Supporting Information for:

Transmission Dynamics of Ebola Virus Disease and Intervention Effectiveness in Sierra Leone

Table of Contents

Supplementary Figures

Fig. S1. Weekly proportion of confirmed and suspected cases in Sierra Leone. The red and blue bars respectively represented the proportion of confirmed and suspected cases, respectively.

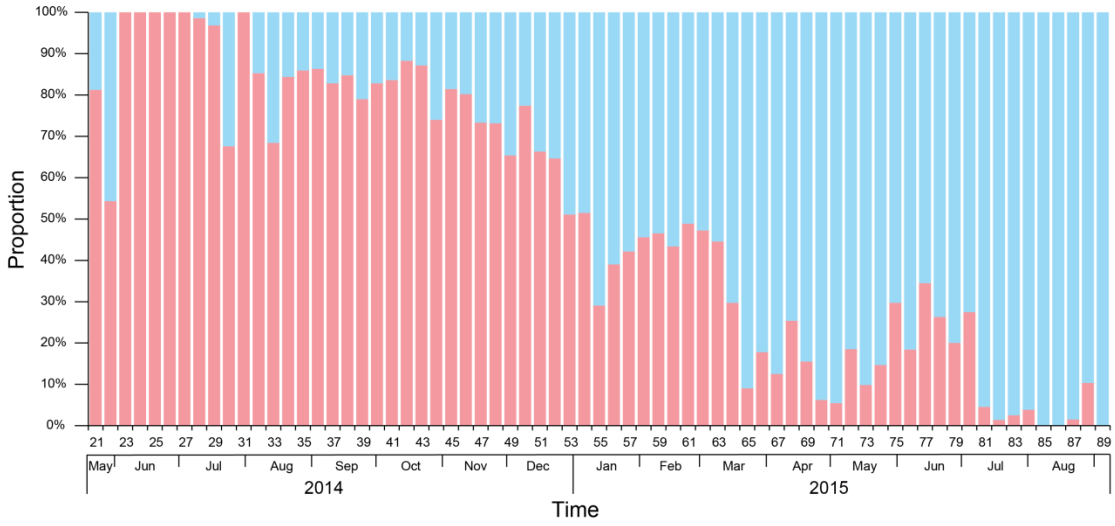


Fig. S2. The monthly number of confirmed cases in healthcare workers and the proportion of them among all confirmed cases in Sierra Leone.

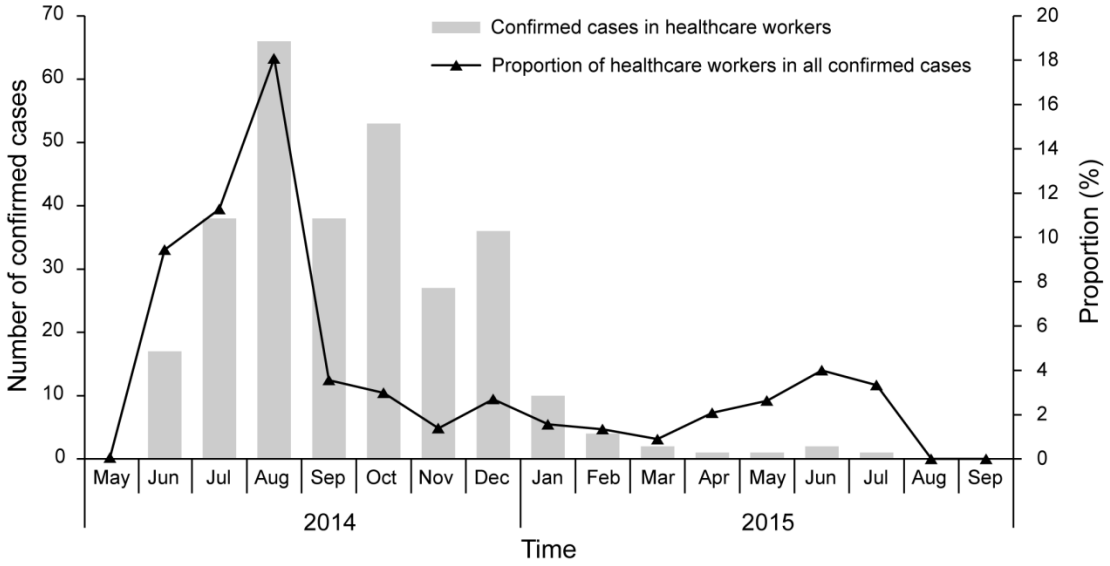


Fig. S3. Map series of monthly incidence rate of EVD at each chiefdom of Sierra Leone from May 2014 to September 2015.

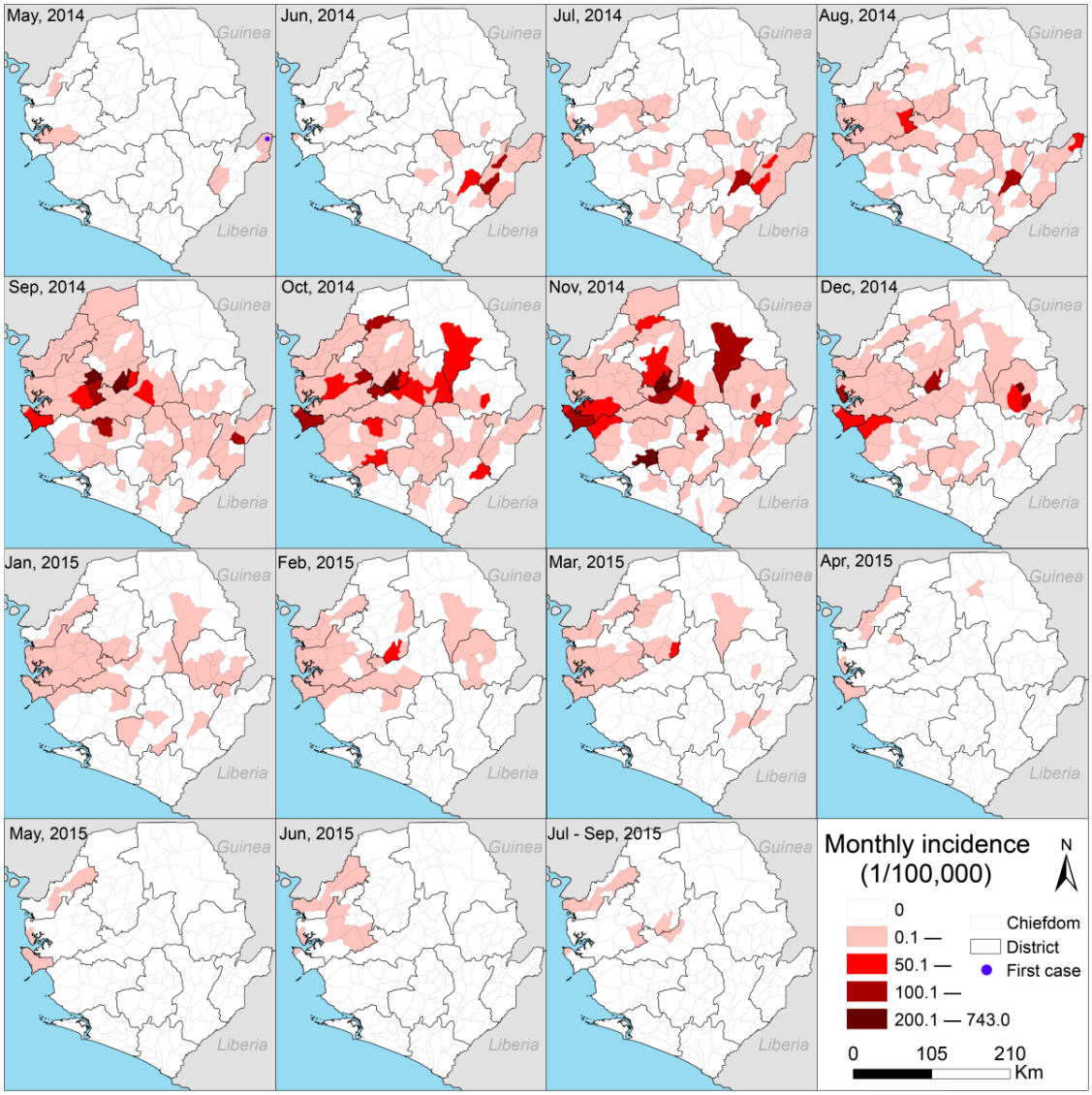


Fig. S4. The heat map of the weekly number of confirmed cases and six categories of epidemic patterns among the 114 affected chiefdoms. The left and right columns represent 54 and 60 affected chiefdoms in western and eastern Sierra Leone, linked to six epidemic patterns in the middle column. The chiefdoms are ordered first by the latitudes (from north to south) of the centroids of the districts and then by the latitudes of the centroids of the chiefdoms, and are numbered accordingly. The identification number of each chiefdom is listed in Table S11.

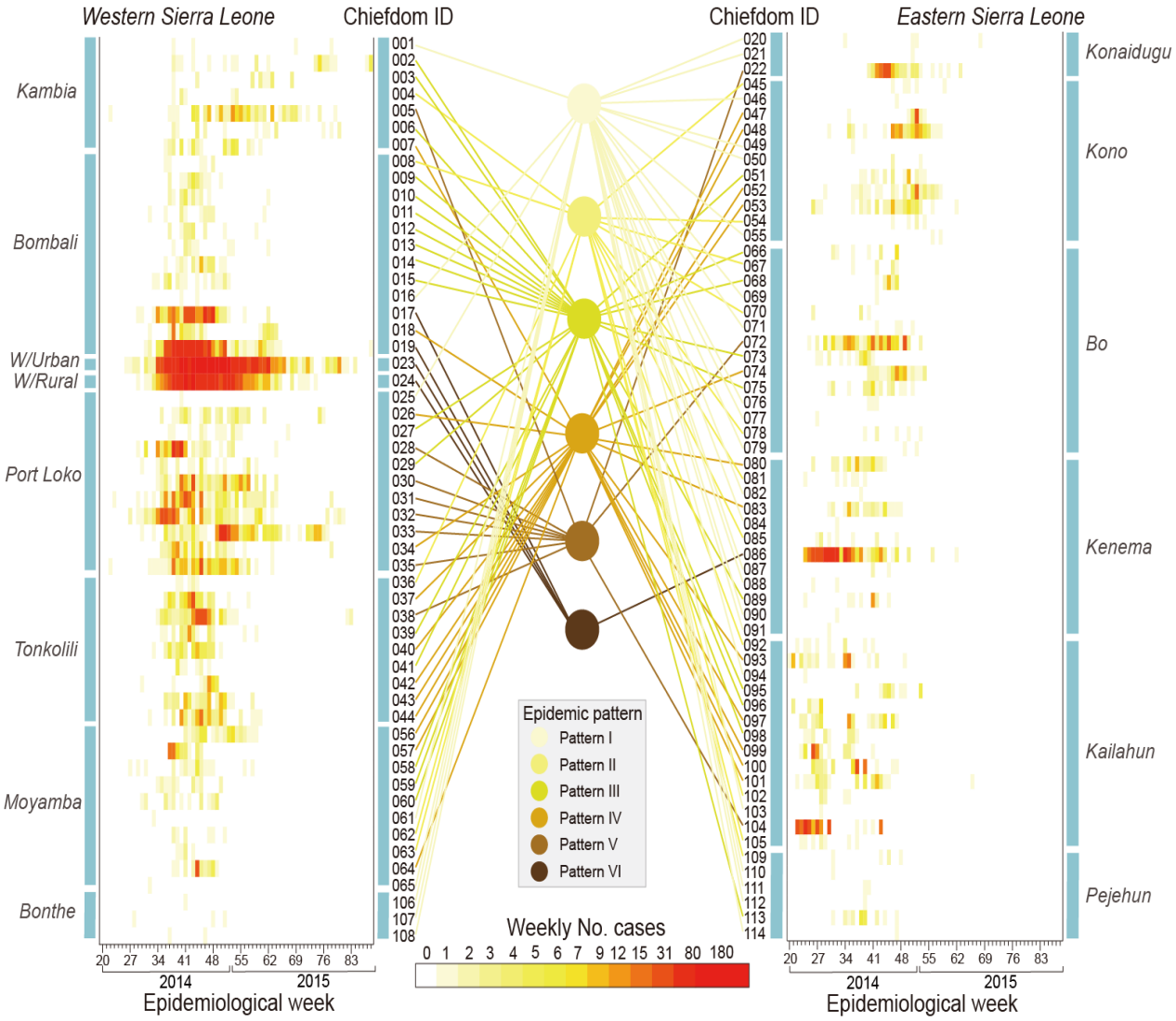


Fig. S5. Model-fitted versus observed weekly numbers of Ebola virus disease cases in Sierra Leone. The fitted values are based on a Poisson transmission model at the chiefdom and week level.

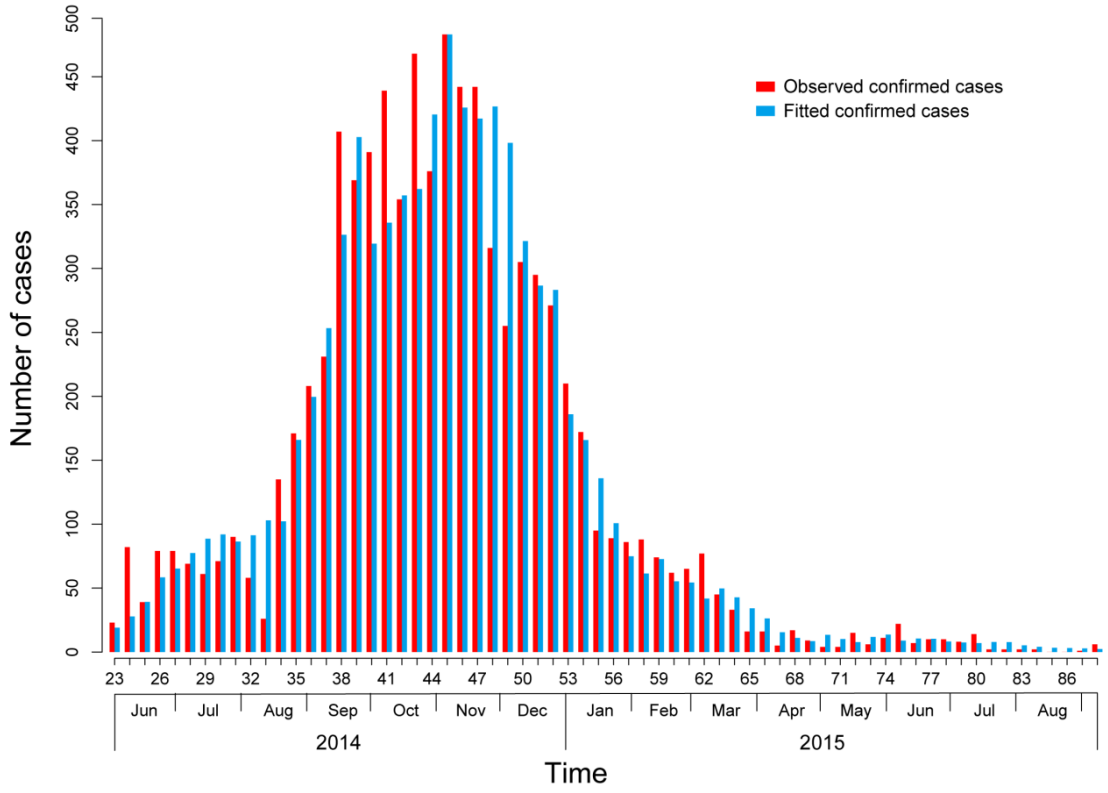


Fig. S6. Association between estimated risk ratios (RR) and socio-environmental factors adjusting for relative humidity. The RRs are estimated based on a Poisson transmission model at the chiefdom and week level for confirmed EVD cases in Sierra Leone from 2 June 2014 to 6 September 2015 (week 23–88). Estimated RR curves are in red. Uncertainty is shown by estimated curves (in gray) based on 50 (randomly selected from 1000) bootstrap samples of the data set. The histograms represent the distribution of the socio-environmental factors. The cross points of horizontal and vertical dash lines indicate the mean values of the socio-environmental factors at which the RR is 1.

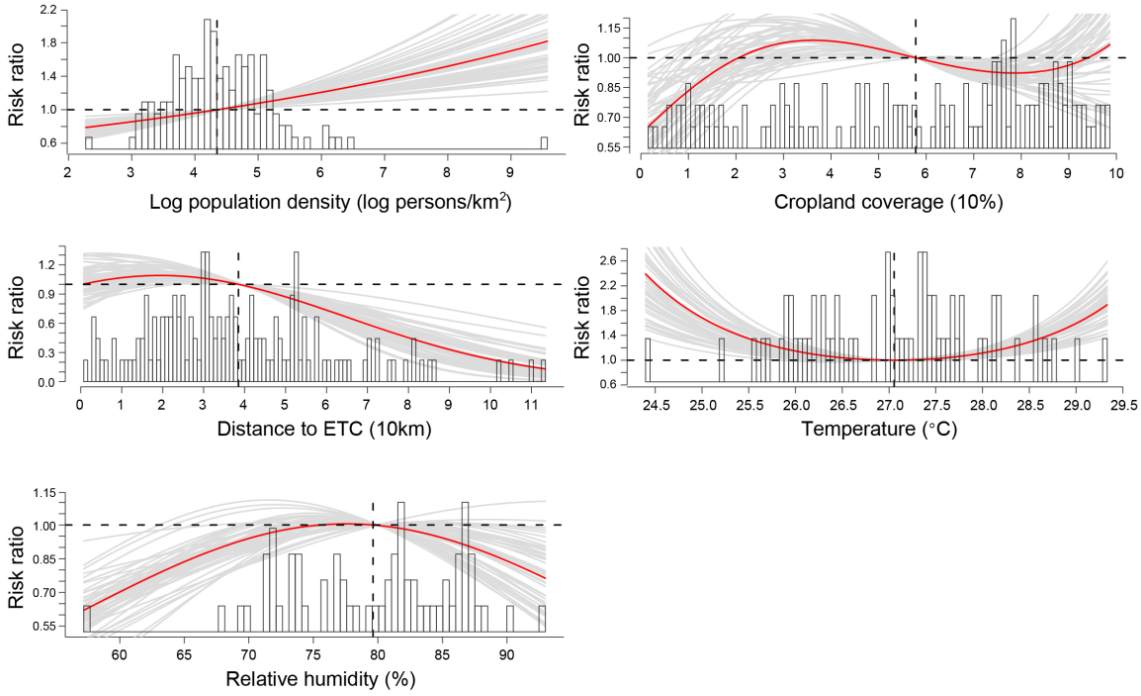


Fig. S7. Association between estimated risk ratios (RR) and socio-environmental factors using both confirmed and suspected cases for sensitivity analysis. The RRs are estimated based on a Poisson transmission at the chiefdom and week level for confirmed and suspected EVD cases in Sierra Leone from 2 June 2014 to 6 September 2015 (week 23–88). The model structure is the same as the final model based on confirmed cases in the primary analysis. Estimated RR curves are in red. Uncertainty is shown by estimated curves (in gray) based on 50 (randomly selected from 1000) bootstrap samples of the data set. The histograms represent the distribution of the socio-environmental factors. The cross points of horizontal and vertical dash lines indicate the mean values of the socio-environmental factors at which the RR is 1.

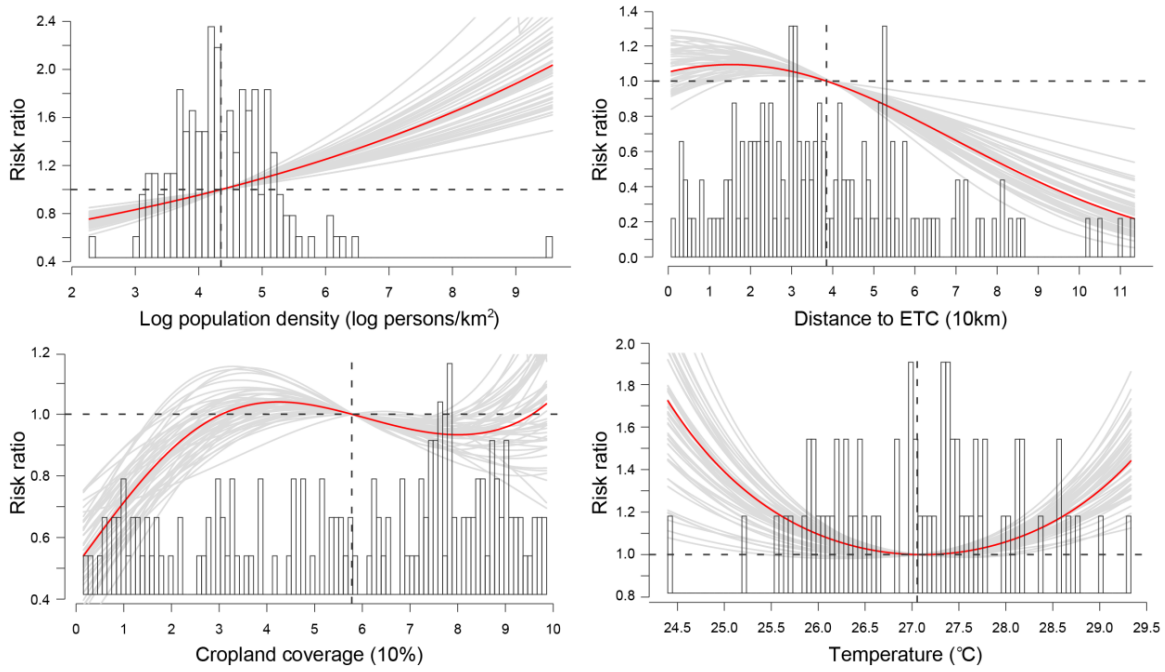


Fig. S8. Estimates for the effects of gender (A and C, odds ratio between males and females) and age group (B and D, odds ratio between children and adults) on infectiousness during household transmission, stratified by mean durations of the incubation and infectious periods. The estimation was also stratified by definition of infection: confirmed EVD cases plus case-reported source contacts (A and B) and clinical EVD cases (with symptoms but not necessarily confirmed) plus case-reported source contacts (C and D).

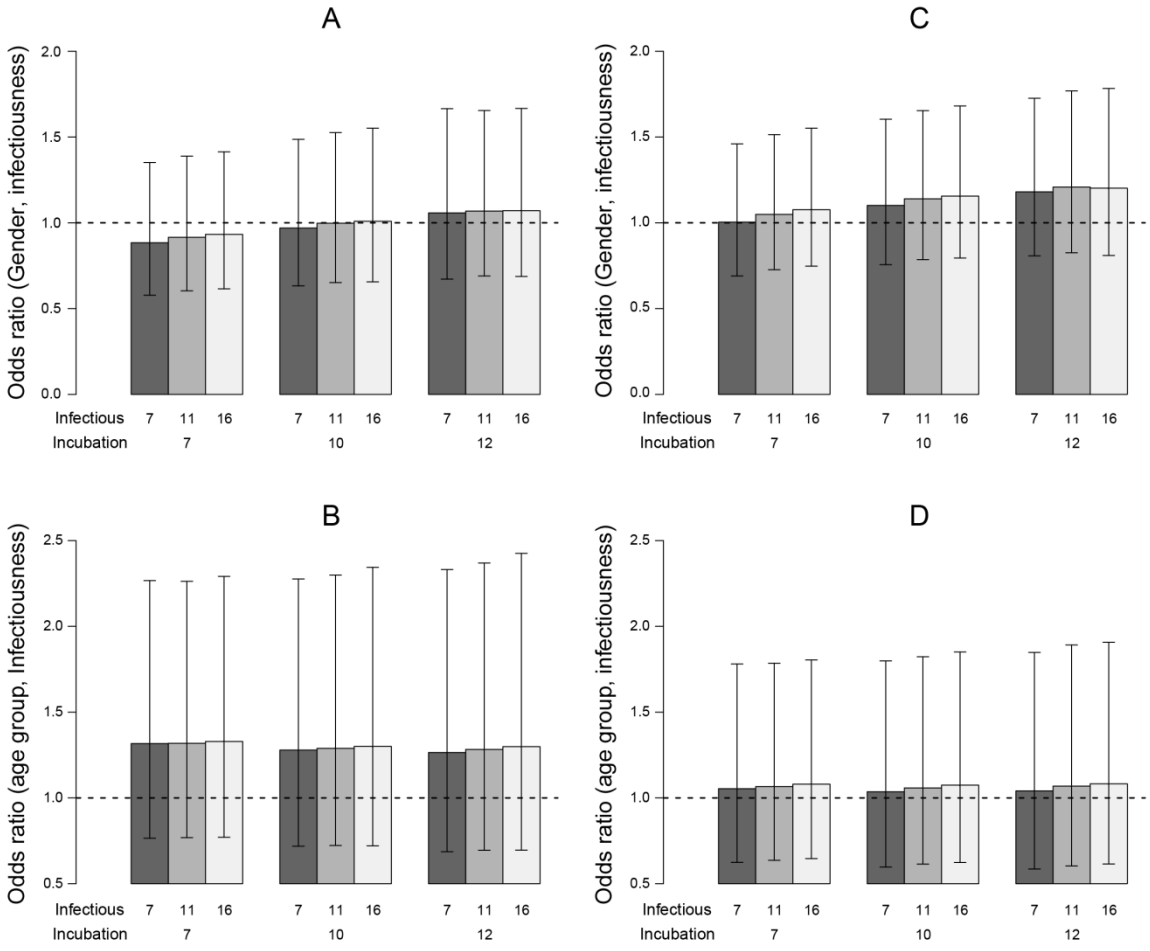
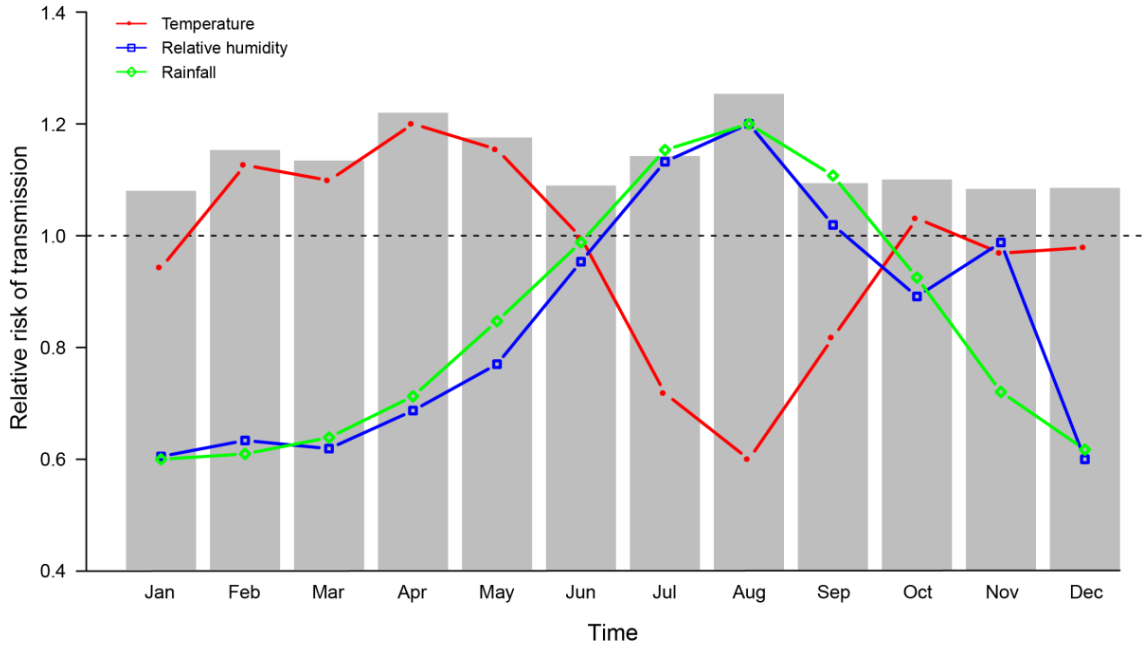


Fig. S9. Temporal trend of Poisson-model-fitted relative risk of EVD transmission and climatic variables in Sierra Leone. Relative risks (gray bars) were calculated at monthly average temperature (in red) and relative humidity (in blue) based on the weather station in Freetown (January–March 2015 for months 1–3 and April–December 2014 for months 4–12), ignoring the two-week lag. Monthly average rainfall (in green) is based on historic data from 1901 to 2009 provided by the World Bank (http://data.worldbank.org/country/sierra-leone#cp_cc).



Supplementary Tables

Table S1. Estimated hazard ratios (HR) of potential risk factors of EVD invasion based on confirmed cases in Sierra Leone. The outcome of the model is the time (days) to be the first confirmed cases for each chiefdom in Sierra Leone. Invasion times for unaffected chiefdoms were considered as right-censored**.**

Variables (Unit)[†]	Median times to invasion (IQR)	Univariate Analysis		Multivariate Analysis	
		Crude HR (95% Cl)	*P* value	Adjusted HR (95% Cl)	*P* value
Intersected by primary road		1.60 (1.08, 2.36)	0.019	1.50 (1.00, 2.23)	0.048
No	124 (82, 476)				
Yes	66 (28, 166)				
Intersected by secondary road		1.75 (1.18, 2.60)	0.005	1.65 (1.11, 2.46)	0.014
No	135 (96, 476)				
Yes	100 (52, 143)				
Intersected by railway		2.14 (1.17, 3.93)	0.014	NS (excluded)	
No	120 (66, 476)				
Yes	71 (55, 108)				
Distance to nearest hospital (10 km)		0.69 (0.62, 0.77)	<0.001	0.71 (0.64, 0.79)	<0.001
< 2.20	68 (42, 113)				
2.21–3.90	111 (48, 144)				
≥ 3.91	476 (123, 476)				
Distance to nearest ETC (10 km)		0.85 (0.78, 0.92)	<0.001	NS (excluded)	
< 2.50	95 (63, 151)				
2.51–4.50	112 (56, 141)				
≥ 4.51	134 (81, 476)				
Population density (100 person per		1.02 (1.004, 1.03)	0.010	NS (excluded)	
< 0.51	212 (131, 476)				

0.52–1.01	115 (75, 152)			
≥ 1.01	66 (36, 119)			
Coverage of built-up (1%)		1.23 (1.06, 1.43)	0.006	NS (excluded)
< 0.004	476 (122, 476)			
0.004–0.030	110 (79, 136)			
≥ 0.031	63 (27, 124)			
Coverage of croplands (10%)		1.002 (0.98, 1.02)	0.889	
< 4.01	188 (112, 476)			
4.01–7.60	112 (57, 158)			
≥ 7.61	88 (43, 134)			
Coverage of forest (10%)		0.86 (0.79, 0.93)	<0.001	NS (excluded)
< 1.61	94 (47, 141)			
1.61–4.50	103 (61, 149)			
≥ 4.51	212 (117, 476)			
Coverage of shrub (10%)		0.85 (0.72, 1.01)	0.059	NS (excluded)
< 0.21	115 (49, 192)			
0.21–0.70	108 (48, 190)			
≥ 0.71	122 (80, 476)			

Abbreviations: CI, confidence interval; HR, hazard ratio; IQR, interquartile range; ETC, Ebola treatment center; NS, not significant.

[†] Each continuous variable was also categorized and mean time to invasion is given for each category.

Table S2. Estimated hazard ratios (HR) of potential risk factors of EVD invasion based on both confirmed and suspected cases.

The outcome of the model is the time (days) to the first confirmed or suspected case for each chiefdom in Sierra Leone. Invasion times for unaffected chiefdoms were considered as right-censored.

Variables (Unit)[†]	Median times to invasion (IQR)	Univariate Analysis		Multivariate Analysis	
		Crude HR (95% Cl)	*P* value	Adjusted HR (95% Cl)	*P* value
Intersected by primary road		1.45 (1.01, 2.07)	0.046	1.61 (1.11, 2.33)	0.013
No	123 (81, 192)				
Yes	65 (14, 136)				
Intersected by secondary road		1.69 (1.19, 2.40)	0.003	1.80 (1.26, 2.58)	0.001
No	131 (94, 308)				
Yes	95 (44, 137)				
Intersected by railroad		2.08 (1.16, 3.70)	0.013	NS (excluded)	
No	116 (65, 189)				
Yes	63 (32, 100)				
Distance to nearest hospital (10 km)		0.72 (0.66, 0.79)	<0.001	0.73 (0.67, 0.79)	<0.001
< 2.20	65 (13, 103)				
2.21–3.90	110 (42, 139)				
≥ 3.91	208 (123, 334)				
Distance to nearest ETC (10 km)		0.85 (0.79, 0.92)	<0.001	NS (excluded)	
< 2.50	94 (58, 138)				
2.51–4.50	106 (48, 138)				
≥ 4.51	134 (81, 320)				
Population density (100 person per		1.04 (1.02, 1.06)	<0.001	1.03 (1.01, 1.05)	0.004
< 0.51	182 (131, 335)				
0.52–1.01	108 (75, 151)				
≥ 1.01	63 (12, 111)				

Coverage of built-up (1%)		1.63 (1.27, 2.10)	<0.001	NS (excluded)
< 0.004	189 (122, 332)			
0.004–0.030	108 (79, 136)			
≥ 0.031	48 (11, 103)			
Coverage of croplands (10%)		1.01 (0.99, 1.03)	0.196	NS (excluded)
< 4.01	183 (109, 327)			
4.01–7.60	103 (53, 141)			
≥ 7.61	88 (40, 130)			
Coverage of forests (10%)		0.86 (0.80, 0.92)	<0.001	NS (excluded)
< 1.61	92 (46, 136)			
1.61–4.50	101 (51, 146)			
≥ 4.51	188 (117, 401)			
Coverage of shrub (10%)		0.90 (0.78, 1.03)	0.123	NS (excluded)
< 0.21	92 (43, 157)			
0.21–0.70	103 (44, 180)			
≥ 0.71	122 (80, 286)			

Abbreviations: CI, confidence interval; HR, hazard ratio; IQR, interquartile range; ETC, Ebola treatment center; NS, not significant.

† Each continuous variable was also categorized and mean time to invasion is given for each category.

Table S3. Summary of six categories of epidemic patterns in chiefdoms with confirmed cases.

Epidemic pattern	No. of Chiefdoms	Median No. of cases (min, max)	Average population size	Average density of population (persons/km^2)
I	31	1 (1, 4)	27045	81.58
II	16	6 (5, 10)	25111	47.68
III	27	20 (11, 34)	32980	83.19
IV	24	49.5 (29, 86)	44149	96.68
V	11	152 (106, 255)	83464	130.52
VI	5	581 (262, 2274)	309937	782.97

Table S4. Observed and Poisson-model-fitted numbers of EVD cases by districts of Sierra Leone from 2 June 2014 to 6 September 2015 (week 23–88).

Districts	Bo	Bombali	Bonthe	Kailahun	Kambia	Kenema	Koinadugu
Observed	315	1049	6	488	240	497	111
Fitted	288	1089	16	509	175	557	82

Districts	Kono	Moyamba	Port Loko	Pujehun	Tonkolili	Western Rural	Western Urban
Observed	263	210	1201	31	489	1146	2274
Fitted	262	198	1162	51	532	1128	2272

Table S5. Parameter estimates for the Poisson transmission model at the chiefdom and week level for the confirmed EVD cases in Sierra Leone.

Transmission rates and θ are shown in natural scale. Risk ratios (RR) are shown for categorical variables: intervention phase and ethnic group. Regression coefficients for the linear equation of the logarithm of transmission rate are shown for continuous variables: temperature, distance to ETC, cropland coverage, and population density. Confidence intervals (CI) are based on 1000 bootstrap samples.

Variables	Level	Estimate (95% CI)
γ_0		0.035 (0.022, 0.050)
γ_1		0.84 (0.64, 1.07)
θ		0.027 (0.018, 0.037)
Intervention phase	≤Sep. 28, 2014	-
I	≥Sep. 29, 2014	0.57 (0.48, 0.70)
II	≥Dec. 29, 2014	0.35 (0.29, 0.43)
Ethinic group	Mende	-
	Temne	1.43 (1.17, 1.76)
	Limba	1.47 (1.01, 2.07)
	Kono	1.47 (1.05, 1.99)
	Other	1.40 (1.07, 1.79)
Temperature	x	0.040 (-0.031, 0.10)
	x^2	0.10 (0.064, 0.15)
Distance to ETC	x	-0.21 (-0.37, -0.064)
	x^2	-0.13 (-0.25, -0.021)
Cropland coverage	x	-0.17 (-0.32, -0.027)
	x^2	0.0065 (-0.14, 0.16)
	x^3	0.11 (0.0027, 0.23)
Log population density	x	0.097 (0.034, 0.17)

Table S6. Parameter estimates for the Poisson transmission model adjusted for relative humidity at the chiefdom and week level for confirmed EVD cases in Sierra Leone. Transmission rates and θ are shown in natural scale. Risk ratios (RR) are shown for categorical variables: intervention phase and ethnic group. Regression coefficients for the linear equation of the logarithm of transmission rate are shown for continuous variables. Confidence intervals (CI) are based on 1000 bootstrap samples.

Variables	Level	Estimate (95% CI)
γ_0		0.035 (0.021, 0.053)
γ_1		0.86 (0.65, 1.12)
θ		0.027 (0.018, 0.037)
Intervention phase	≤Sep. 28, 2014	-
I	≥Sep. 29, 2014	0.57 (0.48, 0.70)
II	≥Dec. 29, 2014	0.37 (0.28, 0.47)
Ethinic group	Mende	-
	Temne	1.42 (1.17, 1.75)
	Limba	1.46 (0.99, 2.05)
	Kono	1.47 (1.06, 2.03)
	Other	1.40 (1.07, 1.79)
Temperature	x	-0.00042 (-0.090, 0.089)
	x^2	0.12 (0.072, 0.17)
Relative humidity	x	-0.031 (-0.15, 0.093)
	x^2	-0.051 (-0.10, -0.0067)
Distance to ETC	x	-0.21 (-0.37, -0.062)
	x^2	-0.12 (-0.26, -0.021)
Cropland coverage	x	-0.17 (-0.32, -0.024)
	x^2	0.0040 (-0.15, 0.15)
	x^3	0.10 (-0.000038, 0.23)
Log population density	x	0.097 (0.034, 0.17)

Table S7. Parameter estimates for the Poisson transmission model at the chiefdom and week level for confirmed and suspected EVD cases in Sierra Leone. The model structure is the same as the final model based on confirmed cases in the primary analysis. Transmission rates and θ are shown in natural scale. Risk ratios (RR) are shown for categorical variables: intervention phase and ethnic group. Regression coefficients for the linear equation of the logarithm of transmission rate are shown for continuous variables. Confidence intervals (CI) are based on 1000 bootstrap samples.

Variables	Level	Estimate (95% CI)
γ_0		0.12 (0.089, 0.15)
γ_1		0.82 (0.65, 1.01)
θ		0.021 (0.013, 0.029)
Intervention phase	≤Sep. 28, 2014	-
I	≥Sep. 29, 2014	0.69 (0.59, 0.83)
II	≥Dec. 29, 2014	0.53 (0.45, 0.63)
Ethinic group	Mende	-
	Temne	1.29 (1.11, 1.51)
	Limba	1.45 (1.11, 1.85)
	Kono	1.16 (0.88, 1.50)
	Other	1.23 (1.03, 1.47)
Temperature	x	-0.0081 (-0.063, 0.040)
	x^2	0.072 (0.037, 0.11)
Distance to ETC	x	-0.18 (-0.27, -0.084)
	x^2	-0.087 (-0.16, -0.023)
Cropland coverage	x	-0.12 (-0.24, -0.0084)
	x^2	-0.035 (-0.13, 0.075)
	x^3	0.096 (0.023, 0.18)
Log population density	x	0.12 (0.068, 0.17)

Table S8. Timeline of rolling out specific interventions by calendar week in Sierra Leone. Data before October 20, 2014 are not available. Data were extracted from weekly situation reports jointly provided by UNMEER and Sierra Leone National Ebola Response Center.

Week		ETC		CCC		No. of Burial Teams	No. of Labs
Start	Stop	No.	Beds	No.	Beds		
2014/10/20	2014/10/26	4	252	0	0	46	6
2014/10/27	2014/11/2	4	288	2	8	48	5
2014/11/3	2014/11/9	9	317	3	28	70	6
2014/11/10	2014/11/16	9	459	15	168	82	6
2014/11/17	2014/11/23	11	500	19	NA	82	5
2014/11/24	2014/11/30	12	550	22	236	95	5
2014/12/1	2014/12/7	12	550	22	NA	101	5
2014/12/8	2014/12/14	13	675	26	299	101	9
2014/12/15	2014/12/21	17	839	26	300	102	9
2014/12/22	2014/12/28	19	896	30	339	107	11
2014/12/29	2015/1/4	21	1096	35	371	125	11
2015/1/5	2015/1/11	23	1174	49	485	125	13
2015/1/12	2015/1/18	23	1174	52	528	125	13
2015/1/19	2015/1/25	23	1174	53	536	125	13
2015/1/26	2015/2/1	23	1174	53	536	125	13
2015/2/2	2015/2/8	23	1174	53	536	125	13
2015/2/9	2015/2/15	23	1174	53	536	125	13
2015/2/23	2015/3/1	18	631	48	470	125	16
2015/3/9	2015/3/15	19	584	34	314	125	13
2015/3/16	2015/3/29	18	574	34	314	122	13
2015/3/30	2015/4/12	16	500	20	197	102	13

Table S9. Distributions of household size in Western Area and in the whole country of Sierra Leone from the 2004 census.

	Household Size									
	1	2	3	4	5	6	7	8	9	≥10
Western Area	12944	15818	17293	17224	15516	12657	10034	7708	5768	19176
Whole Nation	47259	73127	99510	110317	105567	89477	70779	55789	41460	126563

Table S10. Fitted distributions of household size in Western Area and in the whole country of Sierra Leone based on the 2004 census.

	Household Size									
	1	2	3	4	5	6	7	8	9	10—15
Western Area	0.096	0.117	0.129	0.128	0.116	0.094	0.075	0.057	0.043	0.033, 0.026, 0.020, 0.015, 0.012, 0.009
Whole Nation	0.058	0.089	0.121	0.135	0.129	0.109	0.086	0.068	0.051	0.037, 0.029, 0.022, 0.017, 0.013, 0.009

Table S11. Identification number for each chiefdom affected by EVD. The chiefdom ID is used in Fig. 3B and S4.

District	Chiefdom	ID	District	Chiefdom	ID
Kambia	Bramaia	001	Port Loko	Dibia	027
Kambia	Tonko Limba	002	Port Loko	Buya Romende	028
Kambia	Gbinle-Dixing	003	Port Loko	Tms	029
Kambia	Masungbala	004	Port Loko	Lokomasam	030
Kambia	Magbema	005	Port Loko	Maforki	031
Kambia	Samu	006	Port Loko	Marampa	032
Kambia	Mambolo	007	Port Loko	Kaffu Bullom	033
Bombali	Tambakka	008	Port Loko	Masimera	034
Bombali	Sella Limba	009	Port Loko	Koya	035
Bombali	Sanda Loko	010	Tonkolili	Kafe Simiria	036
Bombali	Gbanti-Kamaranka	011	Tonkolili	Konike Sanda	037
Bombali	Biriwa	012	Tonkolili	Kholifa Rowalla	038
Bombali	Sanda Tendaren	013	Tonkolili	Malal Mara	039
Bombali	Ngowahun	014	Tonkolili	Tane	040
Bombali	Safroko Limba	015	Tonkolili	Konike Barina	041
Bombali	Libeisaygahun	016	Tonkolili	Kholifa Mabang	042
Bombali	Makari Gbanti	017	Tonkolili	Gbonkolenken	043
Bombali	Paki Masabong	018	Tonkolili	Yoni	044
Bombali	Bombali Sebora	019	Kono	Sandor	045
Koinadugu	Wara Wara Yagala	020	Kono	Lei	046
Koinadugu	Kasonko	021	Kono	Kamara	047
Koinadugu	Neini	022	Kono	Gbense	048
Western Urban	W/Urban	023	Kono	Gbane Kandor	049
Western Rural	W/Rural	024	Kono	Soa	050
Port Loko	Sanda Magbolontor	025	Kono	Nimiyama	051
Port Loko	Bkm	026	Kono	Nimikoro	052
Kono	Tankoro	053	Kenema	Simbaru	084
Kono	Gbane	054	Kenema	Kandu Leppiama	085
Kono	Gorama Kono	055	Kenema	Nongowa	086

Moyamba	Ribbi	056	Kenema	Small Bo	087
Moyamba	Fakunya	057	Kenema	Niawa	088
Moyamba	Kori	058	Kenema	Gaura	089
Moyamba	Bumpeh	059	Kenema	Koya	090
Moyamba	Kaiyamba	060	Kenema	Tunkia	091
Moyamba	Kowa	061	Kailahun	Kissi Kama	092
Moyamba	Kargboro	062	Kailahun	Kissi Teng	093
Moyamba	Bagruwa	063	Kailahun	Penguia	094
Moyamba	Banta Gbangbatote	064	Kailahun	Yawei	095
Moyamba	Timidale	065	Kailahun	Kissi Tonge	096
Bo	Valunia	066	Kailahun	Luawa	097
Bo	Komboya	067	Kailahun	Peje West	098
Bo	Niawa Lenga	068	Kailahun	Peje Bongre	099
Bo	Selenga	069	Kailahun	Upper Bambara	100
Bo	Badjia	070	Kailahun	Njaluahun	101
Bo	Gbo	071	Kailahun	Mandu	102
Bo	Kakua	072	Kailahun	Dea	103
Bo	Baoma	073	Kailahun	Jawei	104
Bo	Bumpe Ngawo	074	Kailahun	Malema	105
Bo	Tikonko	075	Bonthe	Imperi	106
Bo	Jaiama Bongor	076	Bonthe	Jong	107
Bo	Lugbu	077	Bonthe	Sogbini	108
Bo	Wunde	078	Pujehun	Malen	109
Bo	Bagbo	079	Pujehun	Barri	110
Kenema	Gorama Mende	080	Pujehun	Kpanga Kabonde	111
Kenema	Wandor	081	Pujehun	Panga Krim	112
Kenema	Dodo	082	Pujehun	Makpele	113
Kenema	Lower Bambara	083	Pujehun	Kpaka	114

Supplementary Materials and Methods

Sections

1. Ethical consideration

This work was conducted as part of the surveillance and public health response to contain the EVD outbreak in Sierra Leone, and informed consent was not obtained. The activities were coordinated by the Emergency Operations Center (EOC) jointly established by the Sierra Leone Ministry of Health and Sanitation and the World Health Organization (WHO). All information regarding individual persons has been anonymized in the report.

2. Data Sources

All the individuals, whose blood or nasopharyngeal swab samples were collected and sent to the laboratories appointed by the Sierra Leone Ministry of Health and Sanitation (SLMHS) for testing Ebola virus (EBOV), were defined as persons under investigation (PUIs). For retrospective diagnosis of EVD, nasopharyngeal swab samples were also collected from deceased persons, from whom burial records were obtained. The specimens were tested according to each laboratory's protocol. Results and interpretation were recorded and then reported to the SLMHS. The EBOV-testing records of PUIs included information on individual identification number, name, age, gender, residential address, date of symptom onset, date of specimen collection, date of specimen testing, and interpretation of EBOV-testing result.

More information on household head, clinical manifestations, and contact history was also collected using a standardized WHO case investigation form for most PUIs in three districts: Western Urban, Western Rural and Port Loko, and for a few PUIs in

Kambia and Koinadugu districts from August to December 2014. These data include the name of household head, the names of contacts who were known or suspected Ebola cases, relationship with these contacts, and the dates of contacting them. However, most symptom onset dates of the contacts were not reported. Although limited, these data provide a unique opportunity to assess household transmissibility of the Ebola virus disease (EVD) in Sierra Leone.

We obtained "District" and "Chiefdom" boundary data from the GADM database of Global Administrative Areas version 2.0 (http://www.gadm.org). The 2004 Population and Housing Census information for each chiefdom was obtained from Statistics SL (http://www.statistics.sl). The chiefdom-level population sizes and distribution of age and gender in 2014 were interpolated using the district-level population sizes in 2014 (extracted from http://health.gov.sl/?page_id=583) and the chiefdom-level population sizes in 2004 (Dataset S1). Population density was calculated using the population size divided by the human habitation area. To explore the relationship between the EVD spatial spread and socio-environmental factors, we collected the following data at the chiefdom level: distribution of building and land cover, transportation, locations of hospitals and Ebola treatment centers (ETCs), economic situation (poverty levels), and distribution of ethnic groups. These data were mostly collected from the OpenStreetMap project (http://download.geofabrik.de/africa.html). The distribution of ethnic groups and economic situation were obtained from Wikipedia (http://en.wikipedia.org/wiki/Ethnic_groups_in_Sierra_Leone) and the United States Agency for International Development (USAID)/Sierra Leone, respectively (1, 2).

3. Data Management

A total of 95,089 EBOV-testing records of PUIs from May 2014 to September 2015 were collected. Among them, 14,404 were interpreted as positive for EBOV and 5,100 were commented as pending or indeterminate. From the set of positive PUIs, a total of 8,695 confirmed cases were identified after removing 5,552 duplicated records and 157 records missing too much key information (e.g., name and residential address) which makes it impossible to check duplication. Among these confirmed cases, 8,358 had required information to infer residential chiefdom and disease onset dates and were used in our analyses (Dataset S2).

From the set of pending or indeterminate samples, we obtained 4,179 suspected cases after excluding 359 duplicated records and 562 records of PUIs who were positive for EBOV in a subsequent testing and had been considered as confirmed cases. Further exclusion of 634 individuals without sufficient information to infer residential chiefdom and disease onset dates led to a total of 3,545 suspected cases for our analyses.

There were four date variables in our dataset: symptom onset date, specimen collection date, specimen receipt date and specimen-testing date. The method for inferring the missing symptom onset dates and specimen-related dates was similar to the one used in papers published by the WHO Ebola Response Team (3, 4). To refine the imputation procedures, we applied the algorithm to each district and epidemic phase. Three epidemic phases were considered for this imputation: the early phase before 13 August 2014, the intensified surveillance phase 14 August to 30 October 2014, and the intensified lab-testing phase after 1 November 2014 (Fig. 1 and Table S8). These phases are designed for linking symptom onsets dates to the

specimen-related dates and therefore differ from the ones used for transmission analyses.

The lags between each pair of date variables were evaluated for each case, and unrealistic time lags were flagged, i.e., negative lags from onset to specimen collection, from onset to receipt, from onset to lab-testing, from specimen collection to receipt, from collection to lab-testing, and from receipt to lab-testing. Lags greater than 50 days were also flagged. Unrealistic or unreliable dates were set to missing.

Specifically, missing dates of onset and testing were imputed from the distribution of related time lags based on cases with completely observed dates. For example, for a person with the onset date missing but the specimen collection date observed, we set the onset date to be x days prior to the specimen collection date, where x is the median lag from onset to specimen collection within the district of this person and the epidemic phase to which the specimen collection date belongs. If there were fewer than 10 cases with completely observed relevant dates, the median was then calculated combing all districts in that epidemic phase. If the specimen collection date is also missing but the receipt date or the testing date is available, we use the corresponding lag for imputation of missing onset date. Priority follows the natural order of these dates in proximity to the onset date in general, i.e., specimen collection, receipt and testing. The same rules apply to the imputation of missing dates of specimen collection, receipt and testing.

To identify duplicated records, we created a list of potential duplicate records by finding pairs or groups of individuals that match on surname (encrypted), given names (encrypted), residential district, age, gender and dates. For each variable except for surname, a missing value is considered a match to any other value

including another missing value. For surname, we required non-missing values for matching. We then imposed the constraint that any pairs or groups of individuals are considered matched only if at least two of the following characteristics are matched: 1) date of onset or date of testing or date of specimen collection, 2) district or chiefdom of residence, 3) age (in years), 4)gender, 5) patient ID, and 6) laboratory ID. One additional constraint is that the difference in age between a matched pair must be less than 10 years. Once a pair of records was identified as duplicates, we merged the records into a single one. If some variables are heterogeneous in a matched pair or group, the following rules apply: 1) if a variable is missing in some individuals but identical among others, the identical value is retained; 2) if multiple non-missing values exist for a variable, based on observed dates, the value of the earliest record is retained if the variable is a date; otherwise, the value of the latest record is retained. The rationale is that latter records of the same person may have the most up-to-date information.

4. Temporal and Spatial Analyses

Epidemic curves showing temporal dynamic of the EVD epidemic were created by plotting the daily number of symptom onsets of confirmed and suspected EVD cases, together with a 7-day moving average and the cumulative numbers of affected chiefdoms over time. To display the spatial distribution of EVD in Sierra Leone, each confirmed case was geo-referenced and linked to the digital map of Sierra Leone according to its residential chiefdom using Geographic Information System technologies. A thematic map was created through displaying the cumulative number of confirmed EVD cases in each of affected chiefdoms.

A series of thematic maps of monthly chiefdom-specific incidence rates was created to document the spatial-temporal dynamic of EVD diffusion in Sierra Leone from May 2014 to September 2015. A spatial trend contour plot of the EVD spread was also developed using trend surface analysis on the invasion time of each chiefdom which was defined as the time lag between the first confirmed case for each chiefdom and May 18, 2014, the inferred date of symptom onset of the first confirmed case in Sierra Leone (5, 6). Chiefdom-specific temporal patterns of EVD were presented as heat maps of weekly number of cases over the epidemic period (7, 8). The weighted-average linkage method for the hierarchical cluster analysis was used based on the following five epidemic characteristics: the cumulative number of cases, the cumulative number of weeks affected by EVD, the maximum number of weekly cases, the maximum number of continuous weeks affected by EVD, and the cumulative number of cases within the continuous weeks affected by EVD. A higher number of the category indicates a more sustained outbreak of the disease in the chiefdom. A thematic map was produced to display the spatial distribution of the six categories combined. Typical epidemic curves for the categories IV–VI are also shown.

5. Analysis of time to invasion of EVD at chiefdom level

In this study, the following ten demographic, mobility-related and environmental factors at the chiefdom level were collected to evaluate their possible contribution to the time to invasion of EVD at the chiefdom level: population density, indicators for intersection with primary roads, secondary roads, or railroads, distances to the nearest hospital and ETC, and coverage percentages of each land cover: built-up, cropland,

forest and shrub. Univariate and multivariate Cox proportional hazard models were fitted to the chiefdom-level invasion times of EVD, where unaffected chiefdoms were considered right-censored. Hazard ratios for the continuous variables were calculated using the following units: 10 kilometers for distance to the nearest hospital or ETC from the centroid of each chiefdom, 100 persons per km^2 for population density, 1% for coverage percentage of built-up, and 10% for coverage percentage of croplands, forest and shrub. Univariate analysis was performed to examine the effect of each variable individually, and then the variables with a *P* value < 0.20 in the univariate analysis were included in the multivariate analysis. The multivariate model was selected by comparing the log likelihood of the models and the changes of *P* values of model coefficients when a covariate was included or excluded. Two-sided *P* values under 0.05 were considered as statistically significant. Highly correlated variables (spearman correlation coefficients > 0.7) do not enter the model simultaneously. High correlation was found between population density and percentage coverage of built-up (correlation coefficient = 0.98), so the percentage coverage of built-up was removed from the final model. Each continuous variable was also categorized and the median time to invasion was given for each category, In addition, a sensitity analysis was conducted by including all confirmed and suspected cases in the Cox proportional hazard model. Survival analyses were performed using the Stata package (StataCorp LP, College Station, Texas) (10).

6. Timeline of key intervention programs implemented in Sierra Leone

The planning and implementation of intervention programs against Ebola in Sierra Leone was part of WHO- and UN-led efforts of disease control in West African,

coupled with strategies and campaigns initiated by governmental agencies such as the National Ebola Response Center of Sierra Leone as well as support from non-governmental organizations such as Medecins Sans Frontieres (MSF, also known as Doctors without Borders). Due to logistical constraints, the implementation process was more gradual than swift. The following categories of interventions were employed at increasing levels during the epidemic: 1) Surveillance, e.g., case finding, contact tracing, mobile laboratories and lab materials; 2) Case management, e.g., ETCs, community care centers (CCCs), personal protective equipment (PPE); 3) Safe and dignified burials; 4) Social mobilization, e.g., door-to-door educational outreach, radio programming and a call center, communication with local traditional leaders, reporting symptoms and risky behaviors; 5) Psycho-social support for vulnerable groups, in particular children and women; and 6) Enabling services including logistic support, staffing, training, water sanitation and nutrition. Table S8 gives the timeline for rolling out specific intervention measures including the numbers and bed capacity of ETCs and CCCs after late October 2014, which was extracted from weekly situation reports jointly released by UNMEER and Sierra Leone National Ebola Response Center (11, 12).

We provide in the following a possibly incomplete timeline of key intervention events. In Kailahun District where the first cases were identified, the first ETC was opened by the MSF on 24 June 2014. A mobile laboratory was established in Kenema in early July by Public Health Canada with the help from WHO (13). On 6 August, the President of Sierra Leone declared a national state of emergency, and military personnel were dispatched to enforce quarantines in hardest hit areas. After Freetown in Western Area was invaded in June, the situation deteriorated in August, and more

mobile laboratories were established in Kenema and Western Area in by South Africa and US CDC in August, followed by China in September (13). On 28 August, the WHO published a roadmap to guide and coordinate the international response to the outbreak, aiming to stop ongoing Ebola transmission worldwide within 6–9 months, including urgently strengthen the field response, coordinate the outbreak response, and encourage preparedness in countries at-risk (14). On 18 September the UN Security Council held an emergency session and announced the formation of the UN Mission for Emergency Ebola Response (UNMEER) to scale up the international response including air-transporting staff and materials and establishing new treatment centers and burial teams (15). From 19 to 21 September, the Sierra Leone government initiated a three-day nationwide quarantine campaign, and 28,500 social mobilizers (trained community workers and volunteers) went door-to-door to promote education and infection prevention (16). As part of the UNMEER's strategic plan, starting 1 October, Sierra Leone initiated a campaign to achieve isolation of 70% of cases and 70% safe burials in 60 days and to achieve 100% case isolation and 100% safe burials in 90 days (11). More mobile laboratories and ETCs were established during this period. Meanwhile, as the Western Area became the hardest hit, the NERC intensified the campaign in that region via an initiative called "Operation: Western Area Surge" from 17 December (11, 17). By the end of 2014, it was estimated that Sierra Leone had achieved 95% safe and dignified burials (12). With intervention resources gradually meeting and exceeding needs, the epidemic in West Africa, including Sierra Leone, has been declining during 2015. Sierra Leoneans launched a one-month "Zero Ebola" drive in mid-March 2015, including a three-day stay-at-home exercise during 27–29 March (18). Another wave of campaign to get zero infection in Kambia, Port

Loko districts and Western Area, titled “Operation Northern Push”, was launched on 16 June with reinforced surveillance, contact tracing and quarantine (19). In addition, a phase III cluster-randomized non-blinded trial of the VSV-EBOV candidate Ebola vaccine was started among health workers in Freetown, Sierra Leone in April 2015 (20). Starting in late August, the same vaccine was employed for ring vaccination, in which contacts of newly identified cases within a certain geographic range were vaccinated (21).

Based on this timeline, the formation of UNMEER in late September is a key turning-point in the scale of the fight against Ebola, and fastest pace of improving the quality and quantity of intervention programs occurred during October–December 2014 (Table S8). Indeed, the epidemic started its decline at the end of December 2014 and the beginning of January 2015 (Fig. 1). Accordingly, we divide the interventions into three phases: reference phase (18 May–28 September 2014), intervention phase I (29 September–28 December 2014, weeks 40–52), and intervention phase II (29 December 2014–13 September 2015, weeks 53–89). The dates were chosen to round to the closest calendar weeks.

7. Population-level transmissibility adjusted for risk factors

A Poisson regression designed to account for both case importation and local transmission was used to explore the effect of the following socio-environmental factors on the population-level transmission of EVD at chiefdom and week level: population size and density, weekly average temperature and relative humidity with a two-week lag, distances to nearest primary roads, secondary roads, or railroads, distances to the nearest hospital and ETC, and coverage percentages of cropland,

forest and shrub, poverty level, intervention phase, and primary ethnic groups. Temperature and relative humidity were obtained from the National Climatic Data Center, National Oceanic and Atmospheric Administration of the U.S. They were available only for the capital city Freetown. As Sierra Leone is a relatively small country, we do not expect much spatial variation in climatic conditions, and thus assume the same climatic data apply to all chiefdoms. The auto-correlation between weekly number of cases and each of the climatic variables is the highest at a two-week lag, which is then used in this analysis. Poverty level (high: ≥20% extreme poverty, low: <20% extreme poverty) and intervention phases (reference phase, intervention phase I and intervention phase II) are binary indicators, and poverty level is district-specific. According to the suggestion of the Sierra Leone Ministry of Health and Sanitation, five primary ethnic groups in Sierra Leone are considered in our analysis: Mende (49 chiefdoms), Temne (27 chiefdoms), Kono (12 chiefdoms), Limba (seven chiefdoms) and others (55 chiefdoms).

Let $Y_i(t)$ be the number of symptom onsets in chiefdom i during week t. Let $Z_i(t) = \sum_{d=1}^{D} \omega_d Y_i(t-d)$ be the effective number of infectious cases in chiefdom i who are capable of generating new cases during week t, where ω_d is the probability of the serial interval being d weeks, $d = 1, \cdots, D$. The maximum duration of the serial interval is set to $D = 4$ weeks. Let B_i be the collection of neighboring chiefdoms surrounding chiefdom i. We consider three possible channels of generation of new cases: importation from distant chiefdoms, transmission from cases within the chiefdom, and transmission from neighboring chiefdoms. We assume $Y_i(t) \sim Poisson(N_i \gamma_i(t))$, where N_i is the population size of chiefdom i and

$$\gamma_i(t) = \left\{ \gamma_0 + \gamma_1 \left[Z_i(t) + \theta \sum\nolimits_{j \in B_i} Z_j(t) \right] \right\} e^{\beta' X_i(t)}$$

is the effective rate of generating new cases adjusted for covariates $X_i(t)$, γ_0 is the baseline rate of importing new cases from distant chiefdoms, and γ_1 is the rate a fully infectious local case generates new cases in the same chiefdom. The differential transmission rate across adjacent chiefdoms compared to within chiefdom is captured by θ, which is presumably less than 1. We are interested in estimating γ_0, γ_1, θ and the covariate effects β,and the more interpretable risk ratios e^{β}. The estimation procedure proceeds as follows: 1) choose initial values of parameters $\hat{\gamma}_0^{(0)}$, $\hat{\gamma}_1^{(0)}$ and $\hat{\theta}^{(0)}$; 2) At iteration k, obtain estimate $\hat{\beta}^{(k)}$ using the gam function in the mgcv package (packaged for R by Simon Wood, 2015) with $\log(N_i) + \log\left\{\hat{\gamma}_0^{(k-1)} + \hat{\gamma}_1^{(k-1)}\left[Z_i(t) + \hat{\theta}^{(k-1)}\sum_{j\in B_i} Z_j(t)\right]\right\}$ as the offset; 3) Given $\hat{\beta}^{(k)}$, maximize the Poisson likelihood to obtain $\hat{\gamma}_0^{(k)}$, $\hat{\gamma}_1^{(k)}$ and $\hat{\theta}^{(k)}$; 4) reiterate until convergence in the parameter estimates. Variances of the estimates are obtained using the bootstrap. To ensure identifiability, we only consider up to cubic terms for each covariate. A covariate is retained in the final model only if the z-score for at least one polynomial term of that covariate is ≥5. Our final model includes population density as a covariate in$X_i(t)$, and N_i is thus removed from the model as it is highly correlated with population size in Sierra Leone.

Confirmed cases with symptom onsets during weeks 23–88 (2 June 2014 —6 September 2015) were included in this analysis. Thirty-seven cases before week 23 contribute to the Poisson likelihood as infectors but not as infectees, a consequence of adjusting for selection bias as these cases were likely early infections imported from neighboring countries.

8. **Household transmissibility of EVD**

A household in this study was defined as a group of related family members who live at the same street address and therefore have daily close contact with each other. Based on the individual data on the standardized WHO case investigation form, identification of possible households began by matching individuals who were from the same district and provided the same names (encrypted) of household heads and street addresses (encrypted). Due to the high frequency of identical names, we imposed a relatively stringent matching criterion on street address. If the street addresses are blank or are identical to the names of three most densely populated cities, i.e., “Freetown” (Western Urban District), “Waterloo” (Western Rural District) and “Port Loko” (Port Loko District), the individuals were considered unmatched. Household heads were included in the household of the individual who reported them. Addresses of case-reported contacts were not available but they were assumed living together with the cases, unless the reported relationship suggests that living in the same household is unlikely, e.g., co-workers and neighbors. If an individual cannot be matched to anybody else other than reported names of household head or source contacts, these people also constitute an identified household.

After initial screening, we obtain 775 possible incomplete households. They are incomplete in the sense that any household member who was not lab-tested or reported as household head or EVD source contacts was not captured by our data. Household members found from the data are referred to as observed members. We impute the actual size of each household from the census data, and the imputed household members are referred to as unobserved members. Many source contacts could not be found in the laboratory-test database or standard case report forms,

possibly due to logistic constraints during the process of data collection. In our primary analyses, we assume these subjects were indeed infected. Among these 775 households, 82 had EVD source contacts with neither the date of contact nor the date of symptom onset, making it difficult to infer a reasonable range of symptom onsets. As a result, we excluded these households. We then excluded 59 households containing a single observed member. The majority of these individuals reported themselves as the household heads, implying that these were most likely single-member households and thus contributed no information about secondary transmission. The remaining 634 households constitute the basis for our household transmission analyses. Further exclusion of households with only non-infected members is merely for computational efficiency, and such exclusion depends on the definition of "infection". Two definitions were used: (I) having a positive lab-test or being an source contact, and (II) having an EVD onset date or being a source contact. For both definitions, we assume household members, mostly household heads, who had no symptom onset dates and no lab-test results and were not in the list of source contacts of that household, were not infected. Under definition I and II respectively, we then excluded 238 and 47 households with only non-infected members, leaving 396 and 587 households for the final analyses.

Natural history of Ebola in human cases. The natural history of Ebola infection in humans has been documented for the early phase of the outbreak in West Africa including Sierra Leone (3, 22, 23). However, it is difficult to assess the natural history from our data, due to the lack of exact exposure dates for all cases. We make assumptions about the natural history based on the published estimates (3, 22). We assume the incubation period (time from infection to symptom onset) follows a

gamma distribution truncated at 21 days with mean of 10 days and standard deviation (SD) of 4.7 days. We then vary the shape (α=2.78) and scale (β=3.6) parameters of the truncated gamma by 20% to generate two sensitivity settings: (α=2.22, β=2.88) and (α=3.33, β=4.32), which correspond to (mean=7, SD=4.0) and (mean=12, SD=4.8). We assume the incubation period overlaps with the latent period (time from infection to the onset of infectiousness). The distribution is a mixture of the gamma distribution of the time from symptom onset to hospital discharge and the gamma distribution of the time from onset to death with a mixing weight of 0.5. We assume the infectious period (time that an infected person is infectious to others) starts with the onset of symptoms and also follows a gamma distribution truncated at 30 days with a mean of 11 days and a SD of 6 6 days (3). Similar to the incubation period, the shape and scale parameters of the truncated gamma were varied by 20% to yield sensitivity settings with (mean=7, SD=4.7) and (mean=16, SD=8.0) respectively.

Missing age group, gender and symptom onset dates. Among the 1544 observed household members in the 634 households prior to the exclusion of non-attacked households, 60 were Ebola-positive but with missing dates of symptom onset, and 160 had missing test results, after those who had no symptom onset dates and no lab-test results and were not in the list of source contacts were assigned as non-infected. In addition, either age or gender is unknown for 376 individuals. The proportion of children and the proportion of females were calculated by household size ($\leq$2, 3 and $\geq$4) and infection status (yes, no and unknown). Missing values in age group and gender were then imputed using these proportions. To counterbalance the potential bias in age group, household heads were not included in the calculation of proportions of children as they were generally adults. The number of missing test

results were further reduced by assigning source contacts as infected (definitions of infection I and II) or by assigning individuals with symptom onset dates as infected (definition of infection II). The remaining individuals with missing test results were assumed to have two possible states: infected or uninfected. If an individual was either infected or possibly infected but the symptom onset date is missing, we determine the range of possible symptom onset dates based on observed data, in the order of priority and availability: 1) 4—9 days prior to the specimen collection date, 2) 5—10 days prior to the specimen testing date, or 3) 1—14 days prior to the date of contacting.

Distribution of household sizes Suppose that we found m individuals in our data coming from the same household. We then know that the actual household size, denoted by n, is $n \geq m$. From census data we can obtain the distribution of household size as $\Pr(n = k)$ for a range of reasonable integers $k \geq 1$. To impute the actual household size given that $n \geq m$, we need to compute the conditional probabilities $\Pr(n = k | n \geq m) = \Pr(n = k) / \Pr(n \geq m)$, for $k \geq m$. From the 2004 census data of Sierra Leone, we found the following distributions of household size in Western Area and in the nation as a whole (Table S9). As household sizes of 10 or more are summarized in one category, we fit a right-censored parametric distribution to the frequencies. Out of negative binomial, gamma and Weibull distributions, the gamma distribution provides the best fit. The parametric estimates were then used to derive the probabilities of $\Pr(n = k)$ for $1 \leq k \leq 15$, where we assume 15 is the maximum household size. The exact fitted distributions are given in Table S10. We use this distribution to impute the actual size of each household, and we assume the imputed households' members were not infected.

Household transmission model To account for the uncertainty due to missing data, we use a likelihood-based statistical transmission model coupled with the Expectation-Maximization (EM) algorithm and multiple imputation. The transmission model with the EM algorithm (24) has been implemented in the software tool TranStat (readers can download it from https://dl.dropboxusercontent.com/u/60465401/TRANSTAT/index.htm) to handle missing outcomes, e.g., missing infection status or onset days. We use multiple imputation to deal with missingness in household size and risk predictors, i.e., age group and gender. A similar multiple imputation approach was used to analyze household transmission of the pandemic influenza H1N1 in 2009 (25).

Suppose we observe an epidemic from day 1 to day *T* among a population of *N* individuals in *H* households. For now, also assume there is no missingness in household size and risk predictors. We consider two types of transmission routes. A susceptible person could be infected via casual contact with non-household-member cases in the community or by infectious household members. The daily probabilities of infection of a susceptible person are denoted by *b* if the source is the community at large and by *p* if the source is an infectious household member. These transmission probabilities may be adjusted for covariates, such as age group and gender, via logistic regressions: $\text{logit}(b_{it}) = \text{logit}(b) + X'_{it}\beta_C$ and $\text{logit}(p_{ijt}) = \text{logit}(p) + X'_{it}\beta_S + X'_{jt}\beta_I$, where X_{it} are covariates associated with person i on day t. The coefficient β_C measures the covariate effects on modifying susceptibility for community-to-person transmission, and β_S and β_I measure the effects on susceptibility and infectiousness for household transmission. Suppose that we have observed households ascertained by index cases. The SAR over an infectious period

of D days is $\mathrm{SAR} = 1 - \prod_{k=1}^{D} 1 - p\, g(k)$, where $g(k)$ is the probability that an infected person remains infectious at day k since symptom onset, which is derived from the distribution of the infectious period. The SAR specific to covariate values X_{it} and X_{jt} is given by replacing p with logit^{-1} [logit(p) + $X'_{it}\beta_S + X'_{jt}\beta_I$].

Let $h(i)$ be the collection of members of the household, and $\tilde{t}_i$ be the symptom onset day, of individual i. The probabilities that susceptible individual i escapes infection during day t and up to day t are

$$e_{it} = (1 - b_{it}) \prod_{j \in h(i), j \neq i} \left[1 - p_{ijt}\, g\left(t - \tilde{t}_j + 1\right)\right] \quad \text{and} \quad q_{it} = \prod_{l=1}^{t} e_{il}.$$

Let $f(k)$ be the probability that the incubation period is k days. Let Y_i be the infection status of person i (1=yes, 0=no), and define $\theta = \{b, p, \beta_C, \beta_S, \beta_I\}$. Let $\tilde{t}_{h_i}$ be the collection of observed symptom onset days, and Y_{h_i} the collection of infection status, of all members in the household of individual i. When Y_{h_i} and $\tilde{t}_{h_i}$ are completely observed, the likelihood contribution of individual i is

$$L_{(i)}(\theta; Y_i, \tilde{t}_i) = {q_{iT}}^{1-Y_i} \left(\sum_t f(\tilde{t}_i - t) q_{i(t-1)} [1 - e_{it}] \right)^{Y_i}.$$

In our situation, a household is included in the analysis only if there had been confirmed cases. Households where all members tested negative were excluded. These households appeared in our data because some of the members reported symptoms. It is unclear how representative these households are of the general uninfected households, and therefore they must be excluded to avoid potential bias. Inclusion of only infected households is called case-ascertained design and is commonly seen in household-based clinical studies. With this design, the individual likelihood needs slight modification to correct bias in the estimation of b (26). Let d_{min} and d_{max} be the minimum and maximum duration of the incubation period.

Let t_h^* be the symptom onset day of the index case, defined as the member who has the earliest onset day, in household h. The likelihood for a non-index-case individual should be changed to

$$L_{(i)}(\theta; Y_i, \tilde{t}_i) = q_{i(t_h^*-d_{max},T)}{}^{1-Y_i}\left(\sum_t f(\tilde{t}_i - t) q_{i(t_h^*-d_{max},t-1)}[1-e_{it}]\right)^{Y_i},$$

where $q_{i(t_1,t_2)} = \prod_{l=t_1}^{t_2} e_{il}$ is the cumulative escape probability from t_1 to t_2. The index cases do not contribute their infection events to the likelihood, but the exposure of other household members to the index cases is accounted for in the likelihood.

To account for missingness in Y_i or $\tilde{t}_i$ for some i, let $u_i = (Y_i, \tilde{t}_i)$, and define, for each household h,

$$O_h = \{u_i : i \in h \text{ and } u_i \text{ is completely observed}\} \text{ , and}$$

$$U_h = \{u_i : i \in h \text{ and } u_i \text{ is not completely observed}\},$$

i.e., completely observed outcomes and missing outcomes. For the study population, define $O = \{O_h : h = 1, \cdots, H\}$ and $U = \{U_h : h = 1, \cdots, H\}$. Assuming independence between households, the household-level and population-level complete-data likelihoods are

$$\mathrm{L}_h(\theta; O_h, U_h) = \prod_{i \in h} L_{(i)}(\theta; u_i) \text{ and } L(\theta; O, U) = \prod_{h=1}^{H} \mathrm{L}_h(\theta; O_h, U_h).$$

Let $\{U_{hk}^* : k = 1, \cdots, \delta_h\}$ be the collection of all possible realizations of U_h, where δ_h is the number of such possibilities. If outcomes are observed for all members, then U_h is empty and we let $\delta_h = 1$. Under the assumption that the infection outcomes are missing at random (MAR), the E-M algorithm proceeds as follows: 1) Partition the households into two groups: $\Delta_{OBS} = \{h : \delta_h = 1\}$ and $\Delta_{EM} = \{h : \delta_h > 1\}$. 2) Choose initial values of parameters $\hat{\theta}^{(0)}$. 3) At iteration $r \geq 0$, update the conditional probabilities for all $h \in \Delta_{EM}$

$\lambda_{hk}^{(r)} = \frac{L_h(\hat{\theta}^{(0)};O_h,U_{hk}^*)}{\sum_{l=1}^{\delta_h} L_h(\hat{\theta}^{(0)};O_h,U_{hl}^*)},\ k = 1,\cdots,\delta_h,$ and maximize

$$Q\left(\theta\middle|\hat{\theta}^{(r)}\right) = \sum_{h\in\Delta_{OBS}} \ln \mathrm{L}_h(\theta;O_h) + \sum_{h\in\Delta_{EM}} \sum_{k=1}^{\delta_h} \lambda_{hk}^{(r)} \ln \mathrm{L}_h(\theta;O_h,U_{hk}^*)$$

with regard to θ to find $\hat{\theta}^{(r+1)}$. Repeat this step until convergence in the estimate of θ. Denote the final estimate by $\hat{\theta}$.

The variance of $\hat{\theta}$, denoted by $V_{EM}(\hat{\theta})$, can be calculated using the missing information formula in combination with importance sampling (24). When the household sizes and covariates are also subject to missingness, we use multiple imputation to account for uncertainty in the final variance estimate of θ. Let n_h be the actual size of household h and let $n = (n_1,\cdots,n_H)$. Let $X_i = (X_{i,1},\cdots,X_{i,K})$ be the K-dimensional covariates associated with person i, and let $X = (X_1,\cdots,X_N)$. The covariates used in regression models, X_{ict} and X_{ijt}, are subsets of X_i. When n_h is missing, we sample it from the distributions Table S9 or Table S10, depending on whether the household was in the Western Area or not. When $X_{i,k}$ is missing, we sample it from the empirical frequencies of observed values of $X_{i,k}$ stratified by the observed (truncated) household size (≤2, 3, and ≥4, chosen to ensure sufficient number of observations in each category) and the lab-testing outcome (negative, positive, and unknown). Suppose that we draw M random samples of missing components of n and X, so that we have M sets of n and X denoted as $\left(n^{(m)},X^{(m)}\right),\ m = 1,\cdots,M$. Let $\hat{\theta}_{(m)}$ and $V_{EM}(\hat{\theta}_{(m)})$ be the EM point estimate and variance estimate conditional on the m^{th} set. The final point estimate and variance estimate are given by $\tilde{\theta} = \frac{1}{M}\sum_{m=1}^{M}\hat{\theta}_{(m)}$ and $V\left(\tilde{\theta}\right) = \frac{1}{M}\sum_{m=1}^{M}\hat{\theta}_{(m)}^2 - \tilde{\theta}^2 + \frac{1}{M}\sum_{m=1}^{M} V_{EM}(\hat{\theta}_{(m)})$.

Adjusting for risk factors For within-household person-to-person transmission, we

adjust the transmission probability for age group (1: children (<18 years), 0: adults) and gender (1: male, 0: female) in our primary analysis, and the corresponding model is $\text{logit}(p_{ijt}) = \text{logit}(p) + \beta_1 X_i^{(AGE)} + \beta_2 X_i^{(SEX)}$. In addition, we also explore potential effects of age and gender on modifying infectiousness using the model $\text{logit}(p_{ijt}) = \text{logit}(p) + \beta_1 X_i^{(AGE)} + \beta_2 X_i^{(SEX)} + \beta_3 X_j^{(AGE)} + \beta_4 X_j^{(SEX)}$. Statistical inference about infectiousness generally requires a larger sample size than that about susceptibility (26). In the presence of high-dimensional missing covariates, results from this complex model are reported as a secondary analysis. In the regression for community-to-person transmission, we do not adjust for demographic covariates, because information about such transmission is scarce when only infected households are used in analysis. However, it is unrealistic to assume the community-to-person transmission force is homogeneous spatially and temporally. To account for potential heterogeneity, we stratify b by area (Western Area versus other areas), i.e., incorporate a binary indicator, $X_i^{(WEST)}$ (1=western area, 0=other) in the regression. Area or Province is an administrative geographic unit above district. There are one area and three provinces in Sierra Leone: Western Area (composed of Western Urban and Rural Districts), Eastern Province, Northern Province and Southern Province. The majority of households in our data were from Western Urban District, Western Rural District and Port Loko District. The incidence rates over time are similar between the Western Area and Port Loko District (Fig. S3); however, population density is higher in the Western Area, which may imply stronger force of community-to-person transmission. To reflect time variation, we let b_{it} be proportional (after logit transformation) to the standardized number of onsets of confirmed cases in our study region on day t. The number of onsets is standardized

by first applying a seven-day moving average, then dividing the smoothed number at each day by the overall average smoothed case number. With this standardization, b has an interpretation of average daily infection probability of a susceptible by the community during the outbreak. The standardized onset numbers, $X_t^{(ONSET)}$, is incorporated as an offset, i.e., with a fixed coefficient of 1. Specifically, we use $\text{logit}(b_{it}) = \text{logit}(b) + \beta_0 X_i^{(WEST)} + X_t^{(ONSET)}$. To avoid non-identifiability, we fix $\beta_0 = 2$ in our primary analysis, and perform sensitivity analysis with $\beta_0 = 1$ and $\beta_0 = 3$. The variation of β_0 led to moderate changes in the estimates of the community-to-person transmission risk but gave essentially the same results on all other parameters including the SAR. For example, in a model adjusted for age group and gender effects on both susceptibility and infectiousness, the estimated baseline SAR (between an infectious female adult and a susceptible female adult) is 0.062 when the relative community infection risk is $\beta_0 = 2$ and is 0.061 when $\beta_0 = 1$.

To explore the effect of nationwide interventions on household transmissibility, we then assume the transmission probability $\boldsymbol{p}$ differs between on or before 22 October 2014 and after, and estimate the SARs for the two periods respectively. This time point was the date when the nationwide campaign to achieve 100% case isolation and safe burial had been ongoing for three weeks (the longest incubation period of EVD).

References

1. Washington DC (2009) Food and nutrition technical assistance II project (FANTA-2), academy for educational development (AED). Available at http://pdf.usaid.gov/pdf_docs/PNADH966.pdf. Accessed 30 January 2016.
2. Woldt M, Cadrin M, Jalloh A (2009) USAID office of food for peace Sierra

Leone food security country framework FY 2010–FY2014. Available at http://pdf.usaid.gov/pdf_docs/PDACU314.pdf. Accessed 30 January 2016.

3. WHO-Ebola-Response-Team (2014) Ebola virus disease in West Africa--the first 9 months of the epidemic and forward projections. *N Engl J Med* 371(16):1481–1495.

4. WHO-Ebola-Response-Team (2015) West African Ebola epidemic after one year-slowing but not yet under control. *N Engl J Med* 372(6):584–587.

5. Berke O (1999) Estimation and prediction in the spatial linear model. *Water Air Soil Pollut* 110(3–4):215–237.

6. Berrang-Ford L, Berke O, Abdelrahman L, Waltner-Toews D, McDermott J (2006) Spatial analysis of sleeping sickness, southeastern Uganda, 1970–2003. *Emerg Infect Dis* 12(5):813–820.

7. Yu H, et al. (2013) Characterization of regional influenza seasonality patterns in China and implications for vaccination strategies: spatio-temporal modeling of surveillance data. *PLoS Med* 10(11):e1001552.

8. Xing W, et al. (2014) Hand, foot, and mouth disease in China, 2008–12: an epidemiological study. *Lancet Infect Dis* 14(4):308–318.

9. Hemilton LC ed (2009) Statistics with STATA (Cengage Learning).

10. Rabe-Hesketh S, Everitt B. eds (2006) A handbook of statistical analyses using STATA (New York, NY: CRC Press).

11. United Nations (2015) General assembly (A/69/720). Available at http://reliefweb.int/sites/reliefweb.int/files/resources/N1501016.pdf. Accessed 30 January 2016.

12. UNMEER and Sierra Leone National Ebola Response Center (2014) Sierra

Leone: Ebola emergency weekly situation reports. Available at http://nerc.sl/?q=unmeer-nerc-situation-report. Accessed 30 January 2016.

13. WHO (2015). One year into the Ebola epidemic: a deadly, tenacious and unforgiving virus. Available at http:www.who.int/csr/disease/ebola/one-year-report/introduction/en. Accessed 30 January 2016.

14. WHO (2015). WHO issues roadmap for scaled-up response to the Ebola outbreak. Available at http://www.who.int/csr/disease/ebola/ebola-response-roadmap/en/ Accessed 30 January 2016.

15. Ebola 'threat to world security'- UN Security Council. Available at http://www.bbc.com/news/world-africa-29262968 Accessed 30 January 2016.

16. Sierra Leone to impose three-day quarantine in effort to contain Ebola outbreak. Available at http://www.opposingviews.com/i/health/sierra-leone-impose-three-day-quarantine-effort-contain-ebola-outbreak. Accessed 30 January 2016.

17. Sierra Leone National Ebola Response Center (2014). Operational plan for national emergency response surges in EVD hotspots. Available at http://nerc.sl/sites/default/files/docs/20141206-Ops_Plan_WesternAreaSurge_v1.1.pdf. Accessed 30 January 2016.

18. Sierra Leone National Ebola Response Center (2015). Zero Ebola campaign 3 days stay at home – operational overview, March, 2015. Available at http://nerc.sl/sites/default/files/docs/Zero%20Ebola%20Campaign_Operational%20Overview_260315v1_1.pdf. Accessed 30 January 2016.

19. UNICEF (2015). Sierra Leone EVD weekly situation report (17 June 2015). Available at http://www.unicef.org/appeals/files/UNICEF_Sierra_Leone_EVD_Weekly_SitRep_17_June_2015.pdf. Accessed 30 January 2016.

20. WHO (2015). Ebola vaccines, therapies and diagnostics, questions and answers. Available at http://www.who.int/medicines/emp_ebola_q_as/en/. Accessed 30 January 2016.

21. WHO (2015) Guinea ring vaccination trial extended to Sierra Leone to vaccinate contacts of new Ebola case. Available at http://www.who.int/mediacentre/news/releases/2015/ebola-vaccine-sierra-leone/en. Accessed 30 January 2016.

22. Schieffelin JS, et al. (2014) Clinical illness and outcomes in patients with Ebola in Sierra Leone. *N Engl J Med* 371(22):2092–2100.

23. Chertow DS, et al. (2014) Ebola virus disease in West Africa — Clinical manifestations and management. *N Engl J Med* 371(22):2054–2057.

24. Yang Y, Longini IM, Jr., Halloran ME, Obenchain V (2012) A hybrid EM and monte carlo EM algorithm and its application to analysis of transmission of infectious diseases. *Biometrics* 68(4):1238–1249.

25. Yang Y, et al. (2009) The transmissibility and control of pandemic influenza A (H1N1) virus. *Science* 326(5953):729–733.

26. Yang Y, Longini IM, Jr., Halloran ME (2006) Design and evaluation of prophylactic interventions using infectious disease incidence data from close contact groups. *J R Stat Soc Ser C Appl Stat* 55(3):317–330.