1214 settings. In the case of the joint model, these settings are likely to be those in which
1215 a nowcast of the primary event is useful, where the primary event has observation
1216 error, or when a hazard-based approach would allow additional complexities in the
1217 reporting process to be modelled. However, this model's requirement for a primary
1218 incidence model to be specified may make it more complex to use more generally.
1219 Similarly, the dynamical correction approach's requirement for an untruncated time
1220 series generally means another model would first need to be used to estimate the
1221 incidence time series. This process is likely to introduce hard-to-quantify bias and
1222 make propagating uncertainty difficult (in the absence of a similar joint approach we
1223 have explored for the forward distribution). However, where an independent time
1224 series is available that does not suffer from right truncation this approach would be
1225 more practical. Finally, we provide all the methods we have evaluated as a stan-
1226 dalone, and readily extensible, R package (`epidist`), which additionally provides
1227 functionality to fit distributions that are partially pooled and that vary in discrete

1228 time.

1229 Right-truncation bias was most pronounced for growing epidemics and long de-
1230 lays. Failing to account for this bias typically led to an underestimation of the mean.
1231 This underestimation increased with higher degrees of truncation. We have discussed
1232 several methods that can robustly adjust for right truncation, but when the degree of
1233 right truncation was very large, even these well-performing methods overestimated
1234 the mean. Though their credible intervals still covered the true mean in all cases,
1235 performance could likely be improved by using a more informed prior distribution.
1236 For this reason, it may be helpful for practitioners to track the degree of expected
1237 right truncation in real time. Unfortunately, this was difficult to do, but one heuristic
1238 approach is to plot the forward mean (Fig. 5B) as well as the proportion truncated
1239 (Fig. 5D) over time and compare their changes. Comparing retrospective estimates
1240 with real-time estimates is also helpful but not practical in real-time.

1241 The effects of censoring bias were more subtle on a daily scale. In our simula-
1242 tions, we found that failing to account for censoring typically led to biased estimates
1243 of the standard deviation but unbiased estimates of the mean. When inappropriate
1244 censoring adjustments were used the estimated mean was also biased. An example
1245 common in the literature of an inappropriate censoring adjustment is to use a single-
1246 day discretisation, which only accounts for censoring of a single event and induces
1247 a bias of half the unaccounted-for censoring interval to the mean of the estimated
1248 distribution. The approximate-latent-variable model, the interval-reduced-censoring-
1249 and-truncation model, the dynamical-bias-correction method, and the joint-modeling
1250 method gave relatively unbiased and precise estimates of underlying delay distribu-
1251 tion parameters for the daily reporting scenario.

1252 4.2 Limitations

1253 While we have shown that accounting for both truncation and censoring biases is
1254 critical to accurately estimating epidemiological delay distributions, some methods
1255 can be more computationally costly than others. In particular, latent variable meth-
1256 ods required nearly an order of magnitude more computational time in most cases
1257 compared to non-latent variable methods. The dynamical bias correction method
1258 and the joint model of primary incidence and the forward distribution had similar
1259 computational requirements to the latent model. However, in most instances, these
1260 requirements were manageable with typical research computing resources (i.e., laptop
1261 scale). In settings with time-varying and partially pooled delay distributions, this
1262 may no longer be the case and so non-latent approaches may be favoured despite the
1263 slight increase in bias.

1264 In terms of accounting for dynamical biases, previous studies (Verity et al., 2020a;
1265 Britton and Scalia Tomba, 2019; Park et al., 2021) focused on the exponential growth
1266 phase, simplifying the problem. This approach is acceptable as long as growth stays
1267 roughly constant. However, propagating uncertainty appropriately is difficult with
1268 this approach and stable growth rates are rare in practice. Here, we present a novel

1269 version of the growth rate correction method that accounts for flexible changes in
1270 incidence patterns. Whilst this method performed well in retrospective settings, its
1271 application to real-time epidemics is currently limited due to its dependence on un-
1272 truncated incidence data. It was routinely outperformed for real-time usage by other
1273 methods that directly accounted for truncation and censoring. These approaches are
1274 also more readily implementable using existing software, can account for censoring
1275 windows of varying sizes, and were more numerical stable.

1276 Our case study of the Ebola virus disease epidemic revealed important gaps in
1277 the methods we present here. First, the reporting delays show considerable variation
1278 throughout the epidemic which is significantly larger than any bias due to censoring
1279 or truncation; current methods are not able to fully account for temporal changes
1280 in the delay distribution. While it may seem relatively straightforward to extend
1281 the model to allow for time-varying parameters across primary cohorts, censoring
1282 of the primary events complicates the problem by adding uncertainty to their co-
1283 hort time. The dynamical correction method performed particularly poorly on the
1284 Ebola virus disease data. This was due to left truncation which was caused by di-
1285 viding the data into four observation periods. Properly accounting for this induced
1286 left-truncation would require integrating the entire backward distribution, which is
1287 currently computationally impractical.

1288 A key limitation of our work is that we only consider an idealized daily censor-
1289 ing process—in principle, our methodology and implementation support alternative
1290 censoring periods, except for the dynamic adjustment model. However, we did ex-
1291 plore short delays and high growth rates which has a similar impact to having longer
1292 censoring intervals with long delays and slower growth rates. We found that exponen-
1293 tial growth and truncation affect the distribution of event times within the censoring
1294 window, particularly when delay distributions are short relative to the length of the
1295 censoring window. These effects caused the empirical distribution of event times
1296 within the censoring interval to deviate from the assumed uniform distribution and
1297 we expect larger biases for wider censoring intervals. More work is needed to develop
1298 robust methods for dealing with wider censoring intervals that account for the under-
1299 lying generative process of primary events. More generally the simulated scenarios
1300 we considered were idealised and did not include observation error for either the pri-
1301 mary or secondary events. Testing our models against idealized scenarios allowed us
1302 to identify the detailed sources of biases, but may have favoured methods that did
1303 not try and account for observation error and other real-world sources of biases.

1304 **4.3 Generalisability**

1305 We only considered lognormal distributions in this paper for brevity and because
1306 it is commonly used in the literature, but our findings generalise across other dis-
1307 tributions. Our implementations are also readily extensible to other distributions.
1308 We also did not consider mixtures of distributions, which can better describe some
1309 epidemiological delay distributions that are generated using multiple transmission

1310 or disease progression states (Vink et al., 2014). In addition, we do not include
1311 non-parametric or hazard-based methods in our assessment, although the joint in-
1312 cidence and forward distribution we did consider has been generalised to support
1313 these methods. However, again our key findings generalise to both these settings
1314 and, since our models are implemented using the `brms` package, it would be rela-
1315 tively easy to include these complexities with minimal additional work. Finally, the
1316 dynamical correction method assumes that the incidence is known exactly—a joint
1317 estimation of the incidence pattern and delay distribution, similar to that used in
1318 the joint incidence and forward distribution model, may improve this method’s real-
1319 time performance. Despite these limitations, our conclusions about the importance
1320 of truncation and censoring biases should be carefully considered for any epidemic
1321 analyses, especially when estimating delay distributions.

1322 In this work, we have primarily focused on inferring distributions of non-transmission
1323 intervals (i.e., excluding generation- and serial-interval distributions). Although it
1324 would be possible to apply our methods to infer the mean and standard deviations
1325 of the transmission intervals, there are additional complications that we did not con-
1326 sider. In particular, transmission intervals may not be independent of each other if
1327 they share the same source case. Other problems include identifying intermediate
1328 hosts and the possibility of multiple potential source cases for an infectee. More work
1329 is needed to validate methods for inferring transmission interval distributions.

1330 Estimation of epidemiological delay distributions is a common task in infectious
1331 disease modelling. In this work, we have given particular focus to daily censoring
1332 and right truncation adjustment as these are the most common scenarios researchers
1333 face when estimating delay distributions. When censoring is adjusted for, it is com-
1334 monly assumed to be only censoring for the primary event and not the secondary
1335 event (i.e in the daily setting only account for a day vs two days of censoring). For
1336 example, researchers often account for the censoring in the infection time when esti-
1337 mating incubation-period distributions, but not in the symptom onset time. Right
1338 truncation is rarely adjusted for, and when it is, methods with limited theoretical
1339 support are commonly used which do not, or only partially, account for this bias.
1340 These methods are rarely validated against simulations (Backer et al., 2020; Linton
1341 et al., 2020) but are nonetheless often reused. For example, there has been an in-
1342 creased usage of methods that account for dynamical and right-truncation biases at
1343 the same time (Guo et al., 2023b,a); however, these two biases each pertain to back-
1344 ward and forward distributions, respectively and therefore should not be combined.
1345 Approaches that combine both biases will overcompensate for missing observations
1346 and overestimate the mean. As early estimates of epidemiological distributions are
1347 rarely re-estimated after the initial phase of an epidemic, due to lack of resources
1348 and the difficulty in collecting data, these biased estimates may remain the canonical
1349 ones throughout an epidemic, and beyond, further biasing decision-making.

1350 More work is needed to improve software support for estimating distributions.
1351 Our code base from this work is now part of the `epinowcast` community, a group
1352 of infectious disease researchers aiming to improve epidemic and surveillance tools,

1353 meaning that it will be further developed into a robust tool. New members and
1354 support towards this aim are warmly welcomed. Further work is also needed to
1355 understand optimal methods for modelling time-varying distributions and mixture
1356 distributions with latent components where both may suffer from right truncation.

1357 **4.4 Conclusions**

1358 This study shows that care is required when estimating epidemiological distributions.
1359 We provide theory, methods, and tools to enable practitioners to circumvent common
1360 pitfalls that we have described and compare these methods in a range of simulated
1361 and real-world scenarios. Future epidemic analyses should carefully consider the
1362 different biases outlined in this study and make sure to use methods that can account
1363 for them and that have been robustly validated.

1364 **Acknowledgements** We thank Michael DeWitt for helpful comments on the manuscript.
1365 SF was supported by Wellcome Trust (210758/Z/18/Z).

1366 Supplementary Figures

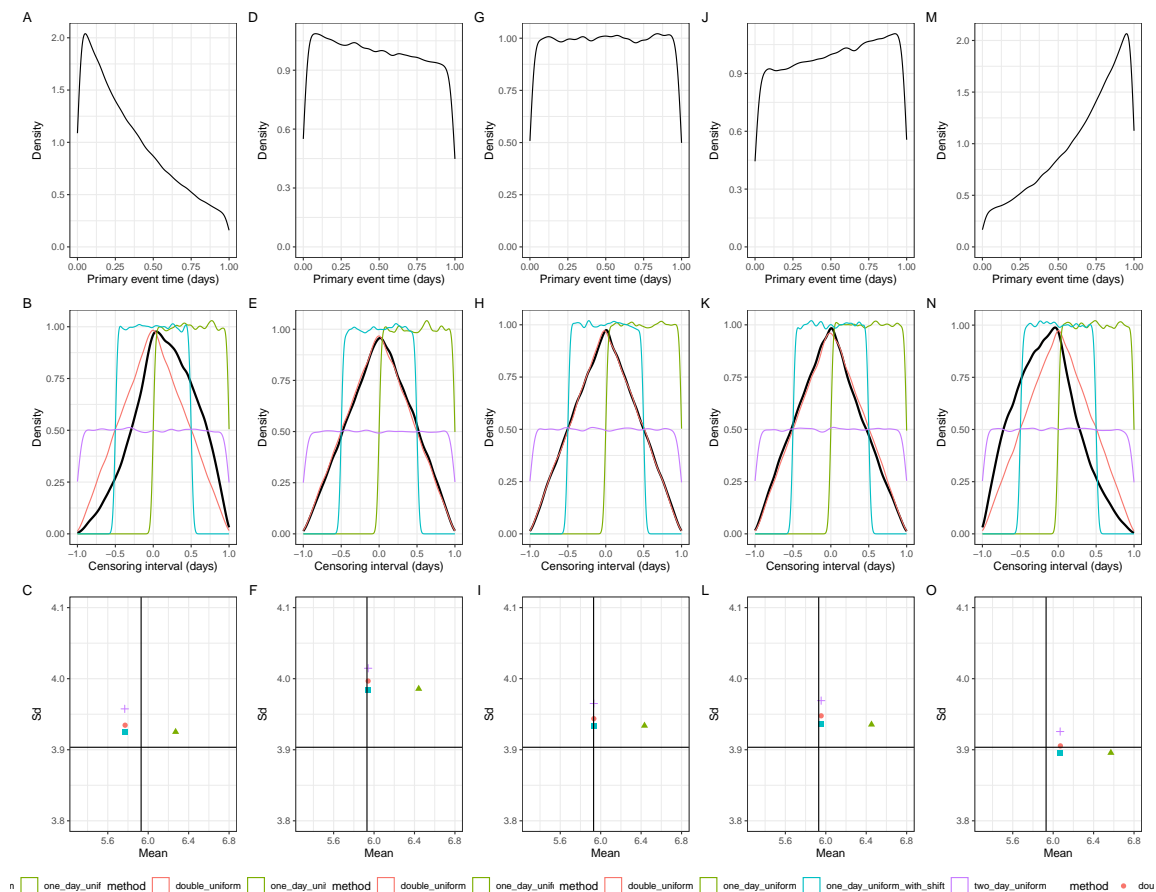


Figure 8: The impact of assumptions about prior distributions on converting discrete-time distributions to continuous-time distributions. (A, D, G, J, M) The distribution of primary event times within a one-day censoring interval across different growth rates. (B, E, H, K, N) The corresponding distribution of weights for the interval-reduced censoring (black lines) against different approximations (colored lines). (C, F, I, L, O) The means and standard deviations of the resulting continuous-time distributions across different assumptions (colored points) against the true mean (vertical lines) and standard deviations (horizontal lines).

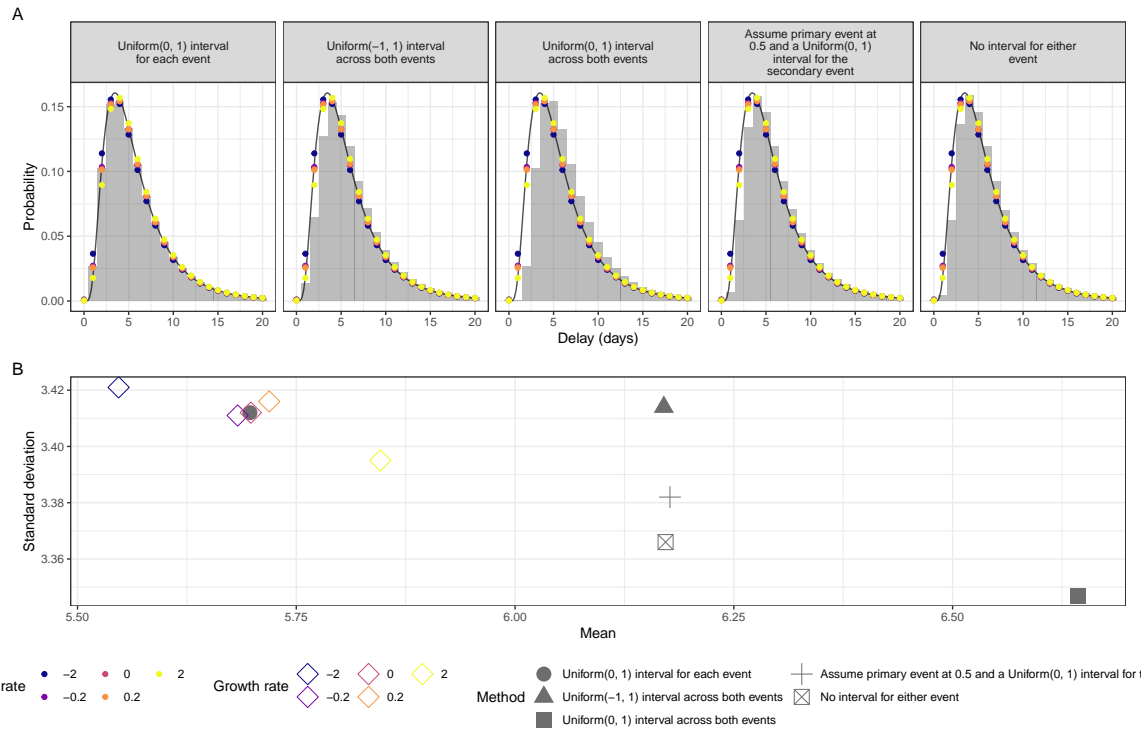


Figure 9: The impact of assumptions about prior distributions on converting continuous-time distributions to discrete-time probability mass functions. (A) Empirically observed probability mass functions (PMFs) from an underlying continuous lognormal distribution delay simulation (mean: 5.9 days, standard deviation: 3.9 days) with daily censoring. Observed PMFs for each interval-reduced censoring window approximation are shown using grey bars, PMFs under different growth rate assumptions for the primary interval are shown using coloured points. The black line indicates the underlying continuous probability density function used for simulation. (B) Empirically observed means and standard deviations from the same simulation as (D). Growth rate-adjusted primary censoring intervals are shown with coloured diamonds, and method-based censoring are indicated using grey shapes.

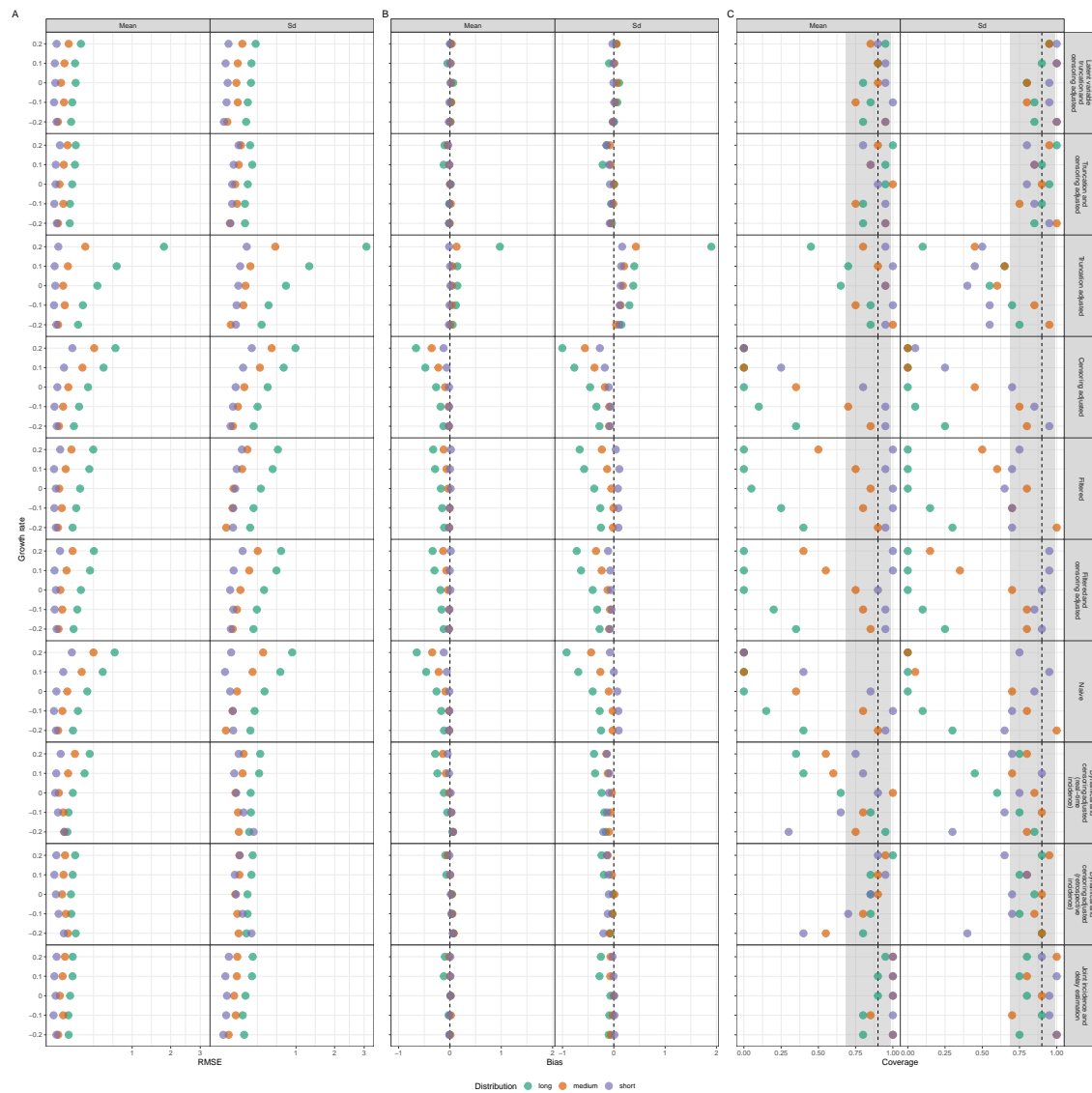


Figure 10: Coloured points represent summary statistics for each distribution scenario (short, medium, and long). (A) RMSE (root mean squared error) of each method. (B) Relative bias of each method. Vertical dashed lines represent the unbiased estimate. (C) Coverage probability of each method. Vertical dashed lines represent the 90% coverage probability. Shaded regions represent the 95% binomial confidence interval around 90% given number of simulations.



Figure 11: (A) Relative CRPS of each method across all sample sizes investigated (10, 100, 200, 400) for the outbreak scenario simulation. Here black diamonds represent the global relative CRPS. Coloured points represent the mean CRPS for each distribution scenario with the shape indicating if the score is for the mean (circle) or standard deviation (Sd, triangle). (B) Relative CRPS of each method stratified by sample sizes (10, 100, 200, 400) for the outbreak scenario simulation.

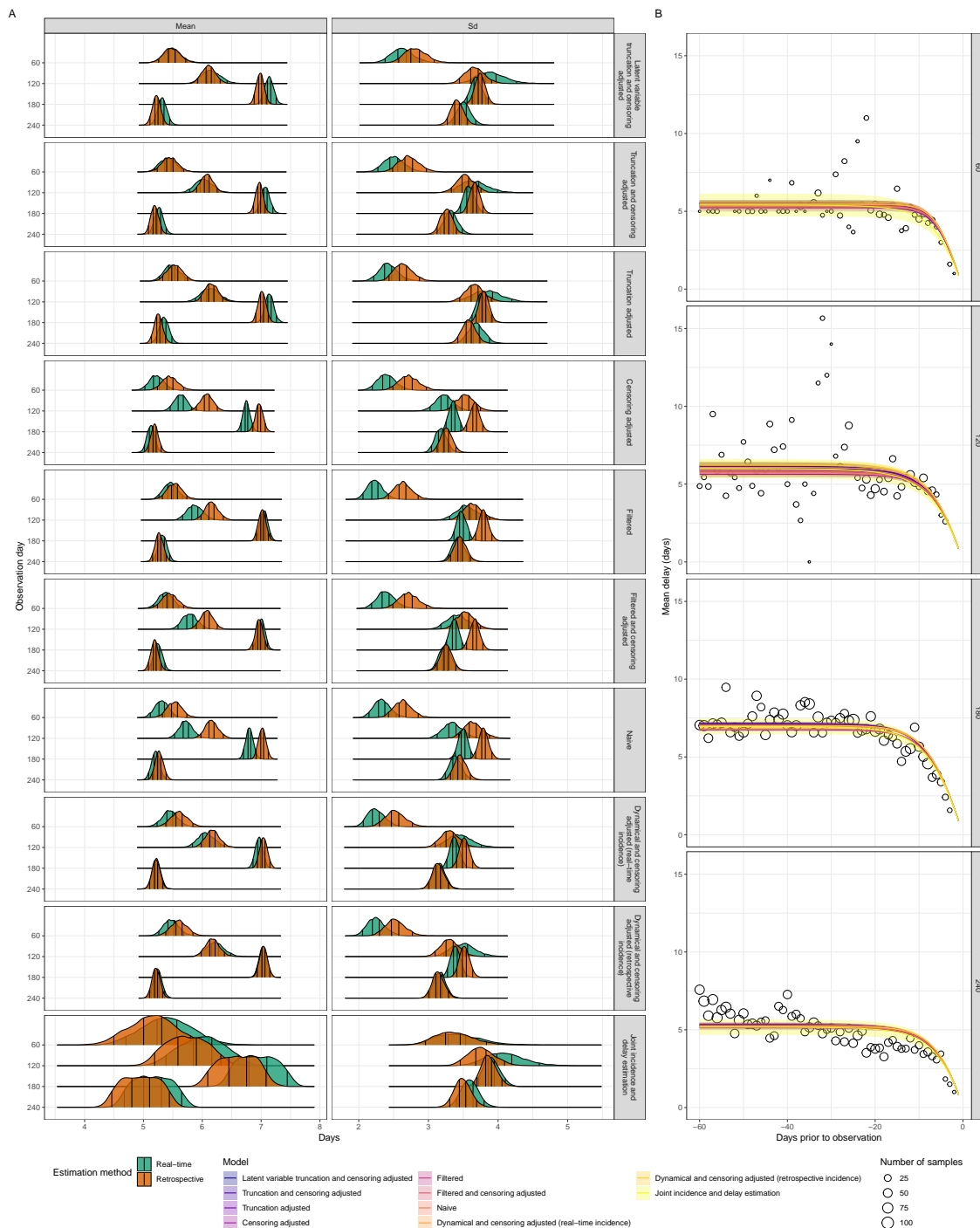


Figure 12: (A) Posterior distributions of mean and standard deviation (Sd). Vertical lines represent the 5%, 35%, 65%, and 95% quantiles respectively. Models are ordered based on the order used in the methods section. (B) Posterior predictions of truncated mean overtime against the observed forward mean. Lines and shaded regions represent posterior median and 95% credible intervals. Points represent the observed mean values. Note that some models are over-plotted here and hence may not be clearly distinguishable.

References

- 1367
- 1368 Abbott, S., Hellewell, J., Thompson, R. N., Sherratt, K., Gibbs, H. P., Bosse, N. I.,
1369 Munday, J. D., Meakin, S., Doughty, E. L., Chun, J. Y., Chan, Y.-W. D., Finger,
1370 F., Campbell, P., Endo, A., Pearson, C. A. B., Gimma, A., Russell, T., Flasche,
1371 S., Kucharski, A. J., Eggo, R. M., Funk, S., and CMMID COVID modelling group
1372 (2020). Estimating the time-varying reproduction number of SARS-CoV-2 using
1373 national and subnational case counts. *Wellcome Open Res.*, 5:112.
- 1374 Abbott, S., Lison, A., and Funk, S. (2021). epinowcast: Flexible hierarchical now-
1375 casting. *Zenodo*.
- 1376 Backer, J. A., Eggink, D., Andeweg, S. P., Veldhuijzen, I. K., van Maarseveen,
1377 N., Vermaas, K., Vlaemynck, B., Schepers, R., van den Hof, S., Reusken, C. B.,
1378 et al. (2022). Shorter serial intervals in SARS-CoV-2 cases with Omicron BA. 1
1379 variant compared with Delta variant, the Netherlands, 13 to 26 December 2021.
1380 *Eurosurveillance*, 27(6):2200042.
- 1381 Backer, J. A., Klinkenberg, D., and Wallinga, J. (2020). Incubation period of 2019
1382 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China,
1383 20–28 January 2020. *Eurosurveillance*, 25(5):2000062.
- 1384 Beesley, L. J., Osthus, D., and Del Valle, S. Y. (2022). Addressing delayed
1385 case reporting in infectious disease forecast modeling. *PLoS Comput. Biol.*,
1386 18(6):e1010115.
- 1387 Betancourt, M. (2017). Diagnosing biased inference with divergences. *Stan Case*
1388 *Studies*, 4.
- 1389 Boettiger, C. (2015). An introduction to Docker for reproducible research. *ACM*
1390 *SIGOPS Operating Systems Review*, 49(1):71–79.
- 1391 Bosse, N. I., Abbott, S., Cori, A., van Leeuwen, E., Bracher, J., and Funk, S.
1392 (2023). Transformation of forecasts for evaluating predictive performance in an
1393 epidemiological context.
- 1394 Bosse, N. I., Gruson, H., Cori, A., van Leeuwen, E., Funk, S., and Abbott, S. (2022).
1395 Evaluating forecasts with scoringutils in r. *arXiv*.
- 1396 Britton, T. and Scalia Tomba, G. (2019). Estimation in emerging epidemics: biases
1397 and remedies. *Journal of the Royal Society Interface*, 16(150):20180670.
- 1398 Brookmeyer, R. and Damiano, A. (1989). Statistical methods for short-term projec-
1399 tions of AIDS incidence. *Statistics in Medicine*, 8(1):23–34.
- 1400 Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package
1401 brms. *The R Journal*, 10(1):395–411.

- 1402 Cain, K. C., Harlow, S. D., Little, R. J., Nan, B., Yosef, M., Taffe, J. R., and
1403 Elliott, M. R. (2011). Bias due to left truncation and left censoring in longitudinal
1404 studies of developmental and disease processes. *American journal of epidemiology*,
1405 173(9):1078–1084.
- 1406 Champredon, D. and Dushoff, J. (2015). Intrinsic and realized generation intervals
1407 in infectious-disease transmission. *Proceedings of the Royal Society B: Biological
1408 Sciences*, 282(1821):20152026.
- 1409 Cori, A., Ferguson, N. M., Fraser, C., and Cauchemez, S. (2013). A new framework
1410 and software to estimate time-varying reproduction numbers during epidemics.
1411 *American journal of epidemiology*, 178(9):1505–1512.
- 1412 Fang, L.-Q., Yang, Y., Jiang, J.-F., Yao, H.-W., Kargbo, D., Li, X.-L., Jiang, B.-G.,
1413 Kargbo, B., Tong, Y.-G., Wang, Y.-W., Liu, K., Kamara, A., Dfafe, F., Kanu, A.,
1414 Jiang, R.-R., Sun, Y., Sun, R.-X., Chen, W.-J., Ma, M.-J., Dean, N. E., Thomas,
1415 H., Longini, Jr, I. M., Halloran, M. E., and Cao, W.-C. (2016). Transmission
1416 dynamics of ebola virus disease and intervention effectiveness in sierra leone. *Proc.
1417 Natl. Acad. Sci. U. S. A.*, 113(16):4488–4493.
- 1418 Flaxman, S., Mishra, S., Gandy, A., Unwin, H. J. T., Mellan, T. A., Coupland, H.,
1419 Whittaker, C., Zhu, H., Berah, T., Eaton, J. W., Monod, M., Imperial College
1420 COVID-19 Response Team, Ghani, A. C., Donnelly, C. A., Riley, S., Vollmer, M.
1421 A. C., Ferguson, N. M., Okell, L. C., and Bhatt, S. (2020). Estimating the effects of
1422 non-pharmaceutical interventions on COVID-19 in europe. *Nature*, 584(7820):257–
1423 261.
- 1424 Fraser, C., Donnelly, C. A., Cauchemez, S., Hanage, W. P., Van Kerkhove, M. D.,
1425 Hollingsworth, T. D., Griffin, J., Baggaley, R. F., Jenkins, H. E., Lyons, E. J.,
1426 et al. (2009). Pandemic potential of a strain of influenza A (H1N1): early findings.
1427 *science*, 324(5934):1557–1561.
- 1428 Fraser, C., Riley, S., Anderson, R. M., and Ferguson, N. M. (2004). Factors that make
1429 an infectious disease outbreak controllable. *Proceedings of the National Academy
1430 of Sciences*, 101(16):6146–6151.
- 1431 Gabry, J. and Češnovar, R. (2021). *cmdstanr: R Interface to 'CmdStan'*. [https://mc-](https://mc-stan.org/cmdstanr)
1432 [stan.org/cmdstanr](https://discourse.mc-stan.org), <https://discourse.mc-stan.org>.
- 1433 Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B.
1434 (2013). *Bayesian data analysis*. CRC press.
- 1435 Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using
1436 multiple sequences. *Statistical science*, 7(4):457–472.

- 1437 Ghani, A. C., Donnelly, C. A., Cox, D. R., Griffin, J. T., Fraser, C., Lam, T. H., Ho,
1438 L. M., Chan, W. S., Anderson, R. M., Hedley, A. J., and Leung, G. M. (2005).
1439 Methods for estimating the case fatality ratio for a novel, emerging infectious
1440 disease. *Am. J. Epidemiol.*, 162(5):479–486.
- 1441 Gneiting, T. and Raftery, A. E. (2007). Strictly Proper Scoring Rules, Prediction,
1442 and Estimation. *Journal of the American Statistical Association*, 102(477):359–
1443 378. DOI: 10.1198/016214506000001437.
- 1444 Gostic, K., Gomez, A. C., Mummah, R. O., Kucharski, A. J., and Lloyd-Smith,
1445 J. O. (2020). Estimated effectiveness of symptom and risk screening to prevent
1446 the spread of COVID-19. *Elife*, 9.
- 1447 Günther, F., Bender, A., Katz, K., Küchenhoff, H., and Höhle, M. (2021). Nowcast-
1448 ing the COVID-19 pandemic in Bavaria. 63(3):490–502.
- 1449 Guo, Z., Zhao, S., Sun, S., He, D., Chong, K. C., and Yeoh, E. K. (2023a). Estimation
1450 of the serial interval of monkeypox during the early outbreak in 2022. *Journal of*
1451 *Medical Virology*, 95(1):e28248.
- 1452 Guo, Z., Zhao, S., Yam, C. H. K., Li, C., Jiang, X., Chow, T. Y., Chong, K. C., and
1453 Yeoh, E. K. (2023b). Estimating the serial intervals of SARS-CoV-2 Omicron BA.
1454 4, BA. 5, and BA. 2.12. 1 variants in Hong Kong. *Influenza and Other Respiratory*
1455 *Viruses*, 17(2):e13105.
- 1456 Hart, W. S., Maini, P. K., and Thompson, R. N. (2021). High infectiousness imme-
1457 diately before COVID-19 symptom onset highlights the importance of continued
1458 contact tracing. *Elife*, 10:e65534.
- 1459 He, X., Lau, E. H., Wu, P., Deng, X., Wang, J., Hao, X., Lau, Y. C., Wong, J. Y.,
1460 Guan, Y., Tan, X., et al. (2020). Temporal dynamics in viral shedding and trans-
1461 missibility of COVID-19. *Nature medicine*, 26(5):672–675.
- 1462 Hellewell, J., Abbott, S., Gimma, A., Bosse, N. I., Jarvis, C. I., Russell, T. W.,
1463 Munday, J. D., Kucharski, A. J., Edmunds, W. J., Sun, F., et al. (2020). Feasibility
1464 of controlling COVID-19 outbreaks by isolation of cases and contacts. *The Lancet*
1465 *Global Health*, 8(4):e488–e496.
- 1466 Höhle, M. and an der Heiden, M. (2014). Bayesian nowcasting during the STEC
1467 O104:h4 outbreak in Germany, 2011. 70(4):993–1002.
- 1468 Kalbfleisch, J. D. and Lawless, J. F. (1989). Inference based on retrospective as-
1469 certainment: an analysis of the data on transfusion-related AIDS. *Journal of the*
1470 *American Statistical Association*, 84(406):360–372.

- 1471 Landau, W. M. (2021). The targets r package: a dynamic make-like function-oriented
1472 pipeline toolkit for reproducibility and high-performance computing. *Journal of*
1473 *Open Source Software*, 6(57):2959.
- 1474 Lauer, S. A., Grantz, K. H., Bi, Q., Jones, F. K., Zheng, Q., Meredith, H. R., Azman,
1475 A. S., Reich, N. G., and Lessler, J. (2020). The incubation period of coronavirus
1476 disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and
1477 application. *Annals of internal medicine*, 172(9):577–582.
- 1478 Lindsey, J. C. and Ryan, L. M. (1998). Methods for interval-censored data. *Statistics*
1479 *in medicine*, 17(2):219–238.
- 1480 Linton, N. M., Kobayashi, T., Yang, Y., Hayashi, K., Akhmetzhanov, A. R., Jung,
1481 S.-m., Yuan, B., Kinoshita, R., and Nishiura, H. (2020). Incubation period and
1482 other epidemiological characteristics of 2019 novel coronavirus infections with right
1483 truncation: a statistical analysis of publicly available case data. *Journal of clinical*
1484 *medicine*, 9(2):538.
- 1485 Lipsitch, M., Donnelly, C. A., Fraser, C., Blake, I. M., Cori, A., Dorigatti, I., Fergu-
1486 son, N. M., Garske, T., Mills, H. L., Riley, S., Van Kerkhove, M. D., and Hernán,
1487 M. A. (2015). Potential biases in estimating absolute and relative Case-Fatality
1488 risks during outbreaks. *PLoS Negl. Trop. Dis.*, 9(7):e0003846.
- 1489 Lipsitch, M., Joshi, K., and Cobey, S. E. (2020). Comment on Pan A, Liu L, Wang
1490 C, et al: Association of Public Health Interventions With the Epidemiology of the
1491 COVID-19 Outbreak in Wuhan, China. *JAMA*.
- 1492 Lison, A., Abbott, S., Huisman, J., and Stadler, T. (2023). Generative Bayesian
1493 modeling to nowcast the effective reproduction number from line list data with
1494 missing symptom onset dates. *arXiv*.
- 1495 Madewell, Z. J., Charniga, K., Masters, N. B., Asher, J., Fahrenwald, L., Still, W.,
1496 Chen, J., Kipperman, N., Bui, D., Shea, M., Saunders, K., Saathoff-Huber, L.,
1497 Johnson, S., Harbi, K., Berns, A. L., Perez, T., Gateley, E., Spicknall, I. H.,
1498 Nakazawa, Y., Gift, T. L., and 2022 Mpox Outbreak Response Team (2023). Se-
1499 rial interval and incubation period estimates of monkeypox virus infection in 12
1500 jurisdictions, united states, May-August 2022. *Emerg. Infect. Dis.*, 29(4):818–821.
- 1501 Marinović, A. B., Swaan, C., van Steenberghe, J., and Kretzschmar, M. (2015).
1502 Quantifying reporting timeliness to improve outbreak control. *Emerging infectious*
1503 *diseases*, 21(2):209.
- 1504 Miura, F., van Ewijk, C. E., Backer, J. A., Xiridou, M., Franz, E., de Coul, E. O.,
1505 Brandwagt, D., van Cleef, B., van Rijckevorsel, G., Swaan, C., et al. (2022).
1506 Estimated incubation period for monkeypox cases confirmed in the Netherlands,
1507 May 2022. *Eurosurveillance*, 27(24):2200448.

- 1508 Nolen, L. D., Osadebe, L., Katomba, J., Likofata, J., Mukadi, D., Monroe, B., Doty,
1509 J., Hughes, C. M., Kabamba, J., Malekani, J., et al. (2016). Extended human-to-
1510 human transmission during a monkeypox outbreak in the Democratic Republic of
1511 the Congo. *Emerging infectious diseases*, 22(6):1014.
- 1512 Overton, C. E., Pellis, L., Stage, H. B., Scarabel, F., Burton, J., Fraser, C., Hall,
1513 I., House, T. A., Jewell, C., Nurtay, A., et al. (2022). EpiBeds: Data informed
1514 modelling of the COVID-19 hospital burden in England. *PLoS Computational*
1515 *Biology*, 18(9):e1010406.
- 1516 Pan, C., Cai, B., and Wang, L. (2020). A bayesian approach for analyzing partly
1517 interval-censored data under the proportional hazards model. *Statistical methods*
1518 *in medical research*, 29(11):3192–3204.
- 1519 Park, S. W., Sun, K., Abbott, S., Sender, R., Bar-On, Y. M., Weitz, J. S., Funk, S.,
1520 Grenfell, B., Backer, J. A., Wallinga, J., et al. (2022). Inferring the differences in
1521 incubation-period and generation-interval distributions of the Delta and Omicron
1522 variants of SARS-CoV-2. *medRxiv*, pages 2022–07.
- 1523 Park, S. W., Sun, K., Champredon, D., Li, M., Bolker, B. M., Earn, D. J., Weitz,
1524 J. S., Grenfell, B. T., and Dushoff, J. (2021). Forward-looking serial intervals cor-
1525 rectly link epidemic growth to reproduction numbers. *Proceedings of the National*
1526 *Academy of Sciences*, 118(2):e2011548118.
- 1527 R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R
1528 Foundation for Statistical Computing, Vienna, Austria.
- 1529 Reich, N. G., Lessler, J., and Azman, A. S. (2010). *coarseDataTools: A collection of*
1530 *functions to help with analysis of coarsely observed data*. R package version 0.6-6.
- 1531 Reich, N. G., Lessler, J., Cummings, D. A., and Brookmeyer, R. (2009). Esti-
1532 mating incubation period distributions with coarse data. *Statistics in medicine*,
1533 28(22):2769–2784.
- 1534 Seaman, S. R., Presanis, A., and Jackson, C. (2022). Estimating a time-to-event
1535 distribution from right-truncated data in an epidemic: a review of methods. *Sta-*
1536 *tistical methods in medical research*, 31(9):1641–1655.
- 1537 Sender, R., Bar-On, Y., Park, S. W., Noor, E., Dushoff, J., and Milo, R. (2022). The
1538 unmitigated profile of COVID-19 infectiousness. *Elife*, 11:e79134.
- 1539 Singanayagam, A., Patel, M., Charlett, A., Bernal, J. L., Saliba, V., Ellis, J., Lad-
1540 hani, S., Zambon, M., and Gopal, R. (2020). Duration of infectiousness and cor-
1541 relation with RT-PCR cycle threshold values in cases of COVID-19, England,
1542 January to May 2020. *Eurosurveillance*, 25(32):2001483.

- 1543 Stan Development Team (2020). *Prior Choice Recommendations*.
1544 <https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations>.
- 1545 Stan Development Team (2021). *Stan Modeling Language Users Guide and Reference*
1546 *Manual, 2.28.1*. <https://mc-stan.org>.
- 1547 Sun, J. (1995). Empirical estimation of a distribution function with truncated and
1548 doubly interval-censored data and its application to AIDS studies. *Biometrics*,
1549 pages 1096–1104.
- 1550 Svensson, A. (2007). A note on generation times in epidemic models. *Math. Biosci.*,
1551 208(1):300–311.
- 1552 Thompson, R. N., Stockwin, J. E., van Gaalen, R. D., Polonsky, J. A., Kamvar,
1553 Z. N., Demarsh, P. A., Dahlqwist, E., Li, S., Miguel, E., Jombart, T., Lessler, J.,
1554 Cauchemez, S., and Cori, A. (2019). Improved inference of time-varying reproduc-
1555 tion numbers during infectious disease outbreaks. *Epidemics*, 29:100356.
- 1556 Tindale, L. C., Stockdale, J. E., Coombe, M., Garlock, E. S., Lau, W. Y. V.,
1557 Saraswat, M., Zhang, L., Chen, D., Wallinga, J., and Colijn, C. (2020). Evidence
1558 for transmission of COVID-19 prior to symptom onset. *Elife*, 9:e57149.
- 1559 Ushey, K. (2021). *renv: Project Environments*. R package version 0.14.0.
- 1560 Verity, R., Okell, L. C., Dorigatti, I., Winskill, P., Whittaker, C., Imai, N., Cuomo-
1561 Dannenburg, G., Thompson, H., Walker, P. G., Fu, H., et al. (2020a). Estimates
1562 of the severity of coronavirus disease 2019: a model-based analysis. *The Lancet*
1563 *infectious diseases*, 20(6):669–677.
- 1564 Verity, R., Okell, L. C., Dorigatti, I., Winskill, P., Whittaker, C., Imai, N., Cuomo-
1565 Dannenburg, G., Thompson, H., Walker, P. G., Fu, H., et al. (2020b). Estimates
1566 of the severity of coronavirus disease 2019: a model-based analysis. *The Lancet*
1567 *infectious diseases*, 20(6):669–677.
- 1568 Verity, R., Okell, L. C., Dorigatti, I., Winskill, P., Whittaker, C., Imai, N., Cuomo-
1569 Dannenburg, G., Thompson, H., Walker, P. G. T., Fu, H., Dighe, A., Griffin, J. T.,
1570 Baguelin, M., Bhatia, S., Boonyasiri, A., Cori, A., Cucunubá, Z., FitzJohn, R.,
1571 Gaythorpe, K., Green, W., Hamlet, A., Hinsley, W., Laydon, D., Nedjati-Gilani,
1572 G., Riley, S., van Elsland, S., Volz, E., Wang, H., Wang, Y., Xi, X., Donnelly,
1573 C. A., Ghani, A. C., and Ferguson, N. M. (2020c). Estimates of the severity of
1574 coronavirus disease 2019: a model-based analysis. *Lancet Infect. Dis.*, 20(6):669–
1575 677.
- 1576 Vink, M. A., Bootsma, M. C. J., and Wallinga, J. (2014). Serial intervals of respi-
1577 ratory infectious diseases: a systematic review and analysis. *American journal of*
1578 *epidemiology*, 180(9):865–875.

1579 Ward, T., Christie, R., Paton, R. S., Cumming, F., and Overton, C. E. (2022).
1580 Transmission dynamics of monkeypox in the United Kingdom: contact tracing
1581 study. *bmj*, 379.

1582 Ward, T. and Johnsen, A. (2021). Understanding an evolving pandemic: An analysis
1583 of the clinical time delay distributions of COVID-19 in the United Kingdom. *PLoS*
1584 *One*, 16(10):e0257978.

1585 Xin, H., Wong, J. Y., Murphy, C., Yeung, A., Taslim Ali, S., Wu, P., and Cowling,
1586 B. J. (2021). The incubation period distribution of coronavirus disease 2019: a
1587 systematic review and meta-analysis. *Clinical Infectious Diseases*, 73(12):2344–
1588 2352.