

1           **Exact calculation of end-of-outbreak probabilities using contact tracing data**

2           NV Bradbury<sup>1,2†</sup>, WS Hart<sup>3†\*</sup>, FA Lovell-Read<sup>3</sup>, JA Polonsky<sup>4</sup>, RN Thompson<sup>1,2,3\*</sup>

3   **Affiliations:**

4   <sup>1</sup>Mathematics Institute, University of Warwick, Coventry, CV4 7AL, UK

5   <sup>2</sup>Zeeman Institute for Systems Biology and Infectious Disease Epidemiology Research

6   (SBIDER), University of Warwick, Coventry, CV4 7AL, UK

7   <sup>3</sup>Mathematical Institute, University of Oxford, Oxford, OX2 6GG, UK

8   <sup>4</sup>Geneva Centre of Humanitarian Studies, University of Geneva, Geneva, 1205, Switzerland

9

10   <sup>†</sup>Contributed equally to this work

11   \*Correspondence to: [william.hart@maths.ox.ac.uk](mailto:william.hart@maths.ox.ac.uk); [robin.thompson@st-hildas.ox.ac.uk](mailto:robin.thompson@st-hildas.ox.ac.uk)

12

## Abstract

13 A key challenge for public health policy makers is determining when an infectious disease  
14 outbreak has finished. Following a period without cases, an estimate of the probability that no  
15 further cases will occur in future (the end-of-outbreak probability) can be used to inform  
16 whether or not to declare an outbreak over. An existing quantitative approach, based on a  
17 branching process transmission model, allows the end-of-outbreak probability to be  
18 approximated from disease incidence time series, the offspring distribution and the serial  
19 interval of the pathogen (the Nishiura method). Here, we show how the end-of-outbreak  
20 probability under the same transmission model can be calculated exactly if data describing  
21 who-infected-whom (the outbreak transmission tree) are available alongside the disease  
22 incidence time series. When such data are available, for example from contact tracing studies,  
23 our novel approach (the traced transmission method) is straightforward to use. We  
24 demonstrate this by applying the traced transmission method to data from previous outbreaks  
25 of Ebola virus disease and Nipah virus infection. For both outbreak datasets considered, we  
26 find that the traced transmission method would have determined that the outbreak was over  
27 more quickly than the Nishiura method. This highlights that consideration of contact tracing  
28 data may allow stringent control interventions to be relaxed quickly at the end of an outbreak,  
29 with only a limited risk of outbreak resurgence.

30

31 Keywords: mathematical modelling; infectious disease epidemiology; outbreaks; end-of-  
32 outbreak declaration; interventions; public health measures; resurgence; local extinction

33

## Introduction

34 Infectious disease outbreaks require coordinated public health responses that limit the  
35 impacts of disease while avoiding unnecessary interventions. After an outbreak is brought  
36 under control, an important consideration is when the outbreak can be declared over safely.  
37 An end-of-outbreak declaration allows public health measures to be relaxed, but such a  
38 declaration must only be made when there is a low risk of a resurgence in cases [1,2]. World  
39 Health Organization (WHO) guidance for diseases such as Ebola virus disease (EVD)  
40 recommends that the acute phase of an outbreak is declared over when no new cases have  
41 been detected over a period of time that is equal to twice the theoretical maximum incubation  
42 period following the recovery or death of the last reported case (42 days for EVD [3]).

43 Simple rules for determining when to declare an outbreak over, based on fixed time periods  
44 without new cases, are straightforward to apply. However, the risk of a resurgence in cases in  
45 fact depends on specific features of the particular outbreak under consideration. Previous  
46 analyses have found that this risk depends on factors including the reproduction number, the  
47 extent of case underreporting, and the time between symptom onset and removal of the last  
48 detected case [4,5]. This indicates that there is a need for quantitative approaches that can be  
49 applied to guide decisions about when to declare an outbreak over, accounting for features of  
50 the outbreak under consideration.

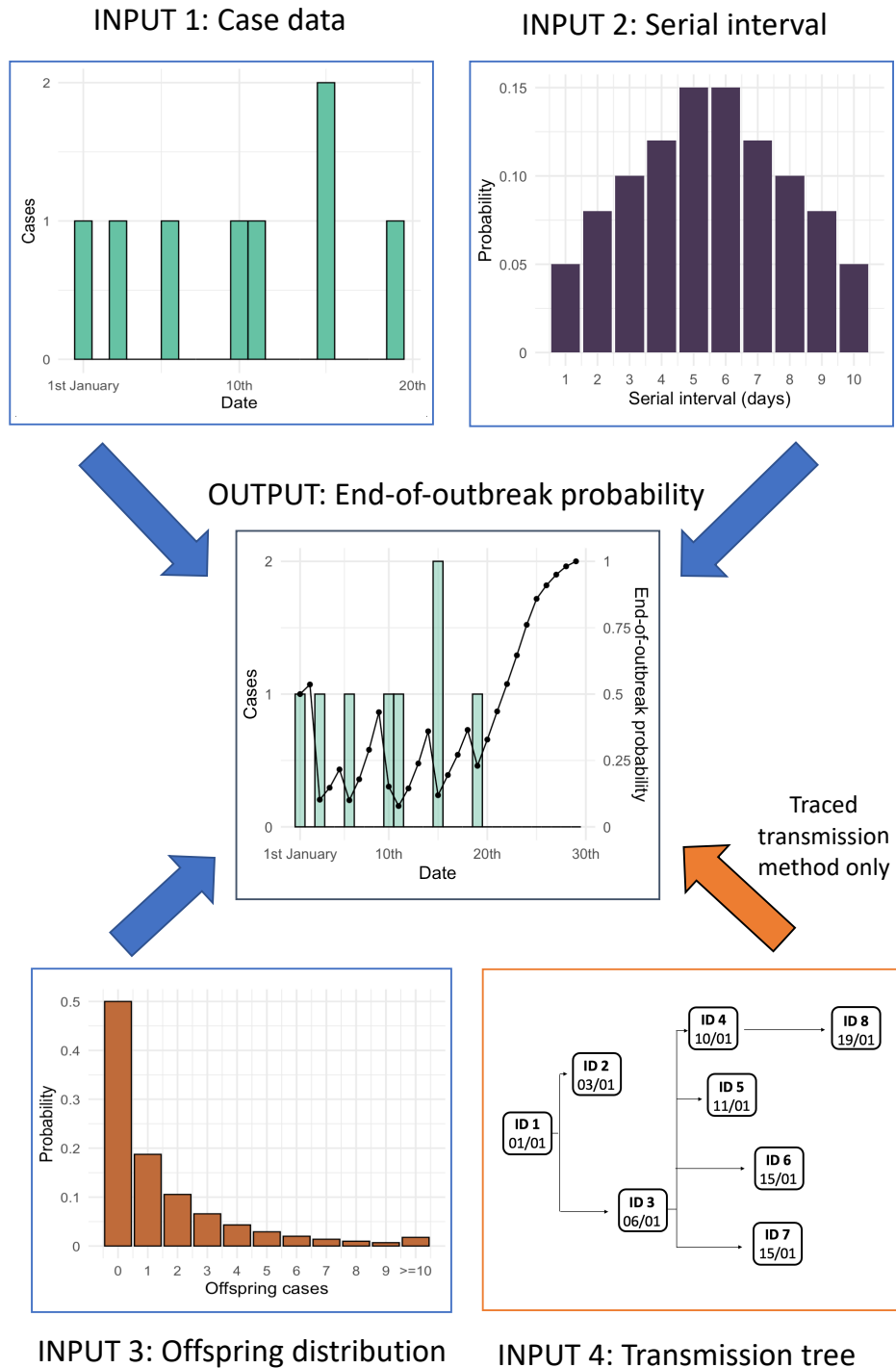
51 There has been recent interest in using mathematical modelling to estimate the probability  
52 that no further cases of disease will occur in future (the *end-of-outbreak probability*), based  
53 on the observed outbreak data up to the current date [1,2]. If the end-of-outbreak probability  
54 can be estimated in real-time during an outbreak, then this quantity facilitates evidence-based  
55 removal of public health interventions. For example, an outbreak could be declared over as  
56 soon as the estimated end-of-outbreak probability exceeds a pre-specified threshold that is set

57 based on the policy maker's level of risk tolerance [1]. Several methods exist for estimating  
58 the end-of-outbreak probability [1,4–14]. The most commonly used approach [6–9], and  
59 therefore the basis from which we began our research here, was introduced by Nishiura *et al.*  
60 [6] (the *Nishiura method*) and is based on a branching process transmission model. The  
61 Nishiura method has the advantage of enabling the end-of-outbreak probability to be  
62 approximated straightforwardly using three inputs (Figure 1): (1) disease incidence time  
63 series (the number of cases recorded on each day of the outbreak up to the current time); (2)  
64 the serial interval distribution (the probability distribution describing the number of days  
65 between the symptom onset dates of an infector-infectee transmission pair); and (3) the  
66 offspring distribution (the probability distribution characterising the number of secondary  
67 cases generated by an infected host).

68 However, even if these inputs are known exactly, and transmission does indeed occur  
69 according to a branching process, the Nishiura method only provides an approximation of the  
70 end-of-outbreak probability (see Methods). Here, we provide a new approach (the *traced*  
71 *transmission method*) for calculating the end-of-outbreak probability exactly under the  
72 branching process transmission model used by Nishiura *et al.* [6] (Figure 1), which can be  
73 applied in scenarios in which information is available about who-infected-whom.

74 Specifically, our approach uses the outbreak transmission tree, which can be obtained or  
75 estimated via contact tracing [7,15], in combination with the inputs to the Nishiura method.  
76 We consider two case studies of outbreaks of viral, zoonotic diseases: an EVD outbreak in  
77 Likati Health Zone, Democratic Republic of the Congo (DRC) in 2017 [16] and an outbreak  
78 of Nipah virus infection in Bangladesh in 2004 [17]. For each outbreak, we compare  
79 estimates of the end-of-outbreak probability obtained using the Nishiura method to analogous  
80 estimates using the traced transmission method. We demonstrate that our exact approach,  
81 with contact tracing data incorporated, leads to different estimates than the Nishiura method

82 while remaining straightforward to apply. To encourage uptake of the novel traced  
83 transmission method to inform when outbreaks of a range of directly transmitted pathogens  
84 can be declared over, we also implement it in an online software application, available via  
85 [https://github.com/nabury/End\\_of\\_outbreak\\_app](https://github.com/nabury/End_of_outbreak_app).



86

87 **Figure 1. Schematic showing the inputs required for the Nishiura method and the traced transmission**

88 **method for estimating the end-of-outbreak probability, and the type of output produced.** Both methods

89 require disease incidence time series (input 1), the serial interval distribution (input 2) and the offspring

90 distribution (input 3). The traced transmission method also requires the outbreak transmission tree (input 4).

91 Blue arrows therefore indicate the inputs required for both methods, and the orange arrow indicates the

92 additional input required for the traced transmission method. The output of both methods is an estimate of the

93 end-of-outbreak probability on a particular day (i.e., the probability that no further cases occur in future, based  
94 on the disease incidence time series data observed up to and including that day).

## 95 Methods

### 96 **Notation**

97 Here, we define the notation used throughout this section and the remainder of this article:

- 98 •  $p(y)$  is the probability mass function of the offspring distribution (in other words, the  
99 probability that a randomly chosen infected individual infects  $y$  other people). In each  
100 outbreak case study, we assumed a negative binomial offspring distribution with mean  
101  $R$  (the reproduction number) and dispersion parameter  $k$ , so that

$$105 \quad p(y) = \frac{\Gamma(k + y)}{y! \Gamma(k)} p_0^y (1 - p_0)^k, \quad \text{for } y = 0, 1, 2, \dots,$$

102 where  $p_0 = R/(R + k)$  and  $\Gamma(z)$  is the Gamma function. The choice of a negative  
103 binomial offspring distribution enables the effect of superspreading to be accounted  
104 for [18]. Specific parameter values used for the two case studies are given below.

- 106 •  $w(x)$  is the probability mass function of the (discrete) serial interval distribution (i.e.,  
107 the probability that the interval between the symptom onset dates of an infector-  
108 infectee transmission pair is  $x$  days, where we assumed that only non-negative serial  
109 intervals can occur), and  $F(x)$  is the corresponding cumulative distribution function.  
110 For each outbreak case study, a discrete serial interval distribution was obtained by  
111 using the method described by Cori *et al.* [19] to discretise a published estimate of the  
112 continuous serial interval distribution (the distribution of time periods between the  
113 precise symptom onset times of infector-infectee transmission pairs). For further  
114 details, see web appendix 11 of that article.

- 115 •  $t$  is the current time, at which we want to estimate the end-of-outbreak probability  
116 (the probability that no cases occur after day  $t$ ).
- 117 • The cases recorded up to and including the current time,  $t$ , are labelled with integer  
118 IDs  $i = 1, 2, \dots, m$  (ordered by symptom onset date). The corresponding symptom  
119 onset dates are denoted by  $t_1, t_2, \dots, t_m$ .
- 120 •  $a_i$  is the number of recorded cases who were infected by individual  $i$  up to (and  
121 including) the current time,  $t$ .

## 122 **The end-of-outbreak probability**

123 Below, we describe the Nishiura method and the traced transmission method for estimating  
124 the end-of-outbreak probability. Both of these methods are based on a branching process  
125 transmission model in which each infected host generates a number of cases that is sampled  
126 from the offspring distribution. These secondary cases then arise in the disease incidence time  
127 series after time periods (following the infector) that are sampled independently from the  
128 serial interval distribution. However, whereas the existing Nishiura method only  
129 approximates the end-of-outbreak probability under this transmission model (as explained  
130 below), the novel traced transmission method enables the end-of-outbreak probability to be  
131 calculated exactly whenever the outbreak transmission tree is known.

### 132 *Nishiura method*

133 Using the notation described in the previous subsection, in the Nishiura method [6] each case  
134 to date,  $i$ , is considered in turn. The offspring and serial interval distributions are used to  
135 calculate the probability,  $q_i$ , that every secondary case generated by individual  $i$  develops  
136 symptoms no later than the current time,  $t$ , assuming nothing is known about the number (or  
137 symptom onset dates) of secondary cases already generated by individual  $i$  (up to time  $t$ ).



138 Specifically,

$$142 \quad q_i = \sum_{y=0}^{\infty} p(y)F(t - t_i)^y .$$

139 The end-of-outbreak probability can then be approximated by

$$143 \quad \text{Prob}(\text{outbreak over on day } t) \approx \prod_{i=1}^m q_i = \prod_{i=1}^m \sum_{y=0}^{\infty} p(y)F(t - t_i)^y .$$

140 If the offspring distribution is a negative binomial distribution, then this formula simplifies to

$$144 \quad \text{Prob}(\text{outbreak over on day } t) \approx \prod_{i=1}^m \left( \frac{1 - p_0}{1 - p_0 F(t - t_i)} \right)^k ,$$

141 as shown in the Supplementary Material.

145 However, this is only an approximation of the end-of-outbreak probability under the assumed  
146 branching process transmission model because the remaining case data are not accounted for  
147 when calculating  $q_i$  (the probability that previous case  $i$  generates no future secondary cases).  
148 Even if the transmission tree is not known, the symptom onset dates of other recorded cases  
149 still provide some information about how many secondary cases may have been generated by  
150 each case  $i$ ; this information is not used in the Nishiura method. In other words, the risk that a  
151 past case generates future infections is assumed to be independent of the number of infections  
152 that the individual has already generated, but this assumption may not hold in the branching  
153 process transmission model on which the Nishiura method is based. Specifically, under a  
154 negative binomial offspring distribution, the probability of an individual generating future  
155 cases increases with the number of cases generated to date – this reflects, for example, that a  
156 more infectious individual is likely to have generated more secondary cases to date (and is  
157 also more likely to generate future cases) than a less infectious individual who developed  
158 symptoms on the same date [18]. Therefore, the Nishiura method only approximates the end-  
159 of-outbreak probability, irrespective of whether or not the transmission tree is known.

160 *Traced transmission method*

161 If the transmission tree up to the current time,  $t$ , is known (so that, in particular, the number  
162 of secondary cases,  $a_i$ , generated to date by each existing case,  $i$ , is known), then assuming  
163 that transmissions occur according to a branching process as described above, the end-of-  
164 outbreak probability can be calculated exactly and is given by

167 
$$\text{Prob}(\text{outbreak over on day } t) = \prod_{i=1}^m \frac{p(a_i)}{\sum_{l=0}^{\infty} \binom{a_i+l}{l} (1 - F(t - t_i))^l p(a_i + l)}.$$

165 This expression is derived in the Supplementary Material. Again, in the case of a negative  
166 binomial offspring distribution, this can be simplified, here giving

168 
$$\text{Prob}(\text{outbreak over on day } t) = \prod_{i=1}^m (1 - p_0(1 - F(t - t_i)))^{(k+a_i)}.$$

169

170 **Outbreak case studies**

171 *Case study 1: Ebola virus disease, Likati, Democratic Republic of the Congo*

172 The first case study we considered is an EVD outbreak that occurred in the Likati Health  
173 Zone of DRC in 2017 [16]. Eight EVD cases were reported between 27 March and 11 May  
174 2017, and symptom onset dates were recorded; five cases were confirmed and the remaining  
175 three were probable. Four of the infected individuals died [16]. The transmission tree was  
176 constructed using contact tracing and the symptom onset date of each case [16] (Figure 2A).  
177 The Ebola offspring distribution was modelled as a negative binomial distribution with  
178 reproduction number  $R = 2.1$  and dispersion parameter  $k = 0.18$  [20] (Figure 2B). We  
179 assumed a gamma-distributed continuous serial interval distribution with mean 15.3 days and

180 standard deviation 9.3 days [20], and discretised this distribution using the method of [19] as  
181 described above (Figure 2C).

182 End-of-outbreak probabilities were estimated each day (based on the data up to and including  
183 that day) from the symptom onset date of the first reported case until 100 days after the  
184 symptom onset date of the last reported case (total duration 146 days), using both the  
185 Nishiura method and the traced transmission method.

### 186 *Case study 2: Nipah virus, Bangladesh*

187 The second case study we considered is an outbreak of Nipah virus infection in the Faridpur  
188 district of Bangladesh between 19 February and 16 April 2004 [17]. Using laboratory testing  
189 and contact tracing, 36 cases were identified of which 23 were laboratory confirmed and 27  
190 died [17]. The number of daily cases (by symptom onset date) peaked at nine on 1 April 2004  
191 [17]. The probable transmission tree is shown in Figure 3A [17]. Two individuals (IDs  $i = 10$   
192 and  $i = 30$ ) were not traced to any other cases [17], and were therefore considered as  
193 imported cases in our analyses (i.e., we assumed that they were not infected by any other case  
194 in the dataset). The individual with ID  $i = 6$  was a local religious leader and had contact with  
195 22 of the cases in this outbreak [17].

196 The offspring distribution for this outbreak was modelled using a negative binomial  
197 distribution with reproduction number  $R = 0.48$  and dispersion parameter  $k = 0.06$  [21,22]  
198 (Figure 3B). We assumed a gamma distributed continuous serial interval distribution with  
199 mean 12.7 days and standard deviation 3.0 days [23], again discretised using the method of  
200 [19] (Figure 3C). End-of-outbreak probabilities were calculated daily from the date of the  
201 first case until 20 days after the last case (total duration 78 days).

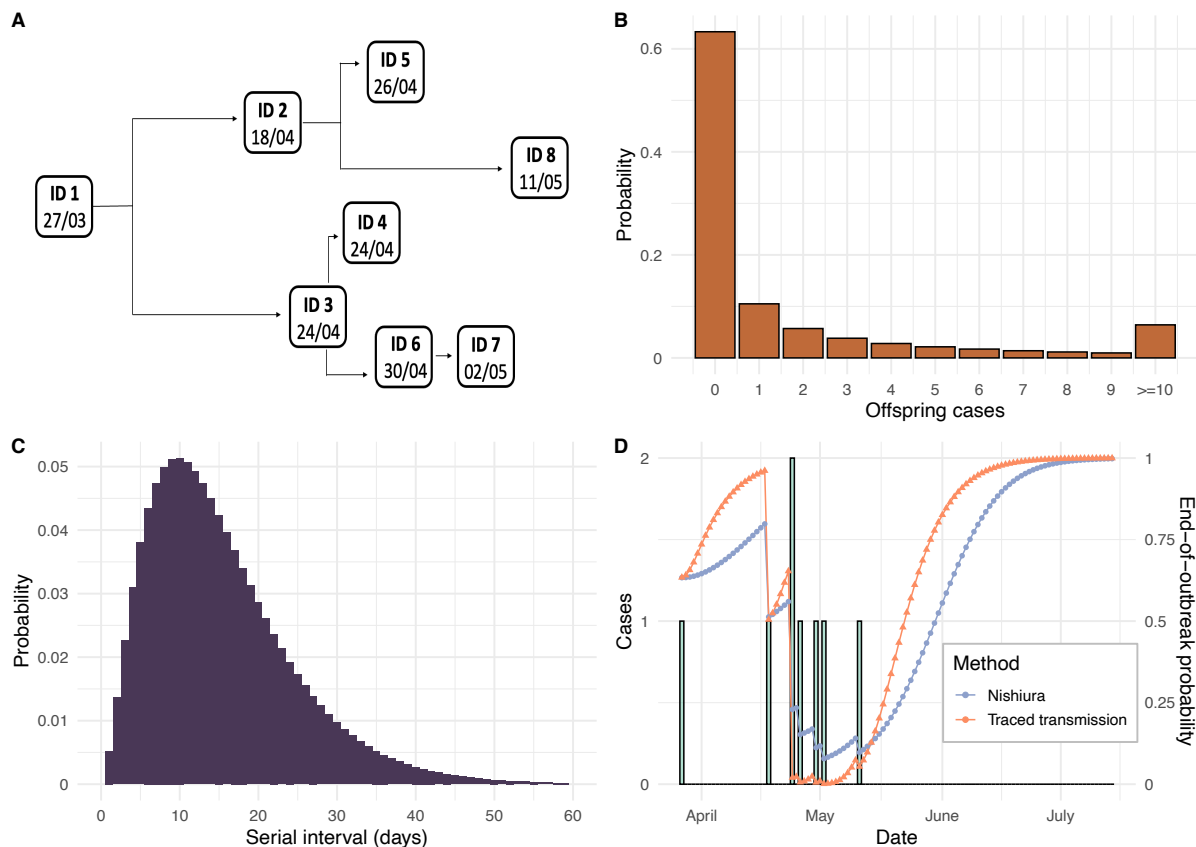
202 In both case studies, we assumed that cases were reported on their symptom onset dates,  
203 thereby neglecting reporting delays when we estimated the end-of-outbreak probabilities. The  
204 symptom onset dates of reported cases are referred to as “case dates” in the remainder of this  
205 article.

## 206 **Results**

### 207 **Real-time estimation of the end-of-outbreak probability**

208 We first used both the Nishiura method and the novel traced transmission method to obtain  
209 end-of-outbreak probability estimates for the 2017 Likati EVD outbreak (Figure 2D). As  
210 would be expected, for both methods, the estimated end-of-outbreak probability increased  
211 over successive days without cases and decreased when new cases occurred.

212 While the general temporal trends were similar under both approaches, the extent of temporal  
213 variations in end-of-outbreak probability estimates was generally more pronounced for the  
214 traced transmission method. For example, on 17 April 2017 there had only been one reported  
215 case of EVD, with symptom onset date 21 days previously. The end-of-outbreak probability  
216 was estimated to be 0.8 using the Nishiura method, compared to a higher value of 0.96 using  
217 the traced transmission approach. On 2 May 2017, following six further cases, the estimated  
218 end-of-outbreak probability reached its lowest value for both approaches: 0.08 for the  
219 Nishiura method, and a lower value of 0.001 for the traced transmission method. Similarly,  
220 the end-of-outbreak probability increased more rapidly following the final case date for the  
221 traced transmission method than for the Nishiura method, with the probability first exceeding  
222 0.99 on 22 June 2017 and 3 July 2017 using the two methods, respectively.



223

224 **Figure 2.** (A) Transmission tree for the 2017 EVD outbreak in the Likati Health Zone of DRC [16]. (B)

225 Offspring distribution assumed for Ebola (negative binomial with reproduction number,  $R = 2.1$  and dispersion

226 parameter,  $k = 0.18$  [20]). (C) Serial interval distribution assumed for Ebola. The continuous serial interval was

227 assumed to be gamma-distributed with mean 15.3 days and standard deviation 9.3 days [20]. This distribution

228 was then discretised using the method from [19]. (D) Estimated daily end-of-outbreak probabilities. Reported

229 cases are represented by the green bars, with the left y-axis showing the daily number of cases. The line plots

230 represent the estimated probability that the outbreak is over for each day of the outbreak for both the Nishiura

231 method (blue) and the traced transmission method (orange). These probabilities are displayed on the right y-

232 axis.

233 We then applied the two methods for estimating the end-of-outbreak probability to the data

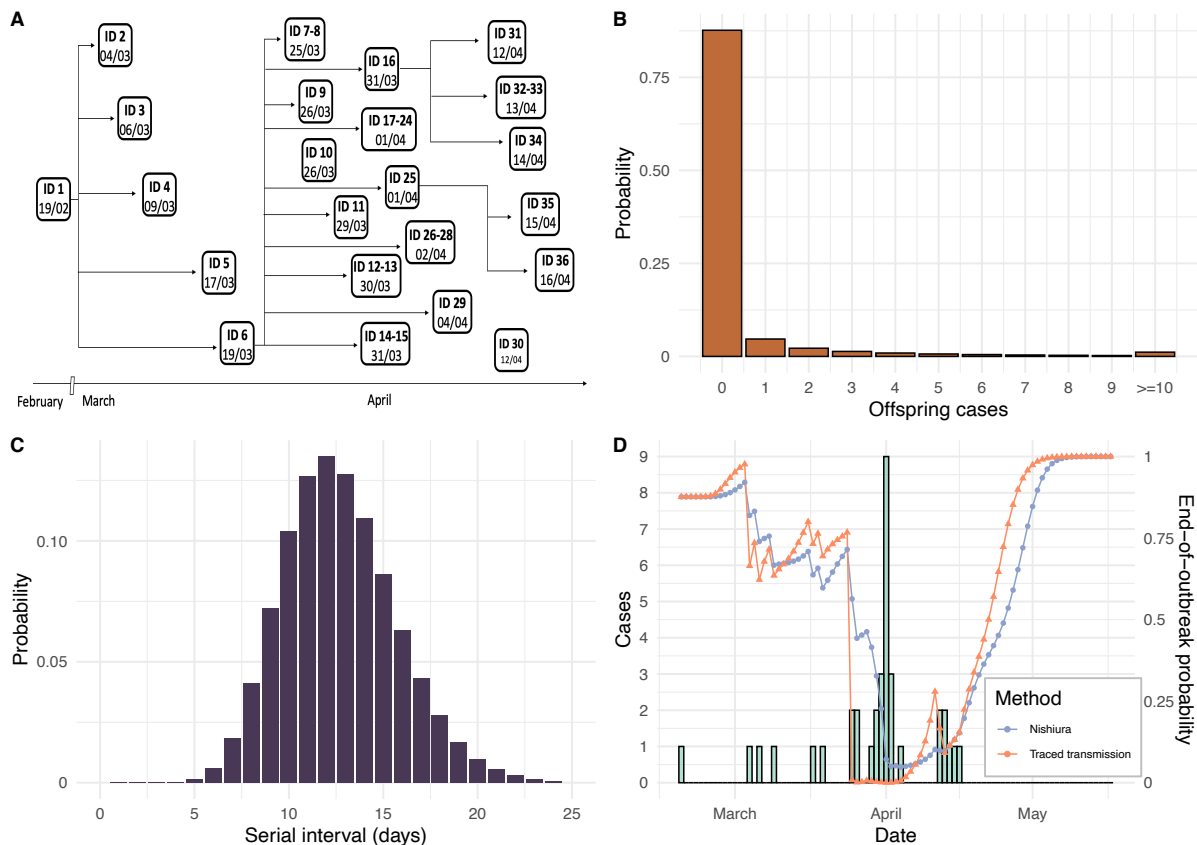
234 from the 2004 outbreak of Nipah virus infection in Bangladesh (Figure 3D). Again, following

235 a cluster of new cases, the estimated end-of-outbreak probability generally fell lower for the

236 traced transmission method than for the Nishiura method – for example, on 1 April 2004

237 (when there were nine new symptomatic cases, the highest daily number during the

238 outbreak), the end-of-outbreak probability was 0.0001 for the traced transmission method and  
 239 0.07 for the Nishiura method. The end-of-outbreak probability also increased more rapidly  
 240 following the final recorded case for the traced transmission method than for the Nishiura  
 241 method.



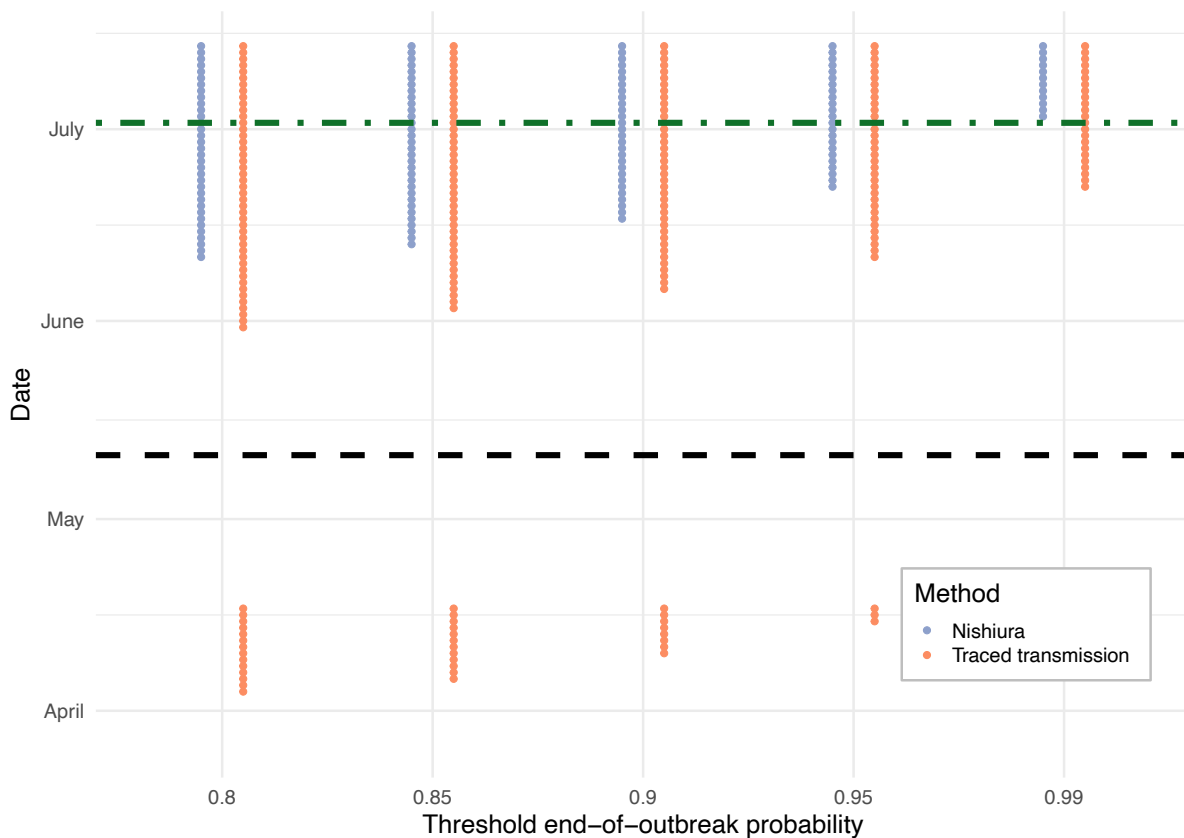
242  
 243 **Figure 3.** (A) Transmission tree for the 2004 outbreak of Nipah virus infection in Bangladesh. (B) Offspring  
 244 distribution assumed for Nipah virus infection (negative binomial with reproduction number,  $R = 0.48$  and  
 245 dispersion parameter,  $k = 0.06$  [21,22]). (C) Serial interval distribution assumed for Nipah virus infection. The  
 246 continuous serial interval was assumed to follow a gamma distribution with mean 12.7 days and standard  
 247 deviation 3.0 days [23]. This distribution was then discretised using the method from [19]. (D) Estimated daily  
 248 end-of-outbreak probabilities. Reported cases are represented by the green bars, with the left y-axis showing the  
 249 daily number of cases. The line plots represent the estimated probability that the outbreak is over for each day of  
 250 the outbreak for both the Nishiura method (blue) and the traced transmission method (orange). These  
 251 probabilities are displayed on the right y-axis.

252

## 253 **End-of-outbreak declaration thresholds**

254 In principle, an outbreak could be considered over whenever the estimated end-of-outbreak  
255 probability exceeds a pre-determined threshold. This threshold can be set according to the  
256 policy maker's willingness to accept a risk of future cases occurring (a lower threshold  
257 corresponds to a faster end-of-outbreak declaration, but with a higher risk that future cases  
258 occur). In Figure 4, we present plots showing the dates on which the 2017 Likati EVD  
259 outbreak would have been considered over for a range of end-of-outbreak probability  
260 threshold values. Results are shown for both the traced transmission method and the Nishiura  
261 method. Equivalent results for the 2004 outbreak of Nipah virus infection in Bangladesh are  
262 shown in Figure S1.

263 For each threshold considered, the two outbreaks would have been declared over earlier  
264 following the final case date using the traced transmission method than using the Nishiura  
265 method. Both methods suggested the EVD outbreak could potentially have been declared  
266 over earlier than the actual end-of-outbreak declaration date of 2 July 2017 [24] (42 days  
267 after the final case recovered, indicated as a green horizontal dash-dotted line in Figure 4) –  
268 the end-of-outbreak probability on 2 July 2017 was 0.998 for the traced transmission method  
269 and 0.989 for the Nishiura method.



270

271 **Figure 4. End-of-outbreak probability thresholds for the 2017 Likati EVD outbreak.** The x-axis represents  
272 a range of end-of-outbreak probability thresholds, and the y-axis shows the outbreak dates on which these  
273 thresholds were exceeded by the estimated end-of-outbreak probability, for both the Nishiura method (blue) and  
274 the traced transmission method (orange). The date of the final recorded case (11 May 2017) is indicated as a  
275 black dashed line, and the actual end-of-outbreak declaration date (2 July 2017) [24] as a green dash-dotted line.

276 We note that, if the traced transmission method was used with an end-of-outbreak probability  
277 threshold of 0.96 or below, then the outbreak would have been considered over early in the  
278 outbreak, in April 2017, after only one case had occurred. A similar phenomenon can be seen  
279 when both the Nishiura method and the traced transmission method are applied to data from  
280 the Nipah virus infection outbreak (Figure S1). The occurrence of further cases when the  
281 transmission models suggested this to be unlikely is discussed below (see Discussion).

282

283



## 284 **Sensitivity of findings to the offspring distribution**

285 Estimates of the reproduction number,  $R$  (representing overall transmissibility), and  
286 dispersion parameter,  $k$  (where lower values of  $k$  correspond to more overdispersed  
287 transmission; in other words, a greater degree of superspreading [18]), vary between  
288 outbreaks, even of the same infectious disease [18,22]. For example, the offspring  
289 distribution may differ due to different strains of a virus or behavioural characteristics of the  
290 affected population. We therefore investigated the effect of the assumed values of  $R$  and  $k$  on  
291 estimates of the end-of-outbreak probability using the Nishiura method and the traced  
292 transmission method, for the EVD and Nipah case studies (Figures S2 and S3). For each  
293 outbreak, we considered values of  $R$  and  $k$  both lower and higher than our assumed baseline  
294 values.

295 For both outbreaks, the assumed value of  $k$  (i.e., the extent of superspreading) had a  
296 particularly pronounced effect on the difference between the end-of-outbreak probability  
297 estimates under the Nishiura and traced transmission methods. In general, this difference was  
298 greater for lower values of the dispersion parameter,  $k$  (Figures S2 and S3). Higher values of  
299  $R$  appeared to lead to larger differences between the methods than lower values of  $R$ .

## 300 **Discussion**

301 Quantitative approaches for estimating the probability that an infectious disease outbreak has  
302 ended help policy advisors determine when the outbreak should be declared over [1,2].

303 Accurate estimation of the end-of-outbreak probability enables resource-intensive  
304 surveillance and control measures to be relaxed or removed as quickly as possible while  
305 limiting the risk of additional cases occurring. Here, we have developed a new approach (the  
306 traced transmission method) for estimating the end-of-outbreak probability in scenarios in

307 which contact tracing enables reconstruction of the outbreak transmission tree (up to the  
308 current date). Our method uses the same branching process transmission model as an existing  
309 approximate approach for estimating the end-of-outbreak probability [6] (the Nishiura  
310 method), but unlike the Nishiura method, the traced transmission method gives the exact end-  
311 of-outbreak probability under this transmission model.

312 While we used the Nishiura method as the basis for our research given its previous use to  
313 calculate end-of-outbreak probabilities during outbreaks of a range of diseases (including  
314 MERS, EVD and COVID-19 [6–9]), we note that other methods for estimating the end-of-  
315 outbreak probability exist [4,5,10–14,25]. These methods have accounted for factors such as  
316 unreported cases [4,5,11] and temporal variations in the reproduction number [4,11]. While  
317 these other methods can be complex, sometimes requiring large numbers of simulations of  
318 stochastic epidemiological models to be run [4,5,25], a benefit of the Nishiura method is its  
319 straightforward application. The traced transmission method is similarly easy-to-use,  
320 allowing the end-of-outbreak probability to be estimated using a simple formula. We have  
321 developed an interactive, web-based app to facilitate future use of our approach (available via  
322 [https://github.com/nabury/End\\_of\\_outbreak\\_app](https://github.com/nabury/End_of_outbreak_app)).

323 To demonstrate our method, we considered outbreaks of the Ebola and Nipah viruses as case  
324 studies. Both these viruses are zoonotic pathogens that cause sporadic outbreaks in humans,  
325 with high case fatality rates of 25-90% [26] and 40-75%, respectively [27]. Outbreaks are  
326 typically met with stringent control measures aiming to break chains of human-to-human  
327 transmission as rapidly as possible [26–28]. The question of when an outbreak can be  
328 declared over so that costly interventions can be safely relaxed or removed is therefore  
329 particularly pertinent to these viruses. Furthermore, as was the case for the two specific

330 outbreaks we considered, intense contact tracing provides an opportunity to reconstruct  
331 outbreak transmission trees.

332 We found that estimates of the end-of-outbreak probability can vary substantially between  
333 our novel traced transmission method and the existing Nishiura method (Figures 2D and 3D).  
334 Specifically, the traced transmission method exhibited larger temporal variations in estimates  
335 of the end-of-outbreak probability during each outbreak, with the probability typically  
336 reaching lower values following clusters of new cases but then increasing more rapidly over  
337 successive days without any cases. The traced transmission approach therefore indicated that  
338 the two case study outbreaks could have been declared over earlier than suggested by the  
339 Nishiura method. The difference between the methods increased when assuming a smaller  
340 dispersion parameter,  $k$ , corresponding to a greater degree of superspreading (Figures S2 and  
341 S3).

342 While the difference in end-of-outbreak probability estimates between the two methods may  
343 be partially attributable to the fact that the traced transmission method leverages more data  
344 than the Nishiura method, this is unlikely to explain the consistent trends described in the  
345 previous paragraph fully. We note that, for the traced transmission method with a negative  
346 binomial offspring distribution, the probability of a recorded case generating no future cases  
347 is smaller if that individual has generated more secondary cases to date. This reflects the fact  
348 that an individual who has already generated more secondary cases may be more infectious  
349 and/or may have more contacts with susceptible individuals, and therefore may be more  
350 likely to generate future cases (either through future transmissions or past transmissions to  
351 individuals yet to develop symptoms), compared an individual who has generated fewer  
352 secondary cases to date. This effect is enhanced for a more overdispersed offspring  
353 distribution, when there is more variation between infected individuals in the number of

354 secondary cases generated. On the other hand, the Nishiura method neglects information  
355 provided by the case data about how many secondary infections each case to date may have  
356 already generated. Even when the transmission tree is not known, some information is  
357 available about possible numbers of secondary cases generated by each case to date through  
358 the disease incidence time series, and this information is not used in the Nishiura method.

359 Both methods for estimating the end-of-outbreak probability considered here suggest a high  
360 probability that the Likati EVD outbreak had ended by the actual date on which the outbreak  
361 was declared over (based on current WHO guidance that recommends waiting for 42 days  
362 following the recovery or safe burial of the last recorded case). In comparison, one previous  
363 study recommended that the current 42 day waiting time guideline needed to be extended to  
364 ensure a high probability of an EVD outbreak being over at the time of an end-of-outbreak  
365 declaration [4], while another study found the appropriate waiting time to depend on the level  
366 of surveillance [5]. This second finding is consistent with our result here that a shorter  
367 waiting time before declaring the Likati EVD outbreak over may have been sufficient, since  
368 intensive contact tracing was undertaken.

369 For both case studies considered, our traced transmission method gives a high end-of-  
370 outbreak probability estimate immediately before the second case occurred (0.96 for the  
371 Likati EVD outbreak and 0.98 for the Bangladesh Nipah outbreak), and a similar  
372 phenomenon can be seen using the Nishiura method for the Nipah outbreak. While the  
373 subsequent occurrences of further cases may have indeed been realisations of unlikely events,  
374 other explanations for this finding are possible. First, the assumed offspring and serial  
375 interval distributions may not have been correct for the specific outbreaks considered – for  
376 example, a higher value of the dispersion parameter would lead to lower end-of-outbreak  
377 probability estimates on the corresponding days (see Figures S2 and S3). Alternatively, long

378 gaps between the first and second recorded cases may have resulted from unrecorded  
379 intermediate cases. This is particularly likely early in an outbreak when intensive surveillance  
380 may not yet be in place. Extension of the traced transmission method to account for  
381 unreported cases and reporting delays, considering the sensitivity of the surveillance system  
382 and the time to put enhanced surveillance in place, is a target for future exploration,  
383 particularly as pathogens such as the Ebola virus tend to emerge in locations with weak  
384 surveillance. In addition, other possible areas for future work include accounting for  
385 uncertainty and/or temporal changes in the offspring and serial interval distributions.

386 In general, careful consideration should be given to the choice of probability threshold for an  
387 end-of-outbreak declaration, to appropriately balance the risk of an incorrect declaration with  
388 the economic and social costs of maintaining stringent outbreak controls for longer than  
389 necessary. One possibility is to take a stepped approach in which an initial end-of-outbreak  
390 declaration is made but surveillance measures are not removed completely until the estimated  
391 end-of-outbreak probability reaches a second, higher, threshold. Current policy for EVD  
392 requires heightened surveillance to be maintained for at least six months following an initial  
393 end-of-outbreak declaration [3].

394 In summary, we have developed a new approach for calculating the end-of-outbreak  
395 probability that robustly accounts for recorded transmission data. Application of our method  
396 indicates that two past outbreaks could have been declared over earlier than suggested by an  
397 existing approximate method. We hope that our approach is useful for informing end-of-  
398 outbreak declarations in future infectious disease outbreaks. The results from this modelling  
399 framework should be used as one of a range of sources of evidence to support public health  
400 decision making.

401

## 402 **COMPETING INTERESTS**

403 We have no competing interests.

## 404 **AUTHORS' CONTRIBUTIONS**

405 NVM – formal analysis, investigation, software, validation, writing – review and editing.

406 WSH – formal analysis, investigation, software, validation, writing – original draft, writing –  
407 review and editing.

408 FAL – formal analysis, writing – review and editing.

409 JAP – methodology, writing – review and editing.

410 RNT – conceptualization, methodology, formal analysis, project administration, supervision,  
411 writing – original draft, writing – review and editing.

## 412 **FUNDING**

413 This work was funded by the UKRI via grant number EP/V053507/1 (NVM and RNT) and  
414 via a Doctoral Prize (WSH, grant number EP/W524311/1).

## 415 **ACKNOWLEDGEMENTS**

416 Thanks to members of the Zeeman Institute for Systems Biology and Infectious Disease  
417 Epidemiology Research at the University of Warwick for useful discussions about this  
418 research.

## 419 **DATA AVAILABILITY**

420 An interactive, web-based Shiny app was developed to conduct end-of-outbreak probability  
421 calculations using the traced transmission approach. This application and underlying R code  
422 are available at [https://github.com/nabury/End\\_of\\_outbreak\\_app](https://github.com/nabury/End_of_outbreak_app). All coding and analysis was  
423 conducted in the R programming language (version 4.0.2).

424

## 425 References

- 426 1. Linton NM, Lovell-Read FA, Southall E, Lee H, Akhmetzhanov AR, Thompson RN,  
427 Nishiura H. 2022 When do epidemics end? Scientific insights from mathematical  
428 modelling studies. *Centaurus* **64**, 31–60. (doi:10.1484/j.cnt.5.130125)
- 429 2. Nishiura H. 2016 Methods to determine the end of an infectious disease epidemic:  
430 a short review. In: *Mathematical and Statistical Modeling for Emerging and Re-*  
431 *emerging Infectious Diseases* (Eds. Chowell G and Hyman JM), 291–301.  
432 (doi:10.1007/978-3-319-40413-4\_17)
- 433 3. World Health Organization. 2020 WHO recommended criteria for declaring the  
434 end of the Ebola Virus Disease Outbreak. See  
435 [https://www.who.int/publications/m/item/who-recommended-criteria-for-](https://www.who.int/publications/m/item/who-recommended-criteria-for-declaring-the-end-of-the-ebola-virus-disease-outbreak)  
436 [declaring-the-end-of-the-ebola-virus-disease-outbreak](https://www.who.int/publications/m/item/who-recommended-criteria-for-declaring-the-end-of-the-ebola-virus-disease-outbreak) (accessed on 17  
437 November 2022).
- 438 4. Djaafara BA, Imai N, Hamblion E, Impouma B, Donnelly CA, Cori A. 2021 A  
439 quantitative framework for defining the end of an infectious disease outbreak:  
440 application to Ebola Virus Disease. *Am J Epidemiol* **190**, 642–651.  
441 (doi:10.1093/aje/kwaa212)
- 442 5. Thompson RN, Morgan OW, Jalava K. 2019 Rigorous surveillance is necessary for  
443 high confidence in end-of-outbreak declarations for Ebola and other infectious  
444 diseases. *Phil Trans Roy Soc B* **374**, 20180431. (doi:10.1098/rstb.2018.0431)
- 445 6. Nishiura H, Miyamatsu Y, Mizumoto K. 2016 Objective determination of end of  
446 MERS outbreak, South Korea, 2015. *Emerg Infect Dis* **22**, 146–148.  
447 (doi:10.3201/eid2201.151383)
- 448 7. Akhmetzhanov AR, Jung S, Cheng H-Y, Thompson RN. 2021 A hospital-related  
449 outbreak of SARS-CoV-2 associated with variant Epsilon (B.1.429) in Taiwan:  
450 transmission potential and outbreak containment under intensified contact  
451 tracing, January–February 2021. *Int J Infect Dis* **110**, 15–20.  
452 (doi:10.1016/j.ijid.2021.06.028)
- 453 8. Lee H, Nishiura H. 2019 Sexual transmission and the probability of an end of the  
454 Ebola virus disease epidemic. *J Theor Biol* **471**, 1–12.  
455 (doi:10.1016/j.jtbi.2019.03.022)
- 456 9. Linton NM, Akhmetzhanov AR, Nishiura H. 2021 Localized end-of-outbreak  
457 determination for coronavirus disease 2019 (COVID-19): examples from clusters  
458 in Japan. *Int J Infect Dis* **105**, 286–292. (doi:10.1016/j.ijid.2021.02.106)
- 459 10. Griette Q, Liu Z, Magal P, Thompson RN. 2022 Real-time prediction of the end of an  
460 epidemic wave: COVID-19 in China as a case-study. In *Mathematics of Public*  
461 *Health: Proceedings of the Seminar on the Mathematical Modelling of COVID-19*  
462 (eds VK Murty, J Wu), pp. 173–195. Cham: Springer International Publishing.  
463 (doi:10.1007/978-3-030-85053-1\_8)
- 464 11. Parag KV, Donnelly CA, Jha R, Thompson RN. 2020 An exact method for  
465 quantifying the reliability of end-of-epidemic declarations in real time. *PLoS*  
466 *Comput Biol* **16**, e1008478. (doi:10.1371/journal.pcbi.1008478)
- 467 12. Parag KV. 2021 Sub-spreading events limit the reliable elimination of  
468 heterogeneous epidemics. *J R Soc Interface* **18**, 20210444.  
469 (doi:10.1098/rsif.2021.0444)

- 470 13. Parag KV, Cowling BJ, Donnelly CA. 2021 Deciphering early-warning signals of  
471 SARS-CoV-2 elimination and resurgence from limited data at multiple scales. *J R*  
472 *Soc Interface* **18**, 20210569. (doi:10.1098/rsif.2021.0569)
- 473 14. Yuan B, Liu R, Tang S. 2022 A quantitative method to project the probability of the  
474 end of an epidemic: application to the COVID-19 outbreak in Wuhan, 2020. *J Theor*  
475 *Biol* **545**, 111149. (doi:10.1016/j.jtbi.2022.111149)
- 476 15. World Health Organization, Centers for Disease Control and Prevention. 2015  
477 Implementation and management of contact tracing for Ebola virus disease. See  
478 [https://apps.who.int/iris/bitstream/10665/185258/1/WHO\\_EVD\\_Guidance\\_Con](https://apps.who.int/iris/bitstream/10665/185258/1/WHO_EVD_Guidance_Con)  
479 [tact\\_15.1\\_eng.pdf](https://apps.who.int/iris/bitstream/10665/185258/1/WHO_EVD_Guidance_Con) (accessed on 17 November 2022).
- 480 16. Nsio J *et al.* 2020 2017 outbreak of Ebola Virus Disease in northern Democratic  
481 Republic of Congo. *J Infect Dis* **221**, 701–706. (doi:10.1093/infdis/jiz107)
- 482 17. Gurley ES *et al.* 2007 Person-to-person transmission of Nipah virus in a  
483 Bangladeshi community. *Emerg Infect Dis* **13**, 1031–1037.  
484 (doi:10.3201/eid1307.061128)
- 485 18. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. 2005 Superspreading and the  
486 effect of individual variation on disease emergence. *Nature* **438**, 355–359.  
487 (doi:10.1038/nature04153)
- 488 19. Cori A, Ferguson NM, Fraser C, Cauchemez S. 2013 A new framework and software  
489 to estimate time-varying reproduction numbers during epidemics. *Am J Epidemiol*  
490 **178**, 1505–1512. (doi:10.1093/aje/kwt133)
- 491 20. van Kerkhove MD, Bento AI, Mills HL, Ferguson NM, Donnelly CA. 2015 A review  
492 of epidemiological parameters from Ebola outbreaks to inform early public health  
493 decision-making. *Sci Data* **2**, 150019. (doi:10.1038/sdata.2015.19)
- 494 21. Devnath P, Masud HMAA. 2021 Nipah virus: a potential pandemic agent in the  
495 context of the current severe acute respiratory syndrome coronavirus 2  
496 pandemic. *New Microbes New Infect* **41**, 100873.  
497 (doi:10.1016/j.nmni.2021.100873)
- 498 22. Taube JC, Miller PB, Drake JM. 2022 An open-access database of infectious disease  
499 transmission trees to explore superspreader epidemiology. *PLoS Biol* **20**,  
500 e3001685. (doi:10.1371/journal.pbio.3001685)
- 501 23. Nikolay B *et al.* 2019 Transmission of Nipah Virus — 14 years of investigations in  
502 Bangladesh. *New Eng J Med* **380**, 1804–1814. (doi:10.1056/NEJMoa1805376)
- 503 24. World Health Organization. 2017 Declaration of the end of Ebola virus disease  
504 outbreak in the Democratic Republic of the Congo.
- 505 25. Eichner M, Dietz K. 1996 Eradication of poliomyelitis: when can one be sure that  
506 polio virus transmission has been terminated? *Am J Epidemiol* **143**, 816–822.
- 507 26. World Health Organization. 2021 Ebola virus disease. See  
508 <https://www.who.int/news-room/fact-sheets/detail/ebola-virus-disease>  
509 (accessed on 24 November 2022).
- 510 27. World Health Organization. 2018 Nipah virus. See [https://www.who.int/news-](https://www.who.int/news-room/fact-sheets/detail/nipah-virus)  
511 [room/fact-sheets/detail/nipah-virus](https://www.who.int/news-room/fact-sheets/detail/nipah-virus) (accessed on 24 November 2022).
- 512 28. Keita M *et al.* 2022 Investigation of and strategies to control the final cluster of the  
513 2018-2020 Ebola virus disease outbreak in the eastern Democratic Republic of  
514 Congo. *Open Forum Inf Dis* **9**, 9. (doi:10.1093/ofid/ofac329)
- 515