9-3-2015

Characterizing Ebola Transmission Patterns based on Internet News Reports.

J. C. Cleaton

Cecile Viboud

Lone Simonsen
George Washington University

A. M. Hurtado

G. Chowell

Clinical Infectious Diseases Advance Access published September 3, 2015

Characterizing Ebola Transmission Patterns based on Internet News Reports

J. C. Cleaton[1,2], C. Viboud[3], L. Simonsen[3,4,5], A.M. Hurtado[2], G. Chowell[1,3]

[1]School of Public Health, Georgia State University, Atlanta, Georgia, USA

[2]College of Liberal Arts and Sciences, Arizona State University, Tempe, Arizona, USA

[3]Division of International Epidemiology and Population Studies, Fogarty International Center, National Institutes of Health, Bethesda, Maryland, USA

[4]Department of Public Health, University of Copenhagen, Copenhagen, Denmark

[5]Department of Global Health, Milken Institute School of Public Health, George Washington University, Washington DC, USA

Corresponding author: Dr. Gerardo Chowell, Email: gchowell@gsu.edu, Fax: 404-413-2344, Ph: 404-413-9322

40-word summary of your article's main point:
We have provided further evidence that information derived from the systematic collection and analysis of data from unstructured news reports from authoritative sources could provide a reliable means to assess epidemiological patterns of evolving infectious disease threats in near real-time.

Abstract

Background: Detailed information on patient exposure, contact patterns, and discharge status, is rarely available in real time from traditional surveillance systems in the context of an emerging infectious disease outbreak. Here we validate the systematic collection of Internet news reports to characterize epidemiological patterns of Ebola virus disease (EVD) infections during the West African 2014-2015 outbreak.

Methods: Based on 58 news reports, we analyzed a total of 79 EVD clusters (286 cases) of size ranging from 1 to 33 cases between January 2014 and February 2015 in Guinea, Sierra Leone and Liberia.

Results and Conclusions: The great majority of reported exposures stemmed from contact with family members (57.3%) followed by hospitals (18.2%) and funerals (12.7%). Our data indicated that funeral exposure was significantly more frequent in Sierra Leone (27.3%) followed by Guinea (18.2%) and Liberia (1.8%) (Chi-square test; P<0.0001). Funeral transmission was the dominant route of transmission until April 2014 (60%) and was replaced with hospital exposure in June-July 2014 (70%), both of which declined after interventions were put in place. The mean reproduction number of the outbreak was 2.3 (95% CI: 1.8, 2.7). The case fatality rate was estimated at 74.4% (95% CI: 68.3, 79.8). Overall our findings based on news reports are in close agreement with those derived from traditional epidemiological surveillance data and with those reported for prior outbreaks. Our findings support the use of real time-information from trustworthy news reports to provide timely estimates of key epidemiological parameters that may be hard to ascertain otherwise.

Introduction

The unprecedented Ebola epidemic in West Africa may have been triggered by a single cross-species transmission event from a fruit bat, an event that has been traced back to December 2013 in a forested Guinean region (1). The Ebola outbreak was not reported to the World Health Organization (WHO) until late March 2014, at which point the outbreak had crossed the porous borders of neighboring Sierra Leone and Liberia (2, 3). The epidemic gained speed over the summer and started abating as behavioral and medical interventions were put in place in fall 2014. As of July 15 2015, the 27,678 illnesses and 11,276 deaths were reported to WHO (3), with the overwhelming majority of cases reported in West Africa and likely an underestimate.

The epidemic has now tailed off, with low incidence counts reported in the region since April 2015 (3). A great amount of knowledge has been generated on the epidemiology and clinical picture of the disease (e.g., (4-9)). However, the nature of and changes in the characteristics of the contact networks and heterogeneity in risk in various exposure settings remain unclear, with information limited in geographic and temporal scope (10, 11). Such critical information was not available to inform mathematical models employed to project the likely trajectory of the outbreak and the likely benefits of various control interventions.

Here we contribute to the growing body of digital epidemiology studies that make use of non-traditional online data sources to enhance the detection, forecasting, and response to infectious disease threats (12-15). Specifically, we show that systematic collection and analysis of Internet news reports from authoritative media outlets and public health authorities can overcome the scarce and patchy information on exposure patterns and chains of transmission available from formal epidemiological surveillance efforts.

Materials and Methods

We reviewed news stories and investigative reports published between March 2014 and March 2015 describing suspected, probable and confirmed cases of Ebola in the three most affected countries (Guinea, Sierra Leone and Liberia) from the WHO website as well as online authoritative media outlets (Supplement). For this purpose, we screened articles from the Ebola section of news sources or used the search keyword: Ebola and perused the links available in retrieved articles for further information. We selected those articles that contained detailed stories about Ebola case clusters arising within families or through funerals and hospitals. Specifically, we selected 21 articles and 2 videos out of 624 *The New York Times* articles from the section Times Topics: the Ebola Outbreak in West Africa published between Mar 25 2014 and Mar 31 2015. We also screened 781 articles from the *The Washington Post* published between Mar 22 2014 and Mar 31 2015 and retrieved by searching the keyword "ebola" anywhere in the article. Of these, we used information from 6 articles to populate our database. We also retrieved useful information from 6 out of 211 EbolaDeeply articles published from Oct 15 2014 to Mar 31 2015 (http://www.eboladeeply.org/). Finally, from the WHO website, we screened 60 articles in the section "Stories from the field on Ebola," and 96 WHO situational reports published between March 2014 and March 2015. Of these, we retrieved relevant information from 11 articles and 5 WHO situation reports.

Each article was carefully assessed and demographic, epidemiologic and clinical information about each case was systematically extracted including: patient age and gender, country where the patient acquired EVD, month and year of symptoms onset, occupational/health-care worker status, disease outcome (recovery or death), nature of exposure to the Ebola virus (family/household, hospital, funeral, care, sexual, hazardous waste, childbirth, zoonotic), as well as the number of secondary cases generated by the index case in each cluster. If the date of symptoms onset was not given in the article we used the article's publication date.

Each Ebola patient was assigned one or several types of Ebola exposures based on information provided in the articles, including family/household, hospital, funeral, care, hazardous waste, sexual, as detailed below. "Family/household" exposure indicates that the family members of the case were ill with Ebola within three weeks before his/her symptoms began or that the article stated that the case got sick with Ebola from contact with a family member. Of note, people living in the same household, such as couples living together, were also considered family. "Hospital" exposure indicates that the case (e.g., patient, visitor, health-care worker) got infected after spending time in health-care settings including hospitals, clinics, Ebola treatment units (ETUs), or isolation facilities.

“Funeral” exposure indicates that the case was exposed to the Ebola virus while preparing for or attending a funeral by touching the body of an Ebola deceased person within the previous 3 weeks. “Care” exposure indicates that a patient was infected after helping another EVD patient outside of the hospital setting, including through transportation to a health-care setting or attending to an Ebola case. This category excludes health-care workers exposed in hospital settings, but does include family members who were stated as taking care of one another. “Hazardous waste” indicates that the article stated that the case got sick with Ebola after handling or cleaning waste products of an Ebola patient. “Sexual exposure” indicates that the case had not been in direct contact with an active EVD case in the past three weeks but had sexual contact with an Ebola recovered case. “Childbirth” indicates that a pregnant woman infected with EVD gave birth and transmitted EVD to the newborn during labor without further contact with the mother or other EVD patients. Patients were considered health-care workers if they worked in health-care settings as doctors, nurses, hygienists, lab technicians, hospital administrators, or hospital guards. Cases were considered hospitalized if they were admitted to a hospital, clinic, ETU, or isolation facility. Some patients were exposed in multiple settings, for example through both care and family, and those cases were counted in both categories. Whenever the means of exposure could not be identified from individual articles, we indicated the exposure information as “unknown.” Exposure was indicated as “zoonotic” for the purported index case (single case) of the epidemic in December 2013 (2) .

Transmission chains of Ebola infection were explicitly provided in the article or inferred based on chronological information on the timing of symptoms of successive cases available from all except for 8 of the selected articles. We also analyzed the number of secondary cases generated by an index case in each Ebola cluster, which was explicitly stated for most clusters, as these were typically limited to family or hospital transmission.

Results

We analyzed a total of 79 EVD clusters (286 cases) that occurred in Guinea, Sierra Leone and Liberia between January 2014 and February 2015 (Supplementary Table S1). A total of 8 clusters with known exposures originated in Guinea, 21 in Sierra Leone and 34 in Liberia (Table S1). All of these clusters except 3 involved a single country only. There was a similar fraction of male and female cases (53.7% vs. 46.3%), which was consistent across the three countries (Chi-square test, P=0.76). The reported cluster size ranged from 1 to 33 cases and included up to 6 disease generations. The mean cluster size was estimated at 4.3 (95% CI: 3.2, 5.3) by fitting a negative binomial distribution to the cluster sizes (Supplementary Figure S1). The curve of monthly cases in our sample was significantly correlated with the outbreak trajectory as reported by WHO (Figure 1; Spearman rho=0.59, P=0.026).

Guinea had the highest rate of cases with unknown exposures (32.3%) followed by Liberia (21.4%) and Sierra Leone (18.5%). The monthly proportion of cases with unknown exposures did not vary significantly during the course of the epidemic (Spearman rho=0.16,P=0.6). Only 5.2% of the cases had multiple exposures, but this proportion did not vary significantly across countries (Chi-square test; P=0.31). The great majority of reported exposures stemmed from contact with family members (57.3%) followed by hospitals (18.2%) and funerals (12.7%) (Table 1). The proportion of cases reporting family exposures was similar between the three countries (Chi-square test; P=0.16). The overall proportion of family exposure was higher among females compared to males (75.8% vs. 54.5%; Chi-square test; P=0.004). The Case fatality rate (CFR) was measured to be 74.4% (95% CI: 68.3, 79.8) across all identified cases; there was no difference in CFR between patients reporting family exposure and non-family exposure (Chi-square test; P=0.46), or between countries (G: 72.4%, S: 72.7%, L: 75.8%; Chi-square test, P=0.87).

The proportion of hospital exposures varied significantly across the three countries with highest proportion reported in Liberia (24.5%) followed by Sierra Leone (13.6%) and Guinea (9.1%) (Chi-square test; P<0.04). A greater fraction of hospital-based transmission involved male individuals compared to females (Chi-square test; P<0.0001). We did not find a significantly different case fatality rate between individuals with and without hospital exposure (Chi-square test; P=0.503)

Funeral exposure was significantly more frequent in Sierra Leone (27.3%) followed by Guinea (18.2%) and Liberia (1.8%) (Chi-square test; P<0.0001). A total of 9 clusters in our data included funeral transmission; in 6 of these 9, funeral transmission was the initial exposure event. Moreover, we did not find a significant difference in the proportion of funeral exposures by gender (Chi-square test; P<0.133).

Overall, the proportion of hospital exposures peaked in April 2014 (70%) and declined to low levels during the subsequent epidemic months (Figure 2). Similarly, the proportion of funeral exposures was highest (60%) in the months of June-July 2014 and then declined during the later months of the epidemic (Figure 2). By August 2014, the family/household became the dominant setting of exposure, accounting for 60% of new cases.

Our data included a total of 46 cases that were health-care workers (16.1%). Of note, only 7 health-care workers in a total of 5 clusters transmitted EVD to 15 individuals: 9 through family exposure, 2 through care outside of the hospital, and 4 through hospital exposure. The proportion of health-care worker cases peaked in June 2014 and declined over the subsequent months of the epidemic (Figure 3). Moreover, health-care workers were more likely to be hospitalized in hospitals or Ebola treatment units compared to non-health care workers (91.7% vs. 59.6%; Chi-square test, P=0.003).

Based on the identified chains of transmission, we estimated 39 individual level reproduction numbers, ranging from 1 to 7. By fitting a negative binomial distribution to our set of reproduction numbers, we estimated the mean reproduction number at 2.3

(95% CI: 1.8, 2.7) (Figure 4). Estimates of the reproduction number did not vary significantly across countries (ANOVA test, P=0.75). The mean R from data during the months January-July 2014 (R=2.8, 95%CI: 1.2, 4.4) was not significantly different to the mean R estimated based on the remainder of the epidemic during August 2014-February 2015 (R=2.1, 95%CI: 1.7, 2.5) (Figure 5).

We did not find a significant change in the monthly estimates of the case fatality rate over the course of the epidemic (Chi-square test; P=0.09). Similarly, the age-specific case fatality rate appeared higher (70% and 82%) among those <15 years and >45 years of age, compared to 56% among individuals aged 15-44 years (Supplementary Figure S2). However, this age difference did not reach statistical significance (Chi-square test; P=0.15) as our sample was reduced to a small fraction of cases with available age information (29%).

Discussion

Here we have harnessed the potential of real-time publicly available Internet reports from the WHO website as well as authoritative media outlets to identify key information on exposure and transmission patterns during the 2014-15 West African Ebola epidemic. Our findings on exposure patterns are in good agreement with those derived from epidemiological surveillance data (10) and with those reported for prior outbreaks [1-3,4], highlighting the predominance of hospital and funeral transmission in the pre-intervention period. Similarly, trends in the frequency of hospital exposures aligned well with the proportion of patients among health-care workers from official WHO statistics. Our mean estimate of the reproduction number based on cluster data lies toward the higher end of the range of previously published estimates around 1.5-2.5 (7, 8, 16-18). This is likely explained as a reporting bias, that news reports favor larger clusters.

Faye et al. (10) analyzed a transmission tree composed of 152 Ebola cases from three areas of Guinea between February and August 2014. The relative exposure patterns for Guinea in our data are consistent with those reported by Faye et al (10). However, our dataset included cases up to February 2015 (10). Specifically, Faye et al. found a predominance of community exposure (72% through family), followed by 13% of hospital exposure, and 13% of funeral exposure (10). In comparison, our Guinean data indicated that 50% of cases were exposed through family, 9.1% via hospitals, and 18.2% via funerals. Moreover, Faye et al. (10) reported the proportion of health-care workers in the transmission tree at 14%, which is similar to the proportion in our data for Guinea (16.7%). The hospitalization rate in our data was also similar to that reported by Faye et al. (80% vs. 81%) (10).

Overall, exposure via family contacts (57%) in our data was the most frequent during the epidemic, which is in line with exposure patterns from prior Ebola outbreaks (19-22) and chains of transmission for the ongoing epidemic in Guinea (February-August 2014) (10). In contrast, a limited outbreak of EVD in Nigeria stemming from a single case importation originating from Liberia was amplified by hospital exposures that primarily affected health-care workers (11/20) (11, 23).

The proportion of funeral and hospital exposures in our data peaked in the months of April and June 2014, respectively, which is a signal of super-spreading events that facilitated rapid spread of the virus across communities. Hospital exposures declined considerably after July 2014 likely as a result of the improvement in infection control measures in health-care settings due to enhanced infection control measures. Similarly, funeral exposures occurred sporadically in later epidemic months, which suggest the positive effect of massive educational campaigns undertaken in the region. The decline in hospital-based transmission is in line with a decline in the proportion of health-care workers in our data and statistics retrieved from the WHO situational reports time (24). Due to our small sample size, it was not possible to characterize temporal changes in exposure patterns by country; however temporal trends in frequency of nosocomial and funeral exposure reported in news articles are a good marker of the effectiveness of interventions.

Sexual transmission appears to have been infrequent during the epidemic, associated with 2.0% of known exposures in our data, but this transmission pathway remains of concern now as the epidemic is waning, because of its potential to generate flare-ups of the disease even after active Ebola transmission chains have been halted. Continued vigilance and monitoring of the populations months after the end of the epidemic will be critical to ensure that disease flare-ups are quickly contained.

Our mean estimate of the reproduction number at 2.3 lies in the higher end of published estimates of the reproduction number including estimates for past outbreaks in Central Africa (25, 26) and estimates derived for the 2014-15 Ebola epidemic using incidence time series in West Africa (7, 8, 16-18) or a transmission tree of Ebola cases in Guinea during March-August 2014 (10). Our estimate is also in agreement with those reported for prior Ebola outbreaks (25, 26). Nevertheless, individual level estimates of the reproduction number for Ebola show substantial heterogeneity (range 1-7 in our data) which underscores the potential role of super-spreading events in sustaining local transmission and a potential opportunity to guide intervention strategies (27). Moreover, it is worth noting that the reproduction number of the Ebola epidemic at district level tends to quickly decline after just a few generations of infections. This could result from the natural history of disease transmission in a highly clustered network (28), or else the benefits of control measures/ interventions combined with behavior changes (6). We were, however, not able to characterize a particular temporal pattern in the effective reproduction number owing to our small sample size.

The higher proportion of cases reported for Liberia (49%) relative to Sierra Leone (28.3%) and Guinea (22.7%) in our news sample likely reflects an American-centric bias towards reporting cases from a country with strong ties to the U.S. Moreover, age information was missing in 71% of the cases, hampering our ability to analyze the transmission patterns according to age. Further, our sample contained a higher proportion of health-care workers (16.1%) compared with official WHO statistics (3.5%). Other potential biases might exist towards writing survivor stories or sensational articles. However, it is reassuring that our regional estimate of the case fatality rate at 74.4%, and associated age patterns, are in good agreement with that reported by the WHO Ebola Response Team (4, 29). Finally, we focused our search on two major newspapers, an Ebola-specific website, and WHO reports, so as to maximize the amount of detailed information on Ebola clusters provided by original news stories. Other media sources such as Twitter and Promed have also provided useful epidemiological data (12, 30) in past outbreaks; however in our experience these media did not focus on detailed stories of Ebola transmission within individual clusters. Casting a broader net would have probably resulted in a larger sample size but would have required systematic text processing to discard noise from relevant information.

Only a few studies have assessed the role of real-time non-traditional online data sources in elucidating epidemiological characteristics of epidemic outbreaks. For instance, disease outbreaks captured from Internet data collected by HealthMap (31) have been shown to provide epidemiological information ahead of formal epidemiological surveillance reports including during the early phase of the 2010 Haitian cholera outbreak (32, 33). More recently, an association was reported between the reproduction number and the volume of news reports documenting public health interventions and aggravating events relating to the Ebola epidemic in Sierra Leone and Liberia (34). Here we have provided further evidence that information derived from the systematic collection and analysis of data from unstructured news reports from authoritative sources may provide a reliable means to assess epidemiological patterns of evolving infectious disease threats in near real-time.

Funding information

CV and GC acknowledge the financial support from the Division of International Epidemiology and Population Studies, The Fogarty International Center, United States National Institutes of Health, funded in part by the Office of Pandemics and Emerging Threats at the United States Department of Health and Human Services. GC also acknowledges support from grants NSF grant 1414374 as part of the joint NSF-NIH-USDA Ecology and Evolution of Infectious Diseases programme, United Kingdom Biotechnology and Biological Sciences Research Council grant BB/M008894/1, RAPID NFS Grants#1518939 and NSF# 1518529 as well as NSF-IIS Grant#1518939. LS, CV and GC acknowledge support from the Research And Policy for Infectious Disease Dynamics program (RAPIDD) of the United States Department of Homeland Security, and LS support from the Lundbeck Foundation, Denmark.

Conflict of interest

Authors declare no conflict of interest.

References

1. Baize S, Pannetier D, Oestereich L, Rieger T, Koivogui L, Magassouba N, et al. Emergence of Zaire Ebola Virus Disease in Guinea - Preliminary Report. The New England journal of medicine. 2014;371(15):1418-25. doi: 10.1056/NEJMoa1404505. PubMed PMID: 24738640.
2. Richards P, Amara J, Ferme MC, Kamara P, Mokuwa E, Sheriff AI, et al. Social pathways for ebola virus disease in rural sierra leone, and some implications for containment. PLoS Negl Trop Dis. 2015;9(4):e0003567. doi: 10.1371/journal.pntd.0003567. PubMed PMID: 25886400; PubMed Central PMCID: PMC4401769.
3. Ebola response roadmap - Situation report - 15 July 2015. Available from: http://apps.who.int/ebola/current-situation/ebola-situation-report-15-july-2015 (accessed on 15 July 2015) 2015. Available from: http://apps.who.int/ebola/en/current-situation/ebola-situation-report-6-may-2015.
4. Team WHOER. Ebola Virus Disease in West Africa - The First 9 Months of the Epidemic and Forward Projections. The New England journal of medicine. 2014;371(16):1481-95. doi: 10.1056/NEJMoa1411100. PubMed PMID: 25244186.
5. Team WHOER, Agua-Agum J, Ariyarajah A, Blake IM, Cori A, Donnelly CA, et al. Ebola virus disease among children in West Africa. The New England journal of medicine. 2015;372(13):1274-7. doi: 10.1056/NEJMc1415318. PubMed PMID: 25806936; PubMed Central PMCID: PMC4393247.
6. Chowell G, Viboud C, Hyman JM, Simonsen L. The Western Africa ebola virus disease epidemic exhibits both global exponential and local polynomial growth rates. PLoS currents. 2015;7. Epub 2015/02/17. doi: 10.1371/currents.outbreaks.8b55f4bad99ac5c5db3663e916803261. PubMed PMID: 25685633; PubMed Central PMCID: PMC4322058.
7. Althaus CL. Estimating the reproduction number of Zaire ebolavirus (EBOV) during the 2014 outbreak in West Africa. PLOS Currents Outbreaks Edition 1 doi: 101371/currentsoutbreaks91afb5e0f279e7f29e7056095255b288. 2014.
8. Nishiura H, Chowell G. Early transmission dynamics of Ebola virus disease (EVD), West Africa, March to August 2014. Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin. 2014;19(36). PubMed PMID: 25232919.
9. Dietz P, Jambai A, Paweska JT, Yoti Z, Ksaizek TG. Epidemiology and risk factors for Ebola virus infection in Sierra Leone - May 23, 2014 -January 31, 2015. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America. 2015. doi: 10.1093/cid/civ568. PubMed PMID: 26179011.
10. Faye O, Boelle PY, Heleze E, Faye O, Loucoubar C, Magassouba N, et al. Chains of transmission and control of Ebola virus disease in Conakry, Guinea, in 2014: an observational study. The Lancet Infectious diseases. 2015;15(3):320-6. doi:

10.1016/S1473-3099(14)71075-8. PubMed PMID: 25619149; PubMed Central PMCID: PMC4373532.

11. Fasina F, Shittu A, Lazarus D, Tomori O, Simonsen L, Viboud C, et al. Transmission dynamics and control of Ebola virus disease outbreak in Nigeria, July to September 2014. Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin. 2014;19(40). PubMed PMID: 25323076.

12. Salathe M, Bengtsson L, Bodnar TJ, Brewer DD, Brownstein JS, Buckee C, et al. Digital epidemiology. PLoS computational biology. 2012;8(7):e1002616. doi: 10.1371/journal.pcbi.1002616. PubMed PMID: 22844241; PubMed Central PMCID: PMC3406005.

13. Brownstein JS, Freifeld CC, Madoff LC. Digital disease detection--harnessing the Web for public health surveillance. The New England journal of medicine. 2009;360(21):2153-5, 7. doi: 10.1056/NEJMp0900702. PubMed PMID: 19423867; PubMed Central PMCID: PMC2917042.

14. Salathe M, Khandelwal S. Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control. PLoS computational biology. 2011;7(10):e1002199. doi: 10.1371/journal.pcbi.1002199. PubMed PMID: 22022249; PubMed Central PMCID: PMC3192813.

15. Lazer D, Kennedy R, King G, Vespignani A. Big data. The parable of Google Flu: traps in big data analysis. Science. 2014;343(6176):1203-5. doi: 10.1126/science.1248506. PubMed PMID: 24626916.

16. Chowell G, Nishiura H. Transmission dynamics and control of Ebola virus disease (EVD): a review. BMC Med. 2014;12(1):196. doi: 10.1186/s12916-014-0196-0. PubMed PMID: 25300956.

17. Towers S, Patterson-Lomba O, Castillo-Chavez C. Temporal Variations in the Effective Reproduction Number of the 2014 West Africa Ebola Outbreak. PLOS Currents Outbreaks. 2014.

18. Fisman D, Khoo E, Tuite A. Early epidemic dynamics of the west african 2014 ebola outbreak: estimates derived with a simple two-parameter model. PLoS currents. 2014;6. doi: 10.1371/currents.outbreaks.89c0d3783f36958d96ebbae97348d571. PubMed PMID: 25642358; PubMed Central PMCID: PMC4169344.

19. Dowell SF, Mukunu R, Ksiazek TG, Khan AS, Rollin PE, Peters CJ. Transmission of Ebola hemorrhagic fever: a study of risk factors in family members, Kikwit, Democratic Republic of the Congo, 1995. Commission de Lutte contre les Epidemies a Kikwit. The Journal of infectious diseases. 1999;179 Suppl 1:S87-91. doi: 10.1086/514284. PubMed PMID: 9988169.

20. Breman JGea. The epidemiology of Ebola hemorrhagic fever in Zaire, 1976. Ebola virus haemorrhagic fever. 1978:103-24.

21. Francesconi P, Yoti Z, Declich S, Onek PA, Fabiani M, Olango J, et al. Ebola hemorrhagic fever transmission and risk factors of contacts, Uganda. Emerging infectious diseases. 2003;9(11):1430-7. doi: 10.3201/eid0911.030339. PubMed PMID: 14718087; PubMed Central PMCID: PMC3035551.

22. Wamala JF, Lukwago L, Malimbo M, Nguku P, Yoti Z, Musenero M, et al. Ebola hemorrhagic fever associated with novel virus strain, Uganda, 2007-2008. Emerging

infectious diseases. 2010;16(7):1087-92. doi: 10.3201/eid1607.091525. PubMed PMID: 20587179; PubMed Central PMCID: PMC3321896.
23. Shuaib F, Gunnala R, Musa EO, Mahoney FJ, Oguntimehin O, Nguku PM, et al. Ebola Virus Disease Outbreak - Nigeria, July-September 2014. MMWR Morb Mortal Wkly Rep. 2014;63(39):867-72. PubMed PMID: 25275332.
24. World Health Organization (WHO). Unprecedented number of medical staff infected with Ebola. Available online at :http://www.who.int/mediacentre/news/ebola/25-august-2014/en/ (25-August-2014) 2014.
25. Chowell G, Hengartner NW, Castillo-Chavez C, Fenimore PW, Hyman JM. The basic reproductive number of Ebola and the effects of public health measures: the cases of Congo and Uganda. J Theor Biol. 2004;229(1):119-26. Epub 2004/06/05. doi: 10.1016/j.jtbi.2004.03.006
S0022519304001092 [pii]. PubMed PMID: 15178190.
26. Legrand J, Grais RF, Boelle PY, Valleron AJ, Flahault A. Understanding the dynamics of Ebola epidemics. Epidemiology and infection. 2007;135(4):610-21. doi: 10.1017/S0950268806007217. PubMed PMID: 16999875; PubMed Central PMCID: PMC2870608.
27. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect of individual variation on disease emergence. Nature. 2005;438(7066):355-9. Epub 2005/11/18. doi: nature04153 [pii]
10.1038/nature04153. PubMed PMID: 16292310.
28. Salathe M, Jones JH. Dynamics and control of diseases in networks with community structure. PLoS computational biology. 2010;6(4):e1000736. doi: 10.1371/journal.pcbi.1000736. PubMed PMID: 20386735; PubMed Central PMCID: PMC2851561.
29. Team WHOER, Agua-Agum J, Ariyarajah A, Aylward B, Blake IM, Brennan R, et al. West African Ebola epidemic after one year--slowing but not yet under control. The New England journal of medicine. 2015;372(6):584-7. doi: 10.1056/NEJMc1414992. PubMed PMID: 25539446; PubMed Central PMCID: PMC4368109.
30. ProMED-mail. A global electronic reporting system for outbreaks of emerging infectious diseases and toxins. Available from: http://www.promedmail.org/ (accessed on July 22, 2015).
31. Brownstein JS, Freifeld CC. HealthMap: the development of automated real-time internet surveillance for epidemic intelligence. Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin. 2007;12(11):E071129 5. PubMed PMID: 18053570.
32. Chunara R, Andrews JR, Brownstein JS. Social and news media enable estimation of epidemiological patterns early in the 2010 Haitian cholera outbreak. The American journal of tropical medicine and hygiene. 2012;86(1):39-45. doi: 10.4269/ajtmh.2012.11-0597. PubMed PMID: 22232449; PubMed Central PMCID: PMC3247107.
33. Bahk CY, Scales DA, Mekaru SR, Brownstein JS, Freifeld CC. Comparing timeliness, content, and disease severity of formal and informal source outbreak

reporting. BMC infectious diseases. 2015;15(1):135. doi: 10.1186/s12879-015-0885-0. PubMed PMID: 25887692; PubMed Central PMCID: PMC4369067.
34. Majumder MS, Kluberg S, Santillana M, Mekaru S, Brownstein JS. 2014 ebola outbreak: media events track changes in observed reproductive number. PLoS currents. 2015;7. doi: 10.1371/currents.outbreaks.e6659013c1d7f11bdab6a20705d1e865. PubMed PMID: 25992303.

Figure legends

Figure 1. Epidemic curves of monthly numbers of Ebola virus disease (EVD) cases based on case reports of the World Health Organization in Guinea, Sierra Leone and Liberia and our sample of cases clusters, January 2014 – February 2015.

Figure 2. Temporal variation in the distribution of Ebola exposures through the family, hospitals and funerals in Guinea, Sierra Leone and Liberia, January 2014 – February 2015.

Figure 3. Temporal variation in the frequency of health-care worker cases among all cases in our sample based on case clusters retrieved from published literature including newspaper reports compared to the proportion of health care worker cases from all Ebola cases reported by the World Health Organization, January 2014 – February 2015.

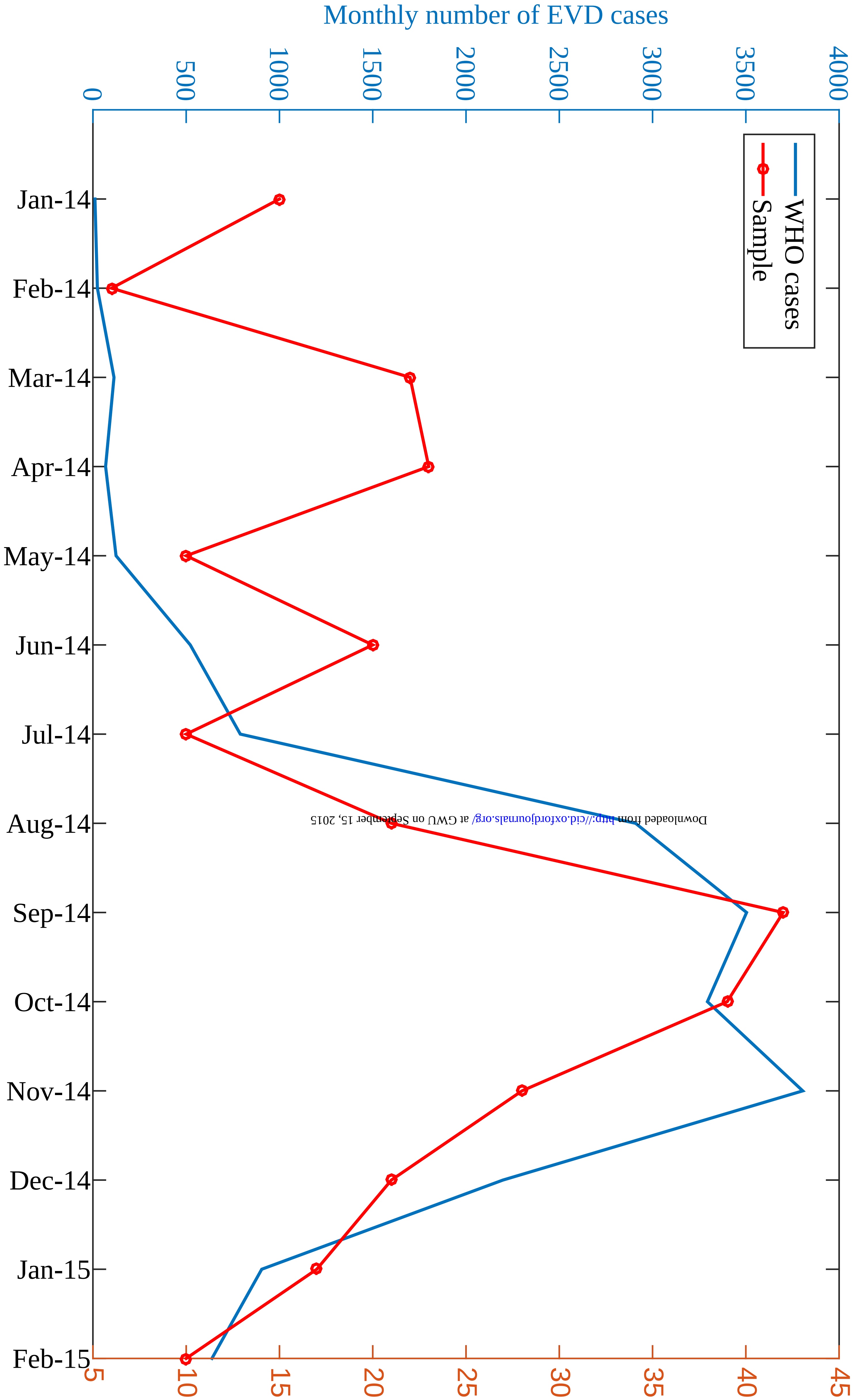
Figure 4. The distribution of the reproduction number estimates for index cases in our sample of clusters, January 2014 – February 2015.

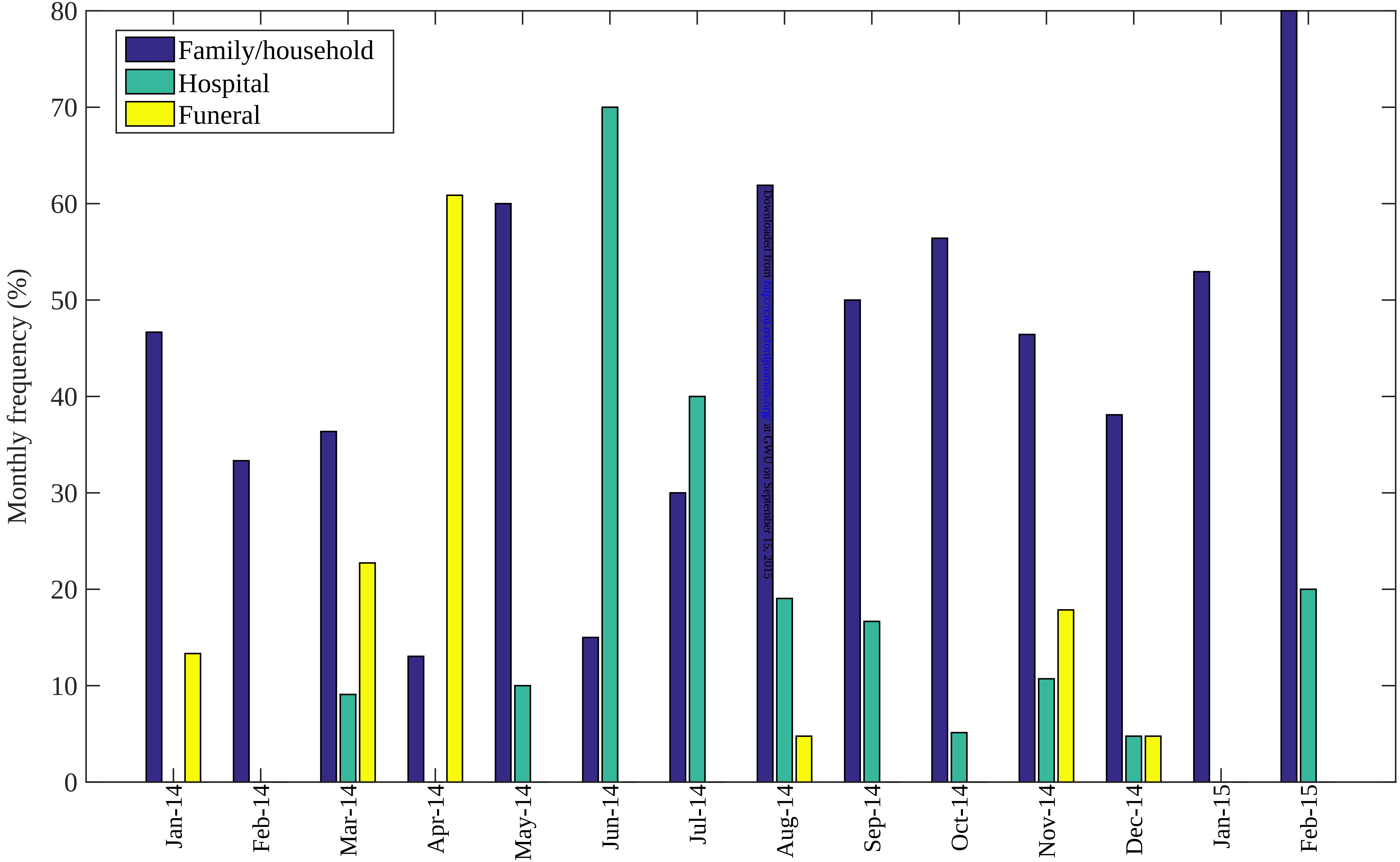
Figure 5. Temporal distribution of the reproduction number estimates available from clusters, January 2014-February 2015. The horizontal dashed line at R=1 is shown for reference.

Table 1. Ebola exposure patterns by country, gender, disease outcome and health-care worker status, December 2013 – February 2015. In the case of multiple exposures, then multiple exposure types were counted.

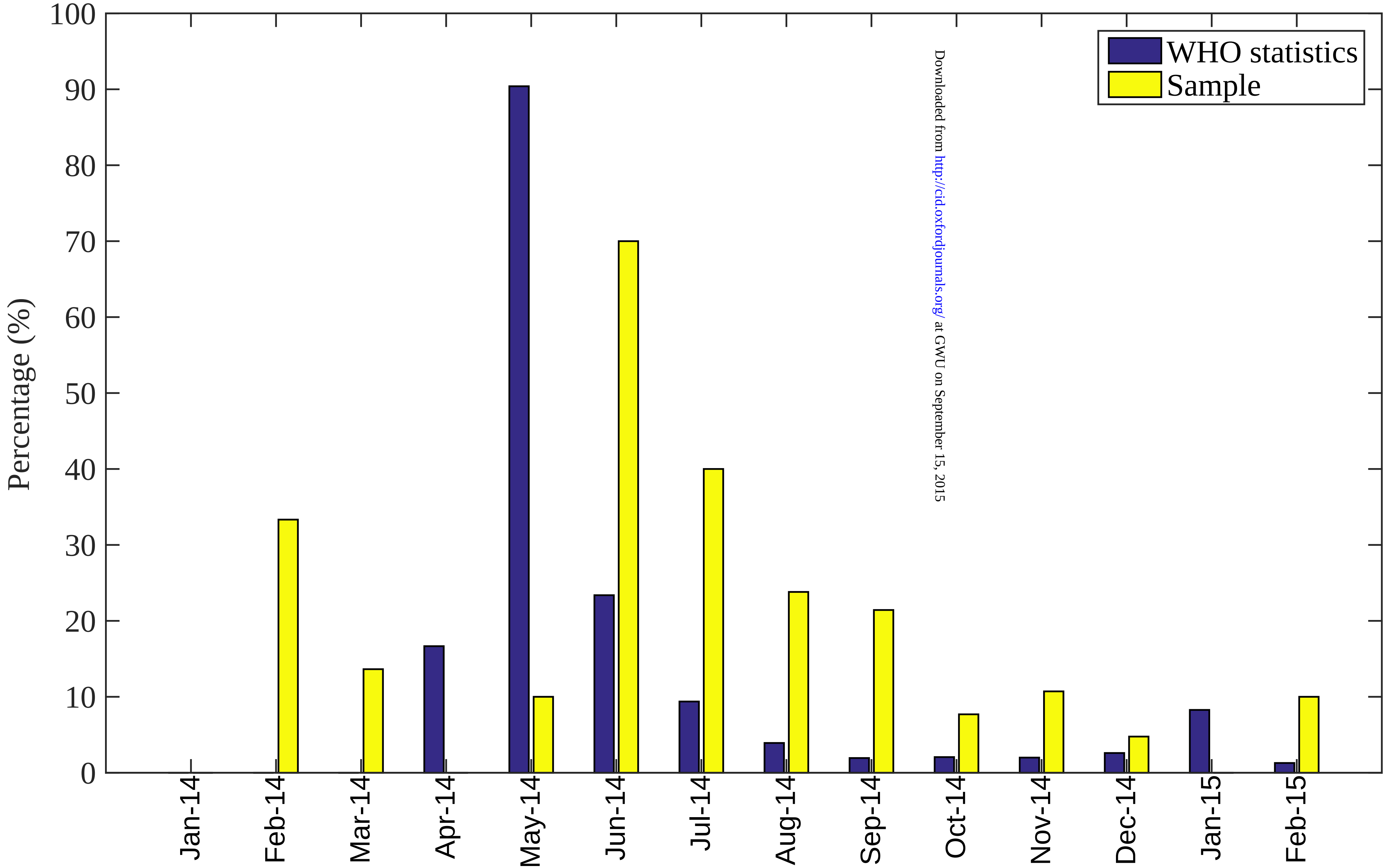
Exposure type	**All**	**Country**			**Gender**		**Disease outcome**		**Health-care worker**	
		Guinea	**Sierra Leone**	**Liberia**	**Male**	**Female**	**Died**	**Survived**	**Yes**	**No**
Family/ household	44.1%	33.8%	42.0%	50.0%	40.0%	56.6%	47.1%	45.0%	8.7%	80.2%
Hospital	14.0%	6.2%	11.1%	19.3%	18.1%	2.5%	15.5%	16.7%	84.8%	0%
Funeral	9.8%	12.3%	22.2%	1.4%	7.6%	3.3%	5.2%	5.0%	0%	6.7%
Care*	5.2%	9.2%	1.2%	5.7%	5.7%	4.9%	7.5%	1.7%	4.3%	73.6%
Sexual	1.7%	0%	4.9%	0.7%	0%	4.1%	0.6%	0%	0%	0%
Hazardous waste	1.4%	4.6%	0%	0.7%	1.0%	2.5%	2.3%	0%	2.2%	2.2%
Childbirth	0.3%	0%	0%	0.7%	0%	0.8%	0.6%	0%	0%	0.7%
Zoonotic	0.3%	1.5%	0%	0%	1%	0%	0.6%	0%	0%	0.7%
Unknown	23.1%	32.3%	18.5%	21.4%	26.7%	25.4%	20.7%	31.7%	0%	35.8%
Multiple exposures	5.2%	4.6%	2.5%	7.1%	2.9%	5.7%	5.7%	5.0%	4.3%	5.2%

* Care exposure indicates that a patient was infected after helping another EVD patient outside of the hospital setting, including through transportation to a health-care setting or attending to an Ebola case.

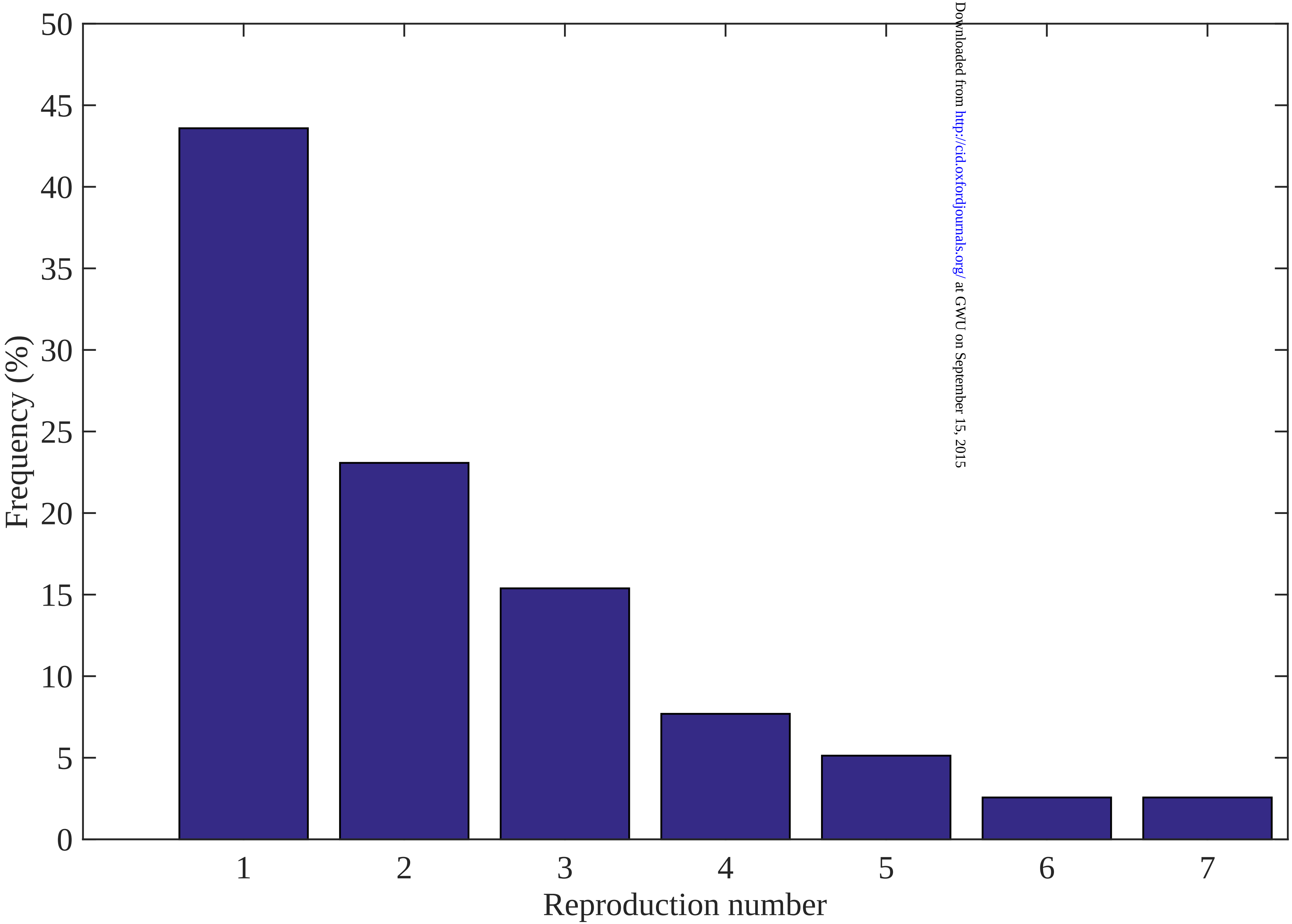

Monthly number of EVD cases
0
500
1000
1500
2000
2500
3000
3500
4000
WHO cases
Sample
Jan-14
Feb-14
Mar-14
Apr-14
May-14
Jun-14
Jul-14
Aug-14
Sep-14
Oct-14
Nov-14
Dec-14
Jan-15
Feb-15
5
10
15
20
25
30
35
40
45

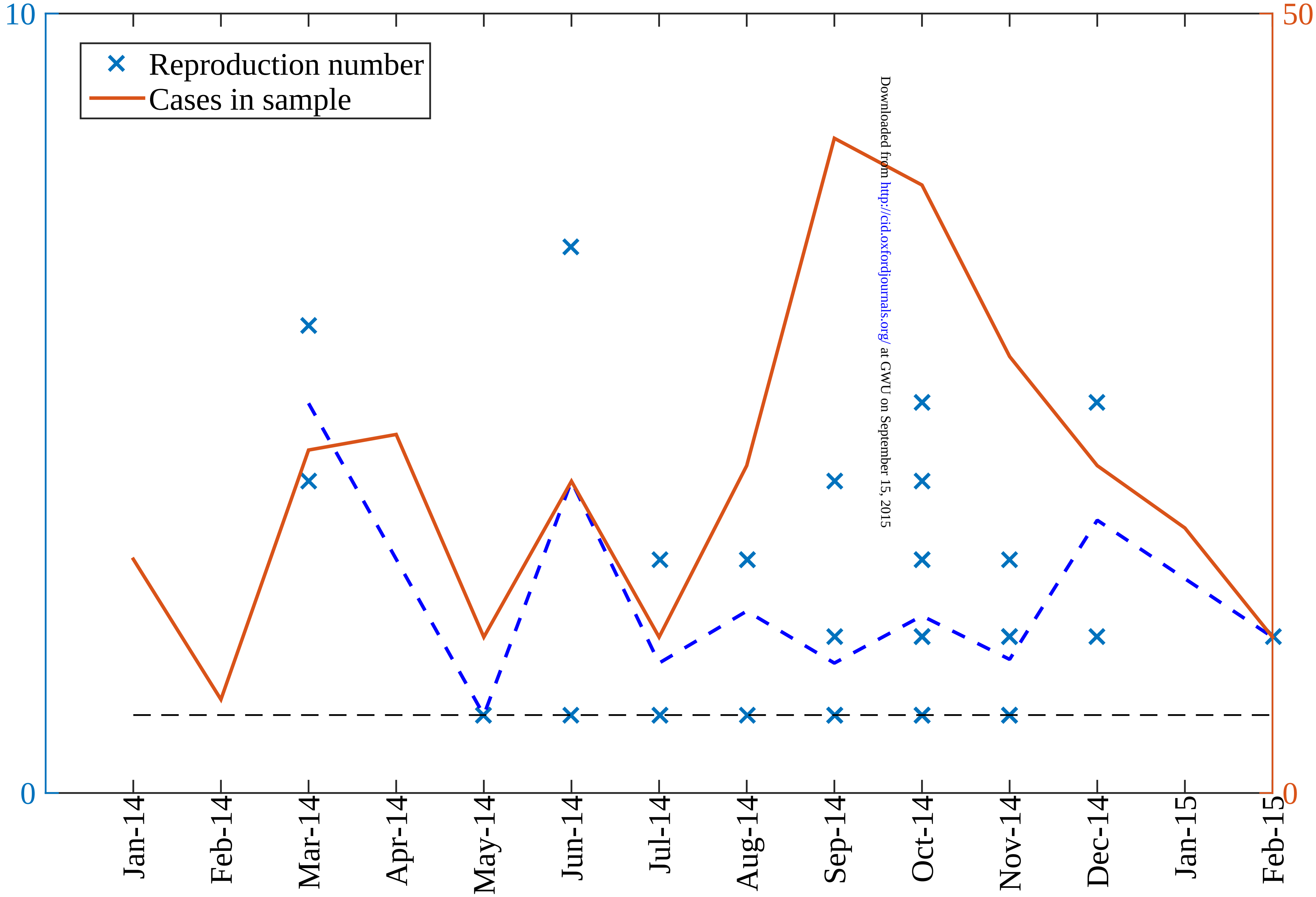


Family/household
Hospital
Funeral
Monthly frequency (%)
0
10
20
30
40
50
60
70
80
Jan-14
Feb-14
Mar-14
Apr-14
May-14
Jun-14
Jul-14
Aug-14
Sep-14
Oct-14
Nov-14
Dec-14
Jan-15
Feb-15



Percentage (%)
0
10
20
30
40
50
60
70
80
90
100
Jan-14
Feb-14
Mar-14
Apr-14
May-14
Jun-14
Jul-14
Aug-14
Sep-14
Oct-14
Nov-14
Dec-14
Jan-15
Feb-15
WHO statistics
Sample


Frequency (%)
0
5
10
15
20
25
30
35
40
45
50
1
2
3
4
5
6
7
Reproduction number


10
0
50
0
Reproduction number
Cases in sample
Jan-14
Feb-14
Mar-14
Apr-14
May-14
Jun-14
Jul-14
Aug-14
Sep-14
Oct-14
Nov-14
Dec-14
Jan-15
Feb-15