Article Summary Line: Unilist capture-recapture methods can be used to assess the completenss of contact tracing during outbreaks of directly-transmitted disease, including Ebola Virus Disease.

Running Title: Completeness of Ebola contact tracing, DRC

Keywords: Ebolavirus, Contact Tracing, Disease Outbreaks, Democratic Republic of the Congo

Novel application of capture-recapture methods to estimate the completeness of contact tracing during a large outbreak of Ebola Virus Disease, Democratic Republic of Congo, 2018-2020

Authors:

Jonathan A. Polonsky[1,2], Dankmar Böhning[3], Mory Keita[2,4], Steve Ahuka-Mundeke[5], Justus Nsio-Mbeta[6], Aaron Aruna Abedi[6], Mathias Mossoko[5], Janne Estill[2], Olivia Keiser[2], Laurent Kaiser[2], Zabulon Yoti[4], Patarawan Sangnawakij[7], Rattana Lerdsuwansri[7], Victor J. Del Rio Vilas[8*]

Affiliations:

[1] World Health Organization, Geneva, Switzerland (J. Polonsky)

[2] Institute of Global Health, Faculty of Medicine, University of Geneva, Geneva, Switzerland (J. Polonsky, M. Keita, J. Estill, O. Keiser, L. Kaiser)

[3] Southampton Statistical Sciences Research Institute, University of Southampton, Southampton, United Kingdom (D. Böhning)

[4] World Health Organization, Regional Office for Africa, Brazzaville, Congo (M. Keita, Z. Yoti)

[5] Institut National de Recherche Biomédicale, Democratic Republic of Congo (S. Ahuka-Mundeke, M. Mossoko)

[6] Ministere de la Santé Publique, Direction Générale de la Lutte contre la Maladie, Kinshasa, Democratic Republic of Congo (J. Nsio-Mbeta, A. Aruna Abedi)

[7] Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat University, Thailand (P. Sangnawakij, R. Lerdsuwansri)

[8] World Health Organization, South East Asia Regional Office, New Delhi, India (V. Del Rio Vilas)

* corresponding author

Abstract

Despite its critical role in containing outbreaks, the efficacy of contact tracing (CT), measured as the sensitivity of case detection, remains an elusive metric. We estimated the sensitivity of CT by applying unilist capture-recapture methods on data from the 2018-2020 outbreak of Ebola virus disease in the Democratic Republic of Congo. We applied different distributional assumptions to the zero-truncated count data to estimate the number of

unobserved cases with a) any contacts and b) infected contacts, to compute CT sensitivity. Geometric distributions were the best fitting models. Our results indicate that CT efforts identified almost all (n=792, 99%) of the cases with any contacts, but only half (n=207, 48%) of the cases with infected contacts, suggesting that CT efforts performed well at identifying contacts during the listing stage, but performed poorly during the contact follow-up stage. We discuss extensions to our work and potential applications for the current COVID-19 pandemic.

Introduction

Contact tracing (CT) is the process by which individuals who are believed to have come into contact with a confirmed case of an infectious disease during their infectious period are located and checked for the presence of the infection or disease. Under traditional approaches, CT involves three distinct steps; *contact identification*, in which potential contacts are identified through interview with the primary case; *contact listing*, in which those identified contacts are listed and communication established with them; and *contact follow-up*, in which those listed contacts are monitored for presence of infection or development of disease over a predefined period [1].

Due to its important role in case detection to monitor and curtail chains of transmission, CT often forms part of the public health response to directly-transmitted infectious diseases [2]. Recently, CT has received widespread attention due to its critical role in the response to outbreaks of diphtheria [3], Ebola virus disease (EVD) [4–6], and the ongoing COVID-19 pandemic [7,8].

From 2018-2020, the Democratic Republic of the Congo (DRC) experienced its twelfth and largest EVD outbreak, the second largest ever experienced globally [9]. EVD is a disease caused by viruses of the genus *Ebolavirus*, family *Filoviridae*. Zoonotic spillover events from the animal reservoir have led to large, explosive outbreaks in West and Central Africa in recent years [9–12]. Owing to its high pathogenicity and virulence, an elimination control strategy is always adopted, aiming to ensure that all cases are identified, isolated, and treated promptly after disease onset, thereby limiting the opportunity for onward community spread. Although CT is a central pillar of control [13], there are no standardised methods to assess a critical aspect of performance, its sensitivity, i.e. the ability to detect all contacts and secondary infections resulting from cases.

One approach to quantifying this metric is to employ capture-recapture (CRC) methods [14,15]. Broadly, this family of methodological approaches permits quantifying the number of individuals missing from lists, and subsequently estimate the sensitivity of the surveillance effort and the probability of case detection. While CRC has previously been used to estimate the number of unobserved cases of disease [16,17], such approaches typically rely on comparison of multiple lists, which are generally not available for contact lists. Therefore, we describe the application of a unilist capture-recapture approach [15] to quantifying the number of unobserved cases and contacts, and describe their sociodemographic profile, helping to identify plausible risk factors that can be used to target limited resources at those unobserved cases most likely to generate onward transmission.

More precisely, we aim to address the following questions, from which we can derive CT sensitivity estimates:

1. How many cases with *any* contacts did CT miss?

2. How many cases with *infected* contacts did CT miss?

Materials and methods

Materials

We included all confirmed and probable EVD cases and contacts (classified according to standardised case definitions [18,19]) identified in Beni Health Zone, DRC between 31 July 2018 and 26 April 2020. Cases were principally detected through three identification mechanisms: i) passive detection at healthcare facilities from clinically suspect individuals presenting symptoms consistent with EVD, ii) house-to-house active case finding by community health workers, and iii) tracing the contacts of EVD cases. CT was coordinated by the Ministry of Public Health with support from WHO, and conducted by locally-recruited teams of contact tracers. Upon detection of a case, efforts to identify and list their contacts were undertaken.

For cases, our data contains basic information on sociodemographic characteristics (age, sex, Health Area of residence), and dates of disease onset and isolation. For contacts, our data contains similar socio-demographic information, and information on the daily follow-up and final status of the contact (either "completed the 21 days follow-up", "confirmed as EVD case", "lost to follow-up", "never seen", and "died during follow-up"). Contacts recorded as "confirmed as EVD case" were those identified by the CT teams during the course of their work. EVD was assumed to be the cause of death for contacts recorded as "died during follow-up" due to the short interval between their contact with an EVD case and their death.

Methods

Exploratory data analysis

We present the distribution of cases according to age, sex, and timing of disease onset. The distribution of the number of contacts per case are described between two distinct epidemic waves, with the Wilcoxon test used to explore differences in continuous variables and Chi-squared test used for categorical variables. “Superspreading”, or overdispersion in the offspring distribution of secondary cases arising from infectious individuals, may have profound impacts on control strategies in low-resource settings [20,21], and we describe the extent of this phenomenon in two ways: firstly by assessing the proportion of infectious individuals linked to 80% of onward transmission using methods described by Endo *et al* [22]; and secondly by estimating the dispersion parameter (k) using methods described by Althaus [23].

A multivariable logistic regression model was used to explore risk factors associated with loss to follow-up, in which previously successfully traced contacts (those identified, listed, and among whom follow-up has begun) become untraceable at some point during the 21 days follow-up period. In such cases, contacts unable to be traced for three consecutive days are recorded as having been lost to follow-up, with no further attempts at tracing made.

To explore characteristics of cases with infected contacts, we calculated the mean number of contacts, mean age, and sex ratio of cases with at least one *listed* contact (among whom we can be confident that at least a minimal investigation was conducted), according to three

categories – those with no infected contacts identified, those with precisely one, and those with two or more.

Capture-recapture modelling

We classified the observed cases according to their number of *listed* contacts (either precisely zero or those with at least one), further classifying this latter category according to the number of *infected* contacts observed (either precisely zero or those with at least one). For each detected case, the CT process generates a list of individuals fitting the definition for a contact (Supplemetary Materials), some of whom may themselves have been infected and will eventually become secondary cases. From this, frequency distributions of cases with any listed contacts, and of cases with infected contacts, can be generated by first excluding (truncating) those cases with zero contacts. For example, the data can be binned into the number of cases with exactly one contact (f_1), two contacts (f_2), and so on to the number of cases with the maximum number of contacts (f_m). Statistically, this leads to a zero-truncated observed count distribution of cases with at least one contact. By applying a unilist CRC approach designed to estimate unobserved population sizes using the distribution of count data within single lists [15], we can infer f_0, the number of unobserved cases with at least one contact. Associated with the observed frequencies ($f_1, f_2, \ldots, f_m$) and unobserved f_0 there are probabilities $p_1, p_2, \ldots, p_m$ and p_0 informing the probability of identifying a case with exactly 1, 2,…, m and 0 contacts, respectively. A conventional approach is to assume that the frequencies arise from a discrete distribution such as the Poisson where $p_0 = e^{-\lambda}$, $p_1 = \frac{e^{-\lambda}\lambda}{1!}$, $p_2 = \frac{e^{-\lambda}\lambda^2}{2!}$, .., $p_m = \frac{e^{-\lambda}\lambda^m}{m!}$. Other common distributions are the negative binomial and the geometric distribution. The geometric distribution has probabilities $p_0 = p, p_1 = p(1-p)$,

168 $p(1-p)^2 \ldots\ p_m = p(1-p)^m$ where p is a probability parameter. Poisson and geometric
169 are special cases of the negative binomial distribution which provides a flexible model
170 family. More details are given in the Supplementary Materials. As the observed distribution
171 contains only positive numbers of contacts we need to consider the associated zero-truncated
172 distribution $p_1/\,(1-p_0),\ p_2/\,(1-p_0)\ ,\ldots\ p_m/\,(1-p_0)$. In other words, we (1) assume
173 that the number of *observed* contacts among cases who actually had contacts follows a
174 parametric distribution (although non-parametric approaches are possible [15,24,25]), (2) find
175 the best-fitting zero-truncated distribution based on the cases with at least one observed
176 contact (we explore the zero-truncated Poisson, negative binomial, and geometric
177 distributions (see Supplementary Materials)), and (3) use the estimated probability p_0 of not
178 observing a case with contacts (calculated from the best-fitting distribution) to inform
179 standard population estimators. Here we use the Horvitz-Thompson estimator to estimate $f0$,
180 the unobserved number of cases

181 $$\widehat{f0} \;=\; n\,\frac{p_0}{1-p_0}$$

182 where n is the number of observed cases with at least one observed contact and p_0 is defined
183 as above. The Horvitz-Thompson estimator provides an unbiased estimate of $f0$, provided
184 that p_0 is correctly specified, hence it is important to use a correctly-specified distribution for
185 the number of observed contacts. We use maximum likelihood for model fitting, selecting the
186 model with the smallest Akaike and Bayesian Information Criteria (AIC and BIC,
187 respectively). Details are given in the Supplementary Materials.
188
189 To estimate 95% confidence intervals (95%CIs), we use a parametric bootstrap, described as
190 follows. Suppose that $\widehat{N}$ is the estimated size of the (observed and unobserved) population
191 under a fitted model. We generate *B* samples of size $\widehat{N}$ using the fitted model and its
192 estimated parameter(s). For each sample, all zeros are truncated and the size estimate $\widehat{N}_b$

computed, for each of the samples *b=1, …, B.* We chose *B=10000* to minimise bootstrap simulation random error. We constructed 95%CIs using the 2.5th percentile of the distribution of $\widehat{N}_b$ as the lower end and the 97.5th percentile as the upper end.

Results

Exploratory data analysis

We identified 913 confirmed and 10 probable EVD cases in Beni HZ. The CT process listed 80,556 contacts, of whom 6224 were duplicates, having been listed as the contact of more than one case, giving 74,181 contacts to trace. In discussion with contact tracing teams, duplicates were identified by matching name and residential location; for operational reasons, these individuals were recorded as a contact of only the earliest-identified primary case with which they were associated. The majority of cases for whom sex and age were available were female (n = 515, 55.8%), while median (interquartile range, IQR) age was 25 (13-38) years. Most contacts (64,545, 87.0%) were successfully traced, leading to the detection of 396 secondary cases. The median delay between last contact with the primary case and first contact by the CT teams was 4 days (interquartile range: 3-6 days).

Disease onset dates spanned the period 31 July 2018 to 26 April 2020, and was bimodally distributed, with two waves peaking in October 2018 and June 2019 (Figure 1A). The second wave followed a period of insecurity in this conflict-affected area that severely hampered response activities, including CT [26].

The median (IQR) number of contacts among all cases was 61 (18 - 120), but this was significantly lower during the first wave than the second (34 vs. 80, $p < 0.001$). Cases occurring in the first wave were more likely to have zero listed contacts than those in the second wave (31.3% vs. 9.6%, X^2 (1, $N = 603$) = 43.2, $p < 0.001$), and second wave cases were more likely to have a large number (> 100) of contacts (Figure 1B). 792 cases (85.8%) reported at least one contact (Figures 2 and 3), among whom the median (IQR) and mean number of contacts was 74 (36 - 134) and 102, respectively.

64,545 contacts (87.0%) were successfully traced, of whom 308 were confirmed as an EVD case and 88 died during follow-up. Therefore, the inferred total number of infected contacts was 396 (308 + 88), or 0.7% of the contacts successfully traced to completion of the follow-up period. Precise detail on the mechanism of identification of confirmed cases among contacts is not available; while we assume these were identified by contact tracers during follow-up, the role of other surveillance activities cannot be excluded.

There was substantial overdispersion in the offspring distribution of secondary cases, with 80% of onward transmission linked to just 13.9% (95%CI 11.4 – 16.2) of primary cases, and all secondary cases concentrated among the contacts of 207 (22.4%) cases. Further, just 99 (10.7%) primary cases led to more than one secondary case (Figures 2 and 4). We estimated k as 0.27 [95%CI 0.20 – 0.33].

Male contacts had slightly (but statistically significantly) greater odds of being lost to follow-up (odds ratio (OR) = 1.06, 95%CI 1.01 – 1.11, Table 1). Contacts in older age groups had significantly greater odds of being lost to follow-up compared to contacts in the youngest age group (0-15 years), with the greatest effect being observed among those aged 60 years and

243 older (OR = 1.65, 95%CI 1.47 – 1.86) and a marginally smaller effect among those aged 45-
244 59 years (OR = 1.55, 95%CI 1.43 – 1.69). Conversely, contacts traced during the second
245 wave had lower odds of being lost to follow-up (OR = 0.83, 95%CI 0.79 – 0.88).
246
247
248 **Capture-recapture modelling**
249
250 *How complete was CT for cases with at least one listed contact?*
251 Among cases with at least one contact listed, the best fitting distribution of the count of cases
252 with any contacts was given by the zero-truncated geometric model, which produced the
253 lowest AIC and BIC (Table S1(a), Supplementary Materials). This distribution was very
254 long-tailed (Figure 5), indicating that the majority of cases with contacts were successfully
255 detected, as with increasing mean of any count distribution, the probability for a zero count
256 becomes smaller. This is seen from the parameterization of the geometric distribution (see
257 Methods) where for $x=0$, i.e. the zero count, its estimated probability p_0 resolves the
258 expression $\frac{1}{1+\mu}(\frac{\mu}{1+\mu})^x$ to return $\frac{1}{1+\mu}$ where μ is the mean of the geometric model; the larger the
259 mean, the smaller the probability of $x=0$.
260
261 We estimated $\widehat{f0}$ (the unobserved number of cases with any contacts) = 8 [95%CI = 8-10],
262 where sample size (n) was 792 and p_0 found as 0.01. The sensitivity of CT to detect cases
263 with any contacts was therefore 792/(792+8) = 0.99 [95%CI 0.99-0.99]. There was no
264 difference in sensitivity by epidemic wave (wave 1 = 0.99 [95%CI 0.99-0.99]; wave 2 = 0.99
265 [95%CI 0.99-0.99]).
266
267 *How complete was CT for cases with infected contacts?*

Among cases with infected contacts, the best fitting distribution of the count of cases with infected contacts was again given by the zero-truncated geometric model, which produced the lowest AIC and BIC (Table S1(b), Supplementary Materials). This distribution is concentrated on the lower counts from 1 to 4 (Figure 6), indicating that a substantial proportion of cases with infected contacts may not have been detected.

We estimated $\widehat{f0}$ (the unobserved number of cases with infected contacts) = 227 [95%CI = 171-241], where sample size (n) was 207 and p_0 found as 0.52. The sensitivity of CT to detect cases with infected contacts was therefore 207/(207+227) = 0.49 [95%CI = 0.43-0.55]. There was a statistically significant difference in sensitivity by epidemic wave, with lower sensitivity during wave 1 (0.24 [95%CI = 0.11-0.38]) than during wave 2 (0.48 [95%CI = 0.40-0.56]).

Among the 792 cases with at least one listed contact, those cases with zero infected contacts had fewer contacts overall, were slightly older, and slightly more likely to be female compared to the other groups (Table 2).

Discussion

Our findings suggest that CT efforts were very successful at identifying cases with at least one contact, but much less successful at identifying cases with contacts who later develop symptoms. This is unsurprising, as the investigation component (typically by interview with cases under treatment and/or their caregivers) is easier to conduct than the tracing component (typically requiring daily visits to a large number of difficult to locate and mobile

individuals). This has important implications, as it is these infected contacts who contribute to ongoing chains of transmission when case investigation and contact tracing is inadequate, and in order to prioritise scarce resources, control efforts should target those cases among whose contacts secondary infections arise [20,21,27]. A high proportion of cases listed at least one contact (~85%), compared to 27% and 44% during EVD outbreaks in Liberia [28] and Sierra Leone [27], suggesting that lessons about enhancing the quality of CT were learned from previous EVD outbreaks [4,5,27,28].

Cases with infected contacts had more contacts on average, which may result from three possible explanations. Firstly, cases with more contacts are more likely to have at least one infected contact among these. Secondly, fewer overall listed contacts may be the result of poorly-conducted case investigations. We found some evidence in support of this, with the mean number of contacts increasing as the epidemic progressed, indicating possible improvements in case investigation quality over time as staff became more accustomed to the procedure and community trust and engagement in the response improved [29].

Thirdly, cases with infected contacts may differ from other cases; in this study, such cases were younger and more likely to be male - demographic factors that have been previously shown to impact transmission of EVD and other diseases [30–32]. Indeed, cases with more contacts have been shown to play a greater role in disease transmission and are more likely to have infected contacts [33,34]. This is particularly true of diseases that demonstrate heterogeneous transmission, including EVD and COVID-19, and our results suggest a high degree of overdispersion and “superspreading”, in line with what has previously been reported during large EVD outbreaks [23]. Overdispersion can lead to rapid expansion, particularly among hidden chains of transmission, and one promising area of research is to

identify correlates of "superspreading" to better target limited resources for greatest impact. Indeed, prior research suggests that, if highly infectious individuals can be predictively identified and targeted, the efficiency of control can be greatly enhanced, such that focussing half of all control effort on the most infectious 20% of cases can improve effectiveness up to threefold [20,21].

While it is possible to estimate the number of unobserved cases with (infected) contacts, it is not possible to identify whether these have been misclassified as having zero (infected) contacts or if they were undetected by the surveillance system in general. However, the greater probability of having zero contacts listed during the first epidemic wave suggests substantial misclassification and suboptimal performance in the period during which surveillance activities were being established, as reported during previous EVD outbreaks [4,27,28]. Indeed, the sensitivity of CT to detect cases with infected contacts was lower, and loss to follow-up greater, during the first epidemic wave, indicating quality improvements of this activity over time, either because the ability to conduct contact follow-up was hampered by the insecurity experienced during the first wave, or because of greater familiarity with, and acceptance of, the process among CT staff and the local population.

Limitations

While the method described herein proposes a robust framework to assess the sensitivity of contact tracing, there are important limitations. There is no gold standard list of contacts against which to validate this method, but the method itself has been validated to estimate

true population size in a variety of other settings [25]. The dataset does not permit the distinction between cases who were confirmed to have no contacts after a thorough case investigation and cases having no listed contacts due to no (or inadequate) case investigation. However, our method may in fact help to identify the magnitude of the misclassification arising from this. The inferences made are exclusively informed by the definition of cases as defined by CT protocols; for example, our results would not inform the sensitivity of CT as applied to asymptomatic EVD cases if these individuals are not part of the testing strategy. Differences in performance between agents could result in strong heterogeneity in the count distribution, which might be detectable. For this reason, we applied Chao's estimator (which allows for heterogeneity), and only if this was significantly different from the model-based estimate would we consider that there is an issue. In our results, this was not the case (Supplementary tables S2 and S3). Finally, we have not adjusted for observed heterogeneity, such as age, sex, profession, geographic location of the cases and delays within the contact tracing process. Further work is planned to incorporate such considerations.

Conclusions

In conclusion, CT is crucial to the containment of certain disease outbreaks. However, as with many surveillance activities, CT has the potential to suffer reduced effectiveness from underreporting and poor sensitivity [4,27,28]. The consequences of poor ascertainment and misclassification can be disastrous in the containment stages of an outbreak, potentially creating explosive expansion among hidden chains of transmission, particularly during containment and de-escalation phases.

We have described a novel application of CRC models to estimate a crucial yet elusive performance indicator of a key component of the public health response to epidemics, namely the sensitivity of contact tracing, as applied to a recent outbreak of EVD. The method demonstrated that the majority of cases with any contacts were observed, suggesting that the case investigation component of CT performed well, while less than half of cases with infected contacts were observed, suggesting that the contact follow-up component of CT performed poorly in this setting. The approach described is disease-agnostic, and can be extended to assess the sensitivity of CT for any disease, including COVID-19, for which CT has been identified as a crucial component of the response activities.

Acknowledgments

We thank the population for their participation conducting this study. We acknowledge the enormous dedication of the various organisations and individuals that responded and provided healthcare to this population and supported the public health response during this outbreak. The authors received no specific funding for this work. The authors alone are responsible for the views expressed in this article and they do not necessarily represent the views, decisions or policies of the institutions with which they are affiliated.

Author Bio

Jonathan Polonsky is an epidemiologist with the World Health Organization, Geneva, Switzerland. His research interests include various aspects of public health response in humanitarian and emergency settings, with a primary focus on infectious disease outbreaks and monitoring the health impact of crises.

References

1. Infection prevention and control: Contact tracing. [cited 10 May 2021]. Available: https://www.who.int/news-room/q-a-detail/contact-tracing

2. Polonsky JA, Baidjoe A, Kamvar ZN, Cori A, Durski K, Edmunds WJ, et al. Outbreak analytics: a developing data science for informing the response to emerging pathogens. Philos Trans R Soc B Biol Sci. 2019;374: 20180276. doi:10.1098/rstb.2018.0276

3. Polonsky JA, Ivey M, Mazhar MKA, Rahman Z, Waroux O le P de, Karo B, et al. Epidemiological, clinical, and public health response characteristics of a large outbreak of diphtheria among the Rohingya population in Cox's Bazar, Bangladesh, 2017 to 2019: A retrospective study. PLOS Med. 2021;18: e1003587. doi:10.1371/journal.pmed.1003587

4. Swanson KC, Altare C, Wesseh CS, Nyenswah T, Ahmed T, Eyal N, et al. Contact tracing performance during the Ebola epidemic in Liberia, 2014-2015. PLoS Negl Trop Dis. 2018;12. doi:10.1371/journal.pntd.0006762

5. Polonsky J, Mboussou F, Haskew C. Lessons learnt from Ebola virus disease surveillance in Équateur Province, May–July 2018. Wkly Epidemiol Rec. 2019; 5.

6. Sikakulya FK, Mulisya O, Munyambalu DK, Bunduki GK. Ebola in the Eastern Democratic Republic of Congo: One Health approach to infectious disease control. One Health Amst Neth. 2020;9: 100117. doi:10.1016/j.onehlt.2019.100117

7. Keeling MJ, Hollingsworth TD, Read JM. Efficacy of contact tracing for the containment of the 2019 novel coronavirus (COVID-19). J Epidemiol Community Health. 2020;74: 861–866. doi:10.1136/jech-2020-214051

8. Althoff KN, Coburn SB, Nash D. Contact Tracing: Essential to the Public Health Response and Our Understanding of the Epidemiology of Coronavirus Disease 2019. Clin Infect Dis Off Publ Infect Dis Soc Am. 2020;71: 1960–1961. doi:10.1093/cid/ciaa757

9. Ilunga Kalenga O, Moeti M, Sparrow A, Nguyen V-K, Lucey D, Ghebreyesus TA. The Ongoing Ebola Epidemic in the Democratic Republic of Congo, 2018-2019. N Engl J Med. 2019;381: 373–383. doi:10.1056/NEJMsr1904253

10. Team WER. Ebola Virus Disease in West Africa — The First 9 Months of the Epidemic and Forward Projections. In: https://doi.org/10.1056/NEJMoa1411100 [Internet]. Massachusetts Medical Society; 15 Oct 2014 [cited 24 Apr 2021]. doi:10.1056/NEJMoa1411100

11. Polonsky JA, Wamala JF, Clerck H de, Herp MV, Sprecher A, Porten K, et al. Emerging Filoviral Disease in Uganda: Proposed Explanations and Research Directions. Am J Trop Med Hyg. 2014;90: 790–793. doi:10.4269/ajtmh.13-0374

12. Barry A, Ahuka-Mundeke S, Ahmed YA, Allarangar Y, Anoko J, Archer BN, et al. Outbreak of Ebola virus disease in the Democratic Republic of the Congo, April–May,

2018: an epidemiological study. The Lancet. 2018;392: 213–221. doi:10.1016/S0140-6736(18)31387-4

13. WHO | Implementation and management of contact tracing for Ebola virus disease. In: WHO [Internet]. World Health Organization; [cited 10 May 2021]. Available: http://www.who.int/csr/resources/publications/ebola/contact-tracing/en/

14. Vergne T, Del Rio Vilas VJ, Cameron A, Dufour B, Grosbois V. Capture-recapture approaches and the surveillance of livestock diseases: A review. Prev Vet Med. 2015;120: 253–264. doi:10.1016/j.prevetmed.2015.04.003

15. Bohning D, Heijden PGM, Bunge J. Capture-Recapture Methods for the Social and Medical Sciences. 2018.

16. Gignoux E, Idowu R, Bawo L, Hurum L, Sprecher A, Bastard M, et al. Use of Capture–Recapture to Estimate Underreporting of Ebola Virus Disease, Montserrado County, Liberia. Emerg Infect Dis. 2015;21: 2265–2267. doi:10.3201/eid2112.150756

17. Isanaka S, Hedt-Gauthier BL, Salou H, Berthé F, Grais RF, Allen BGS. Active and adaptive case finding to estimate therapeutic program coverage for severe acute malnutrition: a capture-recapture study. BMC Health Serv Res. 2019;19. doi:10.1186/s12913-019-4791-9

18. Medley AM, Mavila O, Makumbi I, Nizeyemana F, Umutoni A, Balisanga H, et al. Case Definitions Used During the First 6 Months of the 10th Ebola Virus Disease Outbreak in the Democratic Republic of the Congo - Four Neighboring Countries, August 2018-February 2019. MMWR Morb Mortal Wkly Rep. 2020;69: 14–19. doi:10.15585/mmwr.mm6901a4

19. WHO | Case definition recommendations for Ebola or Marburg virus diseases. In: WHO [Internet]. World Health Organization; [cited 10 May 2021]. Available: https://www.who.int/csr/resources/publications/ebola/case-definition/en/

20. Woolhouse MEJ, Dye C, Etard J-F, Smith T, Charlwood JD, Garnett GP, et al. Heterogeneities in the transmission of infectious agents: Implications for the design of control programs. Proc Natl Acad Sci. 1997;94: 338–342. doi:10.1073/pnas.94.1.338

21. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect of individual variation on disease emergence. Nature. 2005;438: 355–359. doi:10.1038/nature04153

22. Endo A, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, Abbott S, Kucharski AJ, Funk S. Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. Wellcome Open Res. 2020;5: 67. doi:10.12688/wellcomeopenres.15842.3

23. Althaus CL. Ebola superspreading. Lancet Infect Dis. 2015;15: 507–508. doi:10.1016/S1473-3099(15)70135-0

24. Böhning D, Rocchetti I, Maruotti A, Holling H. Estimating the undetected infections in the Covid-19 outbreak by harnessing capture-recapture methods. Int J Infect Dis IJID Off Publ Int Soc Infect Dis. 2020;97: 197–201. doi:10.1016/j.ijid.2020.06.009

25. Alfò M, Böhning D, Rocchetti I. Upper bound estimators of the population size based on ordinal models for capture-recapture experiments. Biometrics. 2021;77: 237–248. doi:10.1111/biom.13265

26. Jombart T, Jarvis CI, Mesfin S, Tabal N, Mossoko M, Mpia LM, et al. The cost of insecurity: from flare-up to control of a major Ebola virus disease hotspot during the outbreak in the Democratic Republic of the Congo, 2019. Euro Surveill Bull Eur Sur Mal Transm Eur Commun Dis Bull. 2020;25. doi:10.2807/1560-7917.ES.2020.25.2.1900735

27. Senga M, Koi A, Moses L, Wauquier N, Barboza P, Fernandez-Garcia MD, et al. Contact tracing performance during the Ebola virus disease outbreak in Kenema district, Sierra Leone. Philos Trans R Soc Lond B Biol Sci. 2017;372. doi:10.1098/rstb.2016.0300

28. Olu OO, Lamunu M, Nanyunja M, Dafae F, Samba T, Sempiira N, et al. Contact Tracing during an Outbreak of Ebola Virus Disease in the Western Area Districts of Sierra Leone: Lessons for Future Ebola Outbreak Response. Front Public Health. 2016;4: 130. doi:10.3389/fpubh.2016.00130

29. Masumbuko Claude K, Underschultz J, Hawkes MT. Social resistance drives persistent transmission of Ebola virus disease in Eastern Democratic Republic of Congo: A mixed-methods study. Schieffelin J, editor. PLOS ONE. 2019;14: e0223104. doi:10.1371/journal.pone.0223104

30. Kiti MC, Kinyanjui TM, Koech DC, Munywoki PK, Medley GF, Nokes DJ. Quantifying Age-Related Rates of Social Contact Using Diaries in a Rural Coastal Population of Kenya. PLoS ONE. 2014;9. doi:10.1371/journal.pone.0104786

31. Johnstone-Robertson SP, Mark D, Morrow C, Middelkoop K, Chiswell M, Aquino LDH, et al. Social mixing patterns within a South African township community: implications for respiratory disease transmission and control. Am J Epidemiol. 2011;174: 1246–1255. doi:10.1093/aje/kwr251

32. le Polain de Waroux O, Cohuet S, Ndazima D, Kucharski AJ, Juan-Giner A, Flasche S, et al. Characteristics of human encounters and social mixing patterns relevant to infectious diseases spread by close contact: a survey in Southwest Uganda. BMC Infect Dis. 2018;18: 172. doi:10.1186/s12879-018-3073-1

33. Kangbai JB. Social network analysis and modeling of cellphone-based syndromic surveillance data for Ebola in Sierra Leone. Asian Pac J Trop Med. 2016;9: 851–855. doi:10.1016/j.apjtm.2016.07.005

34. Hagel C, Weidemann F, Gauch S, Edwards S, Tinnemann P. Analysing published global Ebola Virus Disease research using social network analysis. PLoS Negl Trop Dis. 2017;11: e0005747. doi:10.1371/journal.pntd.0005747

Address for correspondence: Victor J. Del Rio Vilas, World Health Organization, South East Asia Regional Office, New Delhi, India; email: delriov@who.int

Table 1. Multivariable logistic regression for predictors of loss to follow-up of contacts of Ebola Virus Disease cases, Beni Health Zone, Democratic Republic of the Congo, 31 July 2018 - 26 April 2020**.**

			Unadjusted		Adjusted	
Independent variable	**Level**	**N**	**OR (95% CI)**	**p-value**	**OR (95% CI)**	**p-value**
Sex	**Female**	41349	*Reference*	-	*Reference*	-
	Male	37296	1.07 (1.02 – 1.12)	**0.003**	1.06 (1.01 – 1.11)	**0.013**
Age group	**0-14**	20616	*Reference*	-	*Reference*	-
	15-29	26142	1.18 (1.11 – 1.25)	**<0.001**	1.19 (1.12 – 1.27)	**<0.001**
	30-44	17665	1.16 (1.09 – 1.24)	**<0.001**	1.18 (1.10 – 1.26)	**<0.001**
	45-59	6157	1.56 (1.43 – 1.70)	**<0.001**	1.55 (1.43 – 1.69)	**<0.001**
	60+	2599	1.64 (1.46 – 1.84)	**<0.001**	1.65 (1.47 – 1.86)	**<0.001**
Epidemic wave*	**First wave**	14374	*Reference*	-	*Reference*	-
	Second wave	66182	0.85 (0.81 – 0.90)	**<0.001**	0.83 (0.79 – 0.88)	**<0.001**

* Contacts were divided into 2 epidemic waves, according to the date of symptom onset of their associated primary case (first wave = 31 July 2018 – 28 February 2019, second wave = 1 March 2019 - 26 April 2020).

Table 2: Distribution of cases, their median age and sex ratio, and mean total number of contacts, grouped by number of infected contacts, among cases with at least one listed contact, Beni Health Zone, Democratic Republic of the Congo, 31 July 2018 - 26 April 2020.

# infected contacts	# cases	Median age of cases	% Female among cases	Mean (95%CI) total number of contacts
0	585	28.2	59.5	85.7 (79.1 - 92.4)
1	108	23.6	56.7	122 (102.0 - 141.0)
2+	99	25.6	54.6	174 (144.0 - 204.0)

Figure 1A: Epidemic curve by date of symptom onset among Ebola Virus Disease cases, Beni Health Zone, Democratic Republic of the Congo, 31 July 2018 - 26 April 2020. Cases and contacts were divided into 2 epidemic waves, according to the date of symptom onset among cases (first wave = 31 July 2018 – 28 February 2019, second wave = 1 March 2019 - 26 April 2020).

Figure 1B: Distribution of dates of symptom onset among cases, by number of listed contacts. Data were smoothed using a non-parametric (Gaussian) kernel-based estimate, with automatic bandwidth selection (37.6 days).

Figure 2: Flowchart showing breakdown of observed cases by number of listed and infected contacts among Ebola Virus Disease cases, Beni Health Zone, Democratic Republic of the Congo, 31 July 2018 - 26 April 2020.

Figure 3: Frequency distribution of Ebola Virus Disease cases by number of listed contacts, Beni Health Zone, Democratic Republic of the Congo, 31 July 2018 - 26 April 2020.

Figure 4: Frequency distribution of Ebola Virus Disease cases with infected contacts by number of infected contacts, Beni Health Zone, Democratic Republic of the Congo, 31 July 2018 - 26 April 2020.

Figure 5: Observed and fitted (geometric) zero-truncated distribution of the total number of contacts for cases with *at least one contact listed*.

Figure 6: Observed and fitted (geometric) zero-truncated distribution of the total number of infected contacts for cases with *at least one infected contact listed*.