Spatiotemporal Modeling of Cholera, Uvira, Democratic Republic of the Congo, 2016–2020

Ruwan Ratnayake, Jackie Knee, Oliver Cumming, Jaime Mufitini Saidi, Baron Bashige Rumedeka, Flavio Finger, Andrew S. Azman, W. John Edmunds, Francesco Checchi,[1] Karin Gallandat[1]

We evaluated the spatiotemporal clustering of rapid diagnostic test–positive cholera cases in Uvira, eastern Democratic Republic of the Congo. We detected spatiotemporal clusters that consistently overlapped with major rivers, and we outlined the extent of zones of increased risk that are compatible with the radii currently used for targeted interventions.

Cholera outbreaks affect communities that lack access to safe water and adequate sanitation (*1*). Spatiotemporal clustering patterns of cholera indicate a high risk of transmission to the neighboring households of new cases (*2*,*3*). Case-area targeted interventions (CATI), consisting of early, multisectoral response within a 100–500-meter radius around case-households, have been proposed to attenuate clustered transmission (*4*). CATIs, driven by water, sanitation, and hygiene interventions, played a major role in response strategies in Haiti and Yemen, and CATIs including oral cholera vaccination helped suppress outbreaks after vaccination campaigns in Cameroon (*5*,*6*). In the Democratic Republic of the Congo (DRC), health officials evaluated water, sanitation, and hygiene targeting strategies within 500 meters around households with cholera cases (*7*). In Kalemie, DRC, and N'Djamena, Chad, researchers estimated a 200-meter zone of increased risk of infection around cholera cases in the first 5 days (*2*). As CATIs become part of routine practice (*4*,*5*), more insight is needed in delineating the spatiotemporal risk zones required to achieve a substantive effect on transmission.

In Uvira, a city in eastern DRC affected by protracted conflict, population displacement, and flooding, cholera is endemic, and stable transmission is punctuated by seasonal outbreaks (*8*). Citywide interventions include an ongoing piped water supply program with household tap installation beginning in late 2019 (*9*) and mass vaccination in mid-2020 (*10*). Using an enhanced surveillance system with rapid diagnostic testing (RDT), we investigated the location, timing, and prediction of clusters to identify outbreaks earlier and trigger early response. We estimated the extent of spatiotemporal zones of increased risk around cases as a proxy for the ideal radius of CATIs.

The Study

We analyzed suspected cases of cholera during 2016–2020 in patients at cholera treatment centers managed by the Uvira Health Zone. Beginning in April 2016, rectal swab samples were collected from suspected cases and RDT tested (Crystal VC O1/O139; Arkray Inc., https://www.arkray.co.in) after a 6-hour enrichment in alkaline peptone water. We classified cases by avenue of residence (i.e., enumeration areas of mean size 1,177 [range 180–5,711] based on 2017 population sizes) (town of Uvira census data, 2018, unpub. data). We used 2 methods to evaluate spatiotemporal clustering. The space–time scan statistic describes local clustering, where cases exceed expected density within a given area, to identify spatiotemporal clusters and assign relative

Author affiliations: London School of Hygiene & Tropical Medicine, London, UK (R. Ratnayake, J. Knee, O. Cumming, W.J. Edmunds, F. Checchi, K. Gallandat); Ministère de la Santé Publique, Division Provinciale de la Santé du Sud-Kivu, Zone de Santé d'Uvira, Uvira, Democratic Republic of the Congo (J.M. Saidi, B.B. Rumedeka); Epicentre, Paris, France (F. Finger); Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA (A.S. Azman); Geneva University Hospitals, Geneva, Switzerland (A.S. Azman)

DOI: http://doi.org/10.3201/eid3008.231137

[1]These senior authors contributed equally to this article.

risk comparing observed versus expected cases inside and outside the cluster (*11*). To assess capacity for early detection of outbreaks, we simulated real-time detection by scanning prospectively (using few cases) and compared the delay with retrospective scanning (using more cases). We calculated the proportion of years that avenues were included in clusters during 2016–2020. The tau statistic (τ) describes global clustering, or the overall tendency for cases to occur near other cases in time and space (*12*), by using a relative risk of an individual in the population within a given distance band (i.e., 100–150 meters) from an incident case, compared with the risk for any individual in the population, becoming a potentially transmission-related case. This statistic suggests the geographic and temporal extents of increased infection risk. We defined the high-risk and elevated-risk zones as the radius where the moving average's lower 95% CI (high risk) and point estimate (elevated risk) cross 1.0 for ≥30 consecutive meters. We based the main analyses on enriched RDT-positive cases. We conducted sensitivity analyses using suspected cases, and given the use of enumeration areas, using simulated household locations (Appendix, https://wwwnc.cdc.gov/EID/article/30/8/23-1137-App1.pdf). We carried out analyses in R software v.4.1.2 (The R Foundation for Statistical Computing, https://www.r-project.org) by using the rsatscan v.1.0.5 (combined with SaTScan v.10.0.2) and IDSpatialStats v.0.3.12 packages.

Among 5,447 suspected cases, 3,456 (63.4%) were tested and 1,493 (43.2%) were RDT positive. We detected 26 significant spatiotemporal clusters (Table). Mean cluster radius was 652 (range 308–1582) meters, mean size was 20 (range 4–48) cases, and mean duration was 24.8 (range 1–58) days. Clustering occurred in similar locations annually (Figure 1). The first day of a retrospectively detected cluster usually anticipated a seasonal outbreak within 1 week, except for 2016 and 2017, when few cases were RDT tested (Figure 2, panel A). The median delay to the early outbreak signal was 1 day (interquartile range 0–3, maximum 23 days), and median size at signal detection was 3 cases (interquartile range 2–7, maximum 21 cases). Large clusters persisted across 2016–2020 and overlapped with major rivers in north-central and southern Uvira (Figure 2, panel B). We observed no changes in cluster locations in 2019, after household tap implementation began (Figures 1, panels D, E). Sensitivity analysis of suspected cases found more clusters

Table. Statistically significant spatiotemporal clusters of RDT-positive cholera cases detected through annual scanning at the avenue level, Uvira, Democratic Republic of the Congo, 2016−2020*

Year	No.	Cases observed: expected	Population at risk	RR†	Cluster radius, meters	Cluster start date	Cluster duration, d	Signal delay, d‡	Size at signal, no. cases
2016	1	20:1	30,553	20.9§	1,140	Aug 5	18	8	11
	2	28:3	34,232	10.5§	497	Jun 25	48	0	2
	3	17:1	30,758	13.8§	717	Jul 22	23	5	12
	4	15:1	31,240	11.9§	758	Jun 29	23	1	4
	5	4:0	6,579	344.4§	376	Apr 9	1	0	3
	6	14:2	30,082	8.8§	668	Jul 21	30	0	3
	7	9:1	27,452	12.6¶	368	Jul 26	14	3	4
2017	1	48:4	51,012	13.0§	811	Aug 7	40	2	2
	2	32:2	43,992	16.4§	657	Aug 20	23	1	13
	3	32:4	49,794	7.7§	880	Aug 23	44	0	2
	4	13:1	51,016	16.4§	378	Dec 24	7	0	2
	5	12:2	50,635	7.6¶	368	Aug 23	15	12	2
2018	1	20:1	28,884	26.6§	1,116	Oct 26	13	6	9
	2	11:1	31,204	22.7§	475	Feb 13	7	0	3
	3	8:0	25,148	40.6§	662	Aug 28	3	0	4
	4	7:0	17,345	18.6¶	308	Nov 10	10	1	3
2019	1	23:1	33,751	18.6§	743	Sep 10	18	1	7
	2	21:3	33,162	9.0§	755	Sep 7	35	0	12
	3	12:1	16,210	12.3§	309	Apr 27	29	1	2
	4	11:1	16,495	13.2§	527	Sep 7	24	0	2
	5	6:0	15,001	27.8¶	368	Jun 30	6	0	2
2020	1	42:6	60,378	7.8§	1,048	Jul 29	58	2	3
	2	27:3	42,423	8.7§	599	Jul 15	46	23	21
	3	17:1	56,029	19.1§	1,582	Feb 20	9	0	2
	4	30:5	63,207	6.5§	343	May 30	46	2	6
	5	32:6	63,593	5.8§	501	Jun 1	55	4	6

*RDT, rapid diagnostic test; RR, relative risk.
†p values indicate the statistical significance of clusters derived from Monte Carlo simulations.
‡Signal delay indicates the number of days between retrospective detection date with all available data and the earliest prospective detection date.
§p<0.001.
¶p<0.05.

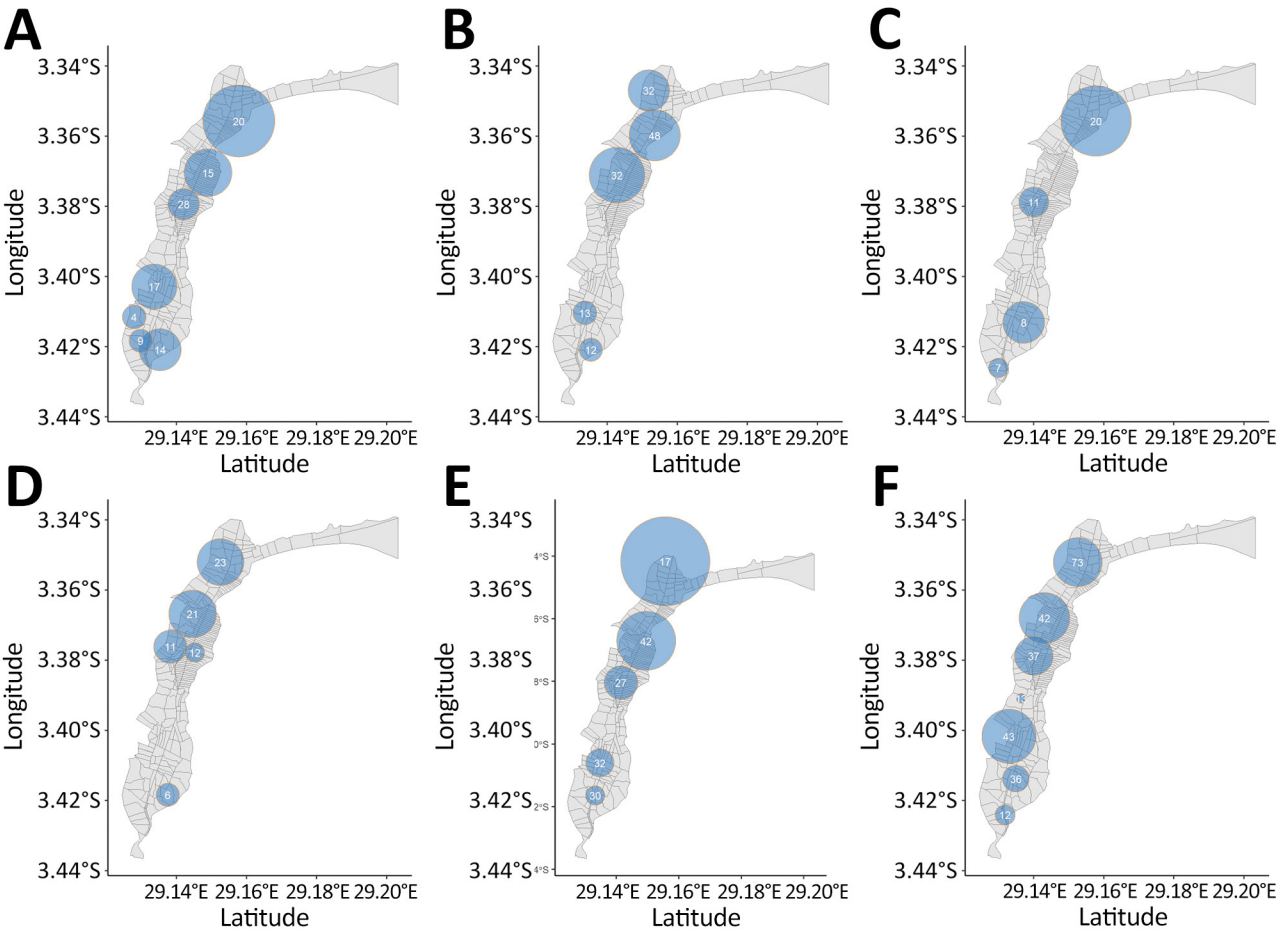


Figure 1. Spatial distribution of spatiotemporal clusters of rapid diagnostic test–positive cholera cases at the avenue level, Uvira, Democratic Republic of the Congo, 2016–2020. A) 2016; B) 2017; C) 2018; D) 2019; E) 2020; F) 2016–2020. Clusters have a relative risk >1 (p<0.05). The sizes of the light blue circles depict the spatial radius and the numbers of cases are shown inside the circles.

(n = 32) in similar locations with similar mean radii (668 [range 331–1,557] meters), larger mean size (42 [range 6–130] cases), and longer duration (27.8 [range 1–59] days) (Appendix Table 2, Figure 5).

In 2016–2020, within 5 days after cases began, the high-risk zone extended to 1,105 meters, and risk remained elevated up to 1,665 meters (maximum moving average τ = 1.8, 95% CI 1.4–2.3) (Figure 3, panel A). During days 1–4, which is more realistic for response, risk zones remained similar (Figure 3, panel D). In 2020, the high-risk zone extended to 585 meters and risk remained elevated up to 1,915 meters (τ = 1.8, 95% CI 1.0–2.9) (Figure 3, panel B). During days 1–4, the risk zones were 425 meters (high risk) and 1,915 meters (τ = 1.7, 95% CI 1.1–2.6) (Figure 3, panel E). Results were similar when we used simulated household locations (during days 0–4) with a moving average τ≥2.0 at 75–275 meters (τ = 2.4, 95% CI 1.7–3.3) and high-risk zone radius (1,415 meters) (Appendix Table 1, Figure 4). Annual results showed lower high-risk (425 meters, except 2017, when it was 875 meters) and elevated (1,125–1,485 meters) zone ranges and no discernable changes after 2019, when household tap implementation began (Appendix Figure 6). Using suspected cases from 2020, the trends remained similar (Figure 3, panels E, F).

Conclusions

We detected spatiotemporal clustering of cholera outbreaks during 2016–2020 in Uvira, DRC, that could inform early mitigation of seasonal outbreaks. The clustering methods produced aligned results compatible with a high-risk radius of ≤500 meters, as previously used for CATI in DRC (*7*,*13*) and similar to clustering in Matlab, Bangladesh, and coastal Sabah, Malaysia (500 meters, ≈5 days after cases began) (*3*,*14*). For RDT-positive cases within 5 days after cases began, we estimated a 1,105-meter high-risk radius, showing that a ≤1,000-meter risk window is optimal. Scan

statistics detected a similar mean cluster radius of 650 meters. The simulated real-time scanning usually signaled an outbreak with a 1-day median delay, which would enable early control.

We used enriched RDT-positive cases to increase specificity, but among study limitations is that we relied on medically attended cases at a cholera treatment center, biasing toward severely dehydrated case-patients and against milder cases. The spatial resolution misses case-pair distances <420 meters, where 5% of distances fell, although simulation of household locations showed similar trends with even higher (τ) across smaller radii. Circular scan statistics have reduced sensitivity to outline the shape of elliptical clusters (potentially along Uvira's coastline), but detection appeared unaffected (*11*).

Conspicuously, the clusters endured annually and overlapped with Uvira's 3 major rivers. According to surveys in 2016, 2017, and 2021, households in those clusters commonly use rivers as a primary water source (K. Gallandat et al., unpub. data) because piped water has remained inconsistent (*15*). Combined with

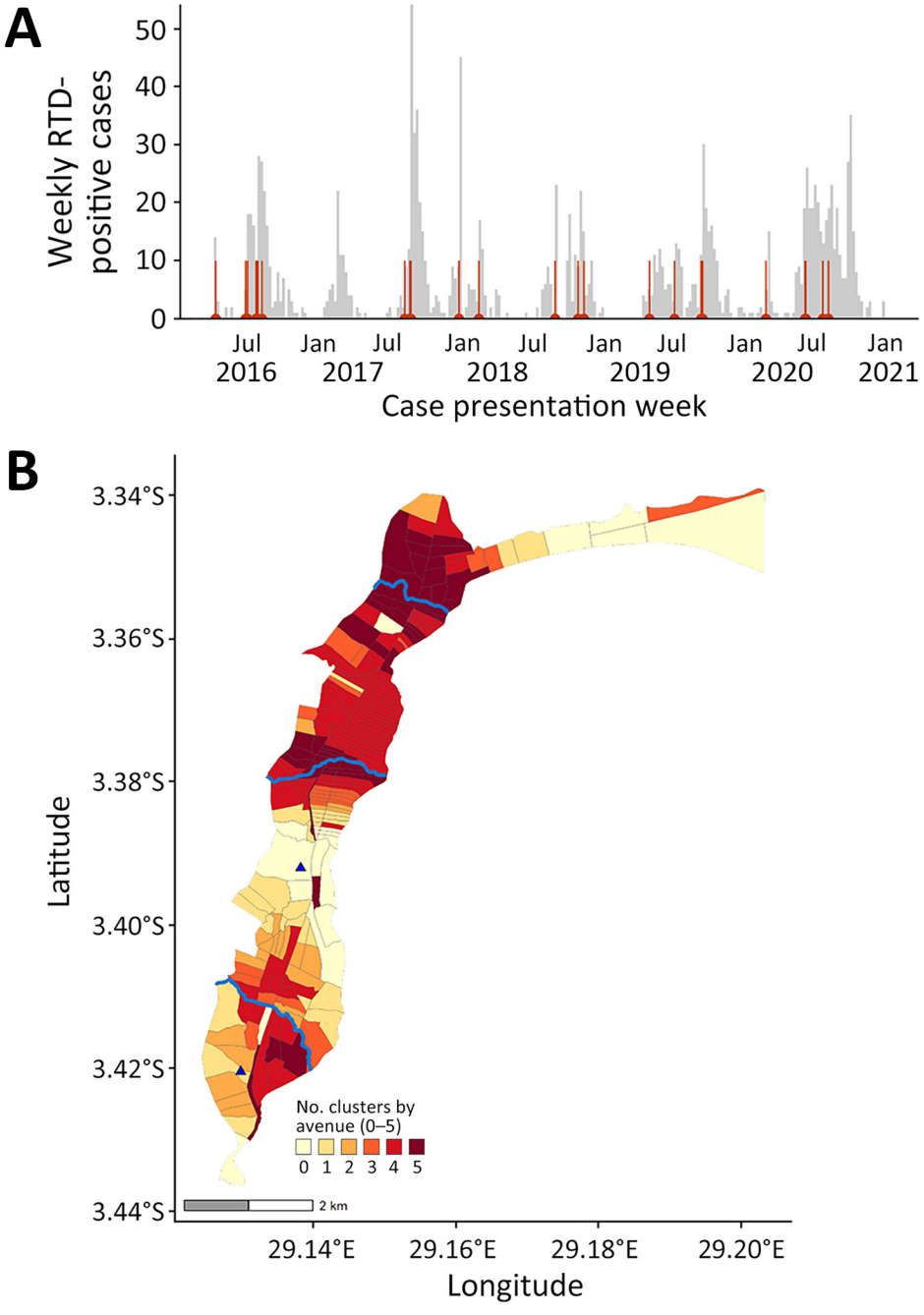


Figure 2. Epidemic curve and cluster persistence in study of spatiotemporal modeling of cholera, Uvira, Democratic Republic of the Congo, 2016–2020. A) Epidemic curve shows weekly numbers of RDT-positive cholera cases based on week of onset and start dates of 26 clusters (red vertical lines). B) Cluster persistence within avenues for RDT-positive cases showing the number of years affected by clustering within avenues and proximity to rivers (blue lines, top to bottom: Kalimabenge River, Mulongwe River, Kanvinvira River). Blue triangles indicate cholera treatment center (top) and unit (bottom). RDT, rapid diagnostic test.

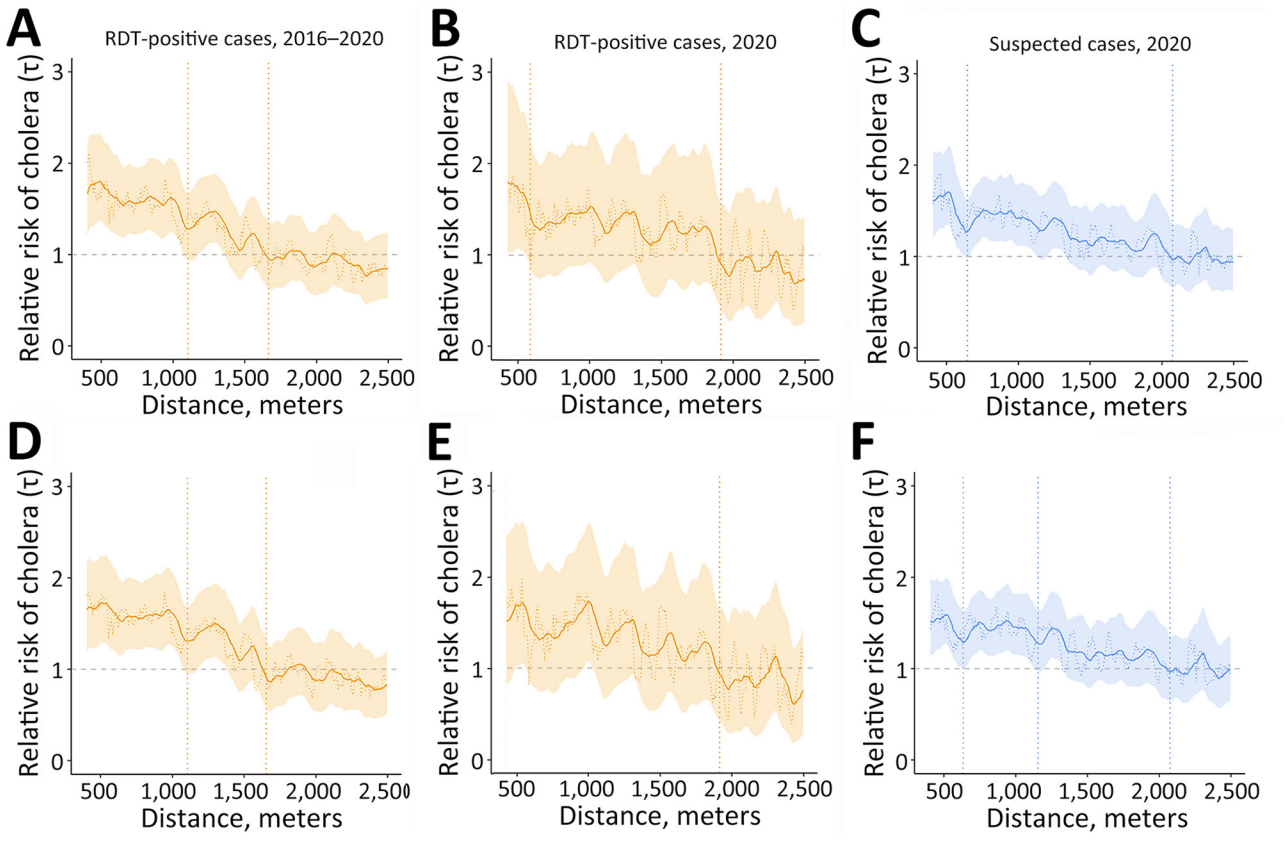


Figure 3. Moving average estimates for RDT-positive and suspected cholera cases in study of spatiotemporal modeling of cholera, Uvira, Democratic Republic of the Congo, 2016–2020. Moving average estimates of (relative risk) and 95% CIs (shading) are shown with point estimates (dashed horizontal lines) for days 0–4 (panels A–C) and days 1–4 (panels, D–F), for RDT-positive cases (orange) and suspected cases (blue), using 1,000 bootstrap samples. The vertical dashed lines indicate the spatial extent of the zone of high-risk where the lower 95% CI crossed 1.0 for ≥30 meters consecutively (first line) and zone of elevated risk where the point estimate crossed 1.0 for ≥30 meters consecutively (second line). RDT, rapid diagnostic test.

the high population density and inadequate sanitation, close-contact, fecal–oral transmission is amplified, producing recurrent clustering. Preventive measures, including piped water and vaccination, could be reinforced in cluster locations. CATI could address containment for new cases in less affected areas to prevent larger outbreaks. Because lakeside cities like Uvira may regularly seed regional outbreaks, targeted disease control strategies may bring substantial public health benefits.

Acknowledgments

We thank the Uvira Health Zone and cholera treatment center/unit (CTC/CTU) collaborators for the support provided to testing and data collection, often under difficult circumstances. We thank John Giles for advice on implementing the IDSpatialStats package. Last, we thank the patients who participated in the main study.

All data and code produced are available online at https://github.com/ruwanepi/Uvira_spatiotemporal.

R.R. is funded by a Doctoral Foreign Study Award from the Canadian Institutes of Health Research (award no. DFS-164266). The trial on which this study sources its data was co-funded by the French Agency for Development (ref. no. EVA/364-2015) and the Veolia Foundation (ref. no. 13/14 HD 1123).

Ethics approval was provided by the London School of Hygiene and Tropical Medicine (#10603-5) and the University of Kinshasa School of Public Health (#ESP/CE/173B/2022) as an amendment to the primary study for which the cholera case data was collected (*9*).

About the Author

Dr. Ratnayake is an infectious disease epidemiologist with a background in public health in humanitarian crises.

This analysis was part of his PhD at the London School of Hygiene and Tropical Medicine on the spatial modelling and field evaluation of case-area targeted interventions for cholera outbreaks. His research links field epidemiology and mathematical modelling to improve public health responses for vulnerable and crisis-affected populations.

References

1. World Health Organization. Multi-country outbreak of cholera: external situation report #11. Geneva, Switzerland; February 12, 2024. [cited 2024 July 2] https://www.who.int/publications/m/item/multi-country-outbreak-of-cholera--external-situation-report--11---12-february-2024
2. Azman AS, Luquero FJ, Salje H, Mbaïbardoum NN, Adalbert N, Ali M, et al. Micro-hotspots of risk in urban cholera epidemics. J Infect Dis. 2018;218:1164–8. https://doi.org/10.1093/infdis/jiy283
3. Debes AK, Ali M, Azman AS, Yunus M, Sack DA. Cholera cases cluster in time and space in Matlab, Bangladesh: implications for targeted preventive interventions. Int J Epidemiol. 2016;45:2134–9. https://doi.org/10.1093/ije/dyw267
4. Ratnayake R, Finger F, Azman AS, Lantagne D, Funk S, Edmunds WJ, et al. Highly targeted spatiotemporal interventions against cholera epidemics, 2000-19: a scoping review. Lancet Infect Dis. 2021;21:e37–48. https://doi.org/10.1016/S1473-3099(20)30479-5
5. Bulit G, Ramos M. Response to cholera outbreaks: case-area targeted interventions (CATI) and community outbreak response teams (CORT) guidelines. New York: UNICEF; 2020. [cited 2024 July 2] https://www.washcluster.net/node/30201
6. Ouamba JP, Fouda Mbarga N, Ciglenecki I, Ratnayake R, Tchiasso D, Finger F, et al. Implementation of targeted cholera response activities, Cameroon. Bull World Health Organ. 2023;101:170–8. https://doi.org/10.2471/BLT.22.288885
7. Bompangue D, Moore S, Taty N, Impouma B, Sudre B, Manda R, et al. Description of the targeted water supply and hygiene response strategy implemented during the cholera outbreak of 2017–2018 in Kinshasa, DRC. BMC Infect Dis. 2020;20:226. https://doi.org/10.1186/s12879-020-4916-0
8. Ingelbeen B, Hendrickx D, Miwanda B, van der Sande MAB, Mossoko M, Vochten H, et al. Recurrent cholera outbreaks, Democratic Republic of the Congo, 2008–2017. Emerg Infect Dis. 2019;25:856–64. https://doi.org/10.3201/eid2505.181141
9. Gallandat K, Macdougall A, Jeandron A, Mufitini Saidi J, Bashige Rumedeka B, Malembaka EB, et al. Improved water supply infrastructure to reduce acute diarrhoeal diseases and cholera in Uvira, Democratic Republic of the Congo: results and lessons learned from a pragmatic trial. PLoS Negl Trop Dis. 2024;18:e0012265. https://doi.org/10.1371/journal.pntd.0012265
10. Malembaka EB, Bugeme PM, Hutchins C, Xu H, Hulse JD, Demby MN, et al. Effectiveness of one dose of killed oral cholera vaccine in an endemic community in the Democratic Republic of the Congo: a matched case-control study. Lancet Infect Dis. 2024;24:514–22. https://doi.org/10.1016/S1473-3099(23)00742-9
11. Kulldorff M, Heffernan R, Hartman J, Assunção R, Mostashari F. A space-time permutation scan statistic for disease outbreak detection. PLoS Med. 2005;2:e59. https://doi.org/10.1371/journal.pmed.0020059
12. Lessler J, Salje H, Grabowski MK, Cummings DA. Measuring spatial dependence for infectious disease epidemiology. PLoS One. 2016;11:e0155249. https://doi.org/10.1371/journal.pone.0155249
13. Ratnayake R, Peyraud N, Ciglenecki I, Gignoux E, Lightowler M, Azman AS, et al.; Epicentre and MSF CATI Working Group. Effectiveness of case-area targeted interventions including vaccination on the control of epidemic cholera: protocol for a prospective observational study. BMJ Open. 2022;12:e061206. https://doi.org/10.1136/bmjopen-2022-061206
14. Maluda MCM, Johnson E, Robinson F, Jikal M, Fong SY, Saffree MJ, et al. The incidence, and spatial trends of cholera in Sabah over 15 years: Repeated outbreaks in coastal areas. PLOS Glob Public Health. 2024;4:e0002861. https://doi.org/10.1371/journal.pgph.0002861
15. Gaiffe M, Dross C, Bwenge Malembaka E, Ross I, Cumming O, Gallandat K. A fuzzy inference-based index for piped water supply service quality in a complex, low-income urban setting. Water Res. 2023;243:120316. https://doi.org/10.1016/j.watres.2023.120316

Address for correspondence: Ruwan Ratnayake, London School of Hygiene & Tropical Medicine, Keppel St, London WC1E 7HT, UK; email: ruwan.ratnayake@lshtm.ac.uk

Article DOI: http://doi.org/10.3201/eid3008.231137

Spatiotemporal Modeling of Cholera, Uvira, Democratic Republic of the Congo, 2016–2020

Appendix

Methods for Local and Global Clustering Statistics

Local Clustering to Identify Recurrent Locations and Timing of Seasonal Outbreaks

We used the space-time scan statistic to retrospectively detect the presence and location of spatiotemporal clusters. We conducted the analysis for the entire period (2016–2020) and according to each year. A relative risk (RR) compares the observed versus expected number of cases inside and outside of a cluster. Poisson distribution of the cases per avenue (or street) was assumed. To find the most likely cluster, candidate clusters were ordered according to a log-likelihood ratio (LLR), where the cluster with the largest LLR is the least likely to be caused by chance and, therefore, is the most likely cluster. The significance of each cluster was evaluated by using Monte Carlo simulation to compare the original dataset with 999 random replicates produced under the null hypothesis.

We examined the entire dataset (i.e., a retrospective scan). We restricted the temporal and spatial windows to capture brief periods (7–60 days) and a radius that included ≤10% of the population at risk. To capture clustering that persisted across years, we also used a longer temporal window (7–365 days) for 2016–2020.

To explore whether the space-time scan statistic produced signals that preceded outbreaks, we conducted prospective scans of each of the clusters that were detected retrospectively. This was done to detect the earliest warning sign that indicated when that cluster would have first been detected. We simulated repeated prospective scans on the retrospective cluster start day and each successive day (up to 4 weeks later). We calculated the median and

interquartile range for the delay between when the prospective scan would have first detected the cluster and the date produced by the retrospective scan, which used more case data. We also calculated the median and interquartile range for the cluster size at first detection. We visualized the timing of the first day of each retrospective cluster on an epidemic curve. To explore when cholera transmission predominated, we calculated the proportion of years that the avenue was included in any cluster during 2016–2020, ranging from 0 (not included in any cluster) to 5 (included in a cluster every year) (*1*).

Methods for Space-Time Scan Statistic

For a given cylinder consisting of a radius centered on an avenue centroid and height of the temporal window of interest, c is the observed number of cases inside the cylinder, $E[c]$ is the expected number of cases for any given cylinder, and C is the total number of cases in Uvira (*2*). RR is calculated as:

$$RR = \frac{\frac{c}{E[c]}}{\frac{(C - c)}{(C - E[c])}}$$

During the scan, a circular scanning window with varying radii and duration moves over the geographic area so that each avenue centroid is at the center of several candidate clusters with different radii and heights. At each cylinder location, the number of cases inside the cylinder is compared with the expected number under a null hypothesis of no clustering (i.e., cases are randomly distributed). To find the most likely cluster, candidate clusters are ordered by the LLR and evaluated by using Monte Carlo simulation as previously described.

Global Clustering to Inform Risk Boundaries

We estimated the tau (τ) statistic for the entire period (2016—2020) and annually to quantify the spatial extent of the risk zone around an index case (*3*). Because the dataset only contained the date of the visit to the cholera treatment center/cholera treatment unit as opposed to the date of symptom onset, this statistic represented the risk of developing medically attended disease, which we assumed indicated severe dehydration and diarrhea compared with mild dehydration and diarrhea. This approach defines clustering according to how likely any pair of cases are potentially transmission-related within a given distance between the cases. Accordingly, we first classified each pair of cases as potentially transmission-related if their

dates of case presentation were within 0–4 days of each other (≈1 serial interval) (*4*). τ is the RR that a person in the population within a given distance (d_1, d_2) band (e.g., 100 m, 150 m) from an incident case becomes a potentially transmission-related case compared with the risk for any person in the population becoming a potentially transmission-related case. A τ value >1 indicates evidence of clustering within the given distance band.

As we lacked individual household locations for cases, τ reflects the spatial scale of the avenues. We estimated τ with a moving window of 50 m computed every 10 m at distances starting at 420 m (because 5% of inter-avenue centroids fell below this value) to 2,500 m (the approximate width of Uvira). We calculated 95% CIs by using the 2.5th and 97.5th quantiles from 1,000 bootstrap replicates. We evaluated τ over a 5-day window, which included the date of case presentation, and a 4-day window, which excluded the date of case presentation, to provide a more realistic response on day 5 (*5*). To smooth the artifactual fluctuations resulting from the resolution of data and the smaller sample size of annual datasets, we calculated a moving average over the previous 10 m. We defined the high-risk zone around incident cases as the radius up to which the moving average's lower 95% CIs crossed 1.0 for ≥30 consecutive meters. We defined the elevated-risk zone around incident cases as the radius up to which the moving average point estimate crossed 1.0 for ≥30 consecutive meters. To explore the potential bias from using centroids compared with household locations, we conducted a simulation study where we randomly assigned household locations within each case-patient's avenue and then estimated τ by using a lower distance range (75–2,500 m).

Methods for τ Statistic

$\hat{\tau}(d_1, d_2)$ as an RR is approximated by dividing the odds that cases within the band are transmission-related $\hat{\theta}(d_1, d_2)$ by the same odds among cases in the general population (*3*,*5*,*6*), regardless of distance $\hat{\theta}(0, \infty)$.

The τ equation is: $\hat{\tau}(d_1, d_2) = \frac{\hat{\theta}(d_1, d_2)}{\hat{\theta}(0, \infty)}$

The odds for numerator $\hat{\theta}(d_1, d_2)$ are calculated as: $\hat{\theta}(d_1, d_2) = \frac{\sum i \sum j * I_1(i,j)}{\sum i \sum j * I_2(i,j)}$

The numerator tallies the number of case pairs (i, j) within the given distance band that are transmission-related (within 0–4 days), using indicator variable $I_1(i, j)$ = 1 for notation. The

denominator tallies the number of case pairs (i, j) within the given distance band that are not transmission-related (occurring after 4 days), using indicator variable $I_2(i,j) = 1$ for notation. The equivalent odds $\hat{\theta}(0,\infty)$ is estimated for the entire population.

Simulations to Compare Centroid-Geotagged Cases and Cases with Simulated Individual Household Locations

The case data used in this study are geocoded by X, Y coordinates, indicating 216 avenues (or streets) within the centroid belonging to a residence (Appendix Figure 1). In this simulation, we assessed whether using centroids versus simulated individual household locations affected trends in the τ statistic and to what extent.

Simulation Methods

We used the dataset of 1,493 rapid diagnostic tests (RDTs) that showed positive cholera cases from 2016–2020, displaying those results in space and time (Appendix Figure 2). The X, Y coordinates in this dataset were perturbed randomly by adding a random normal distribution that had an arbitrarily defined SD of 100. The points were plotted as maps to visually compare the spatial spread of cases between datasets 1 and 2 (Appendix Figure 3). The main τ analysis was run for each dataset. This produced the RR and 95% CI (τ statistic) of the next RDT-positive case being within a specific distance to another case compared with the risk of the case occurring anywhere else during days 0–4. A moving average was applied in distance spans of 10 m, 25 m, and 50 m to smooth fluctuations. To assess the similarity between the datasets, the moving average trend lines were evaluated visually by graphing and by comparing Pearson correlations.

Findings and Interpretation

The 2 datasets showed similar τ trends (Appendix Figure 4). Both the lower CIs of the moving average τ and the moving average τ point estimates (where τ consecutively crossed 1.0 for ≥30 consecutive meters) differed between the centroid and household datasets (Appendix Table 1). The Pearson correlation coefficients for the moving average τ point estimates were significant and nearly identical.

Overall, the centroid dataset showed a similar descending trend in risk over distance, central tendencies, and correlation coefficients compared with the simulated household dataset. The centroid dataset however showed 8.3% lower τ threshold estimate for the moving average τ

point estimate and 21.9% lower 95% CI moving average than the household simulation dataset. The simulated households compared to the centroid dataset had a higher maximum moving average τ estimate (equivalent to 2.0<RR<2.5) from 75–275 m (a distance segment that was unmeasured in the centroid dataset).

Software

Analyses were performed in R software version 4.1.2 (The R Project for Statistical Computing, https://www.r-project.org) using the rsatscan version 1.0.5 (https://github.com/Kenkleinman/rsatscan) and the IDSpatialStats version 0.3.12 (https://github.com/HopkinsIDD/IDSpatialStats) (*6*) R packages. rsatscan is used in tandem with SaTScan software version 10.0.2 (https://www.satscan.org) to calculate the space-time scan statistics.

References

1. Cleary E, Boudou M, Garvey P, Aiseadha CO, McKeown P, O'Dwyer J, et al. Spatiotemporal dynamics of sporadic Shiga toxin-producing *Escherichia coli* enteritis, Ireland, 2013–2017. Emerg Infect Dis. 2021;27:2421–33. PubMed https://doi.org/10.3201/eid2709.204021

2. Kulldorff M, Heffernan R, Hartman J, Assunção R, Mostashari F. A space-time permutation scan statistic for disease outbreak detection. PLoS Med. 2005;2:e59. PubMed https://doi.org/10.1371/journal.pmed.0020059

3. Lessler J, Salje H, Grabowski MK, Cummings DAT. Measuring spatial dependence for infectious disease epidemiology. PLoS One. 2016;11:e0155249. PubMed https://doi.org/10.1371/journal.pone.0155249

4. Azman AS, Rudolph KE, Cummings DAT, Lessler J. The incubation period of cholera: a systematic review. J Infect. 2013;66:432–8. PubMed https://doi.org/10.1016/j.jinf.2012.11.013

5. Azman AS, Luquero FJ, Salje H, Mbaïbardoum NN, Adalbert N, Ali M, et al. Micro-hotspots of risk in urban cholera epidemics. J Infect Dis. 2018;218:1164–8. PubMed https://doi.org/10.1093/infdis/jiy283

6. Giles JR, Salje H, Lessler J. The IDSpatialStats R Package: quantifying spatial dependence of infectious disease spread. R J. 2019;11:308–27. https://doi.org/10.32614/RJ-2019-043

Appendix Table 1. Differences in points where τ crosses RR = 1.0 for ≥30 consecutive meters consecutively*

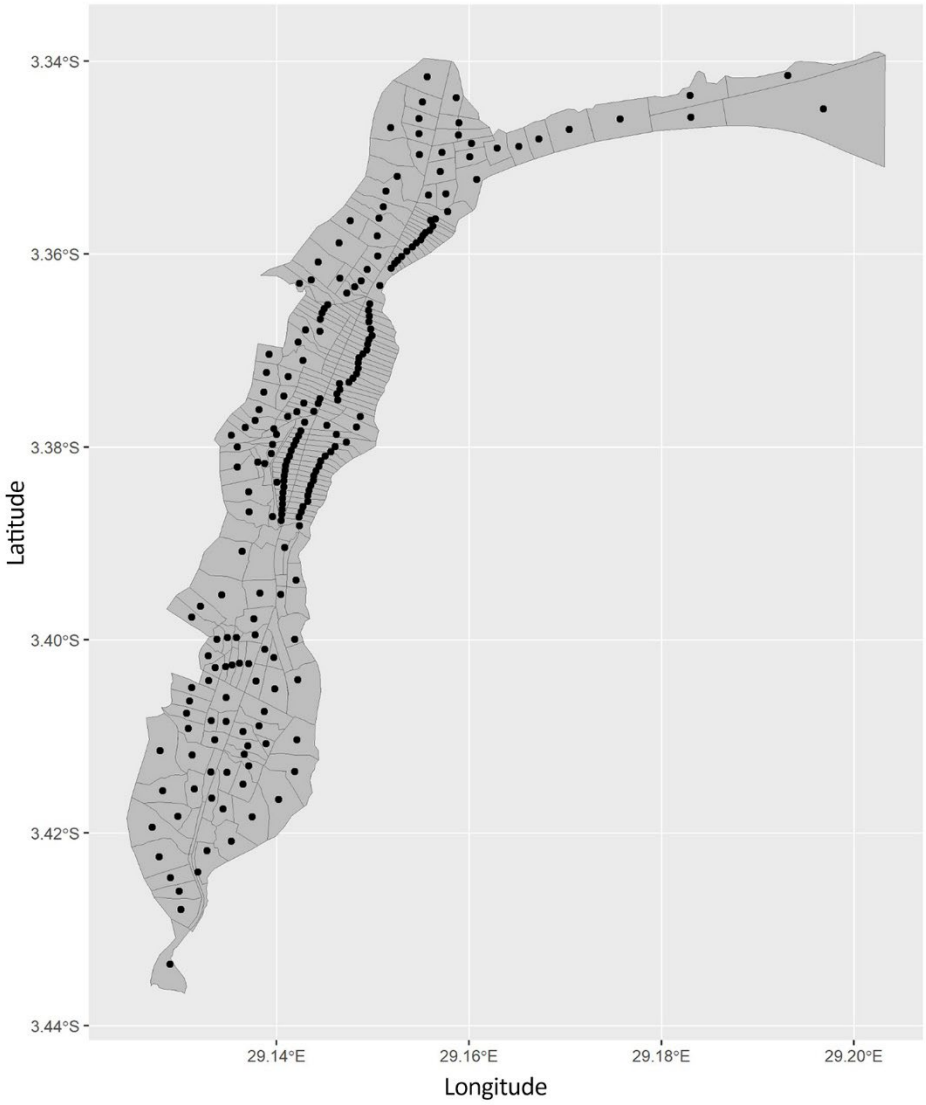
Dataset	Min τ	Max τ	Mean τ	Moving average τ<1.0 (>30m)	Moving average τ LCI <1.0 (>30m)	Pearson correlation coefficient (95% CI)
Centroid	0.52	3.01	1.01	1,665 m	1,105 m	−0.87 (−0.89 to −0.85)
Simulated household†	0.55	2.40	1.05	1,815 m	1,415 m	−0.88 (−0.90 to −0.86)

*LCI, lower confidence interval; Max, maximum; Min, minimum; τ, tau statistic.
†Household locations were simulated and compared with the centroid dataset.

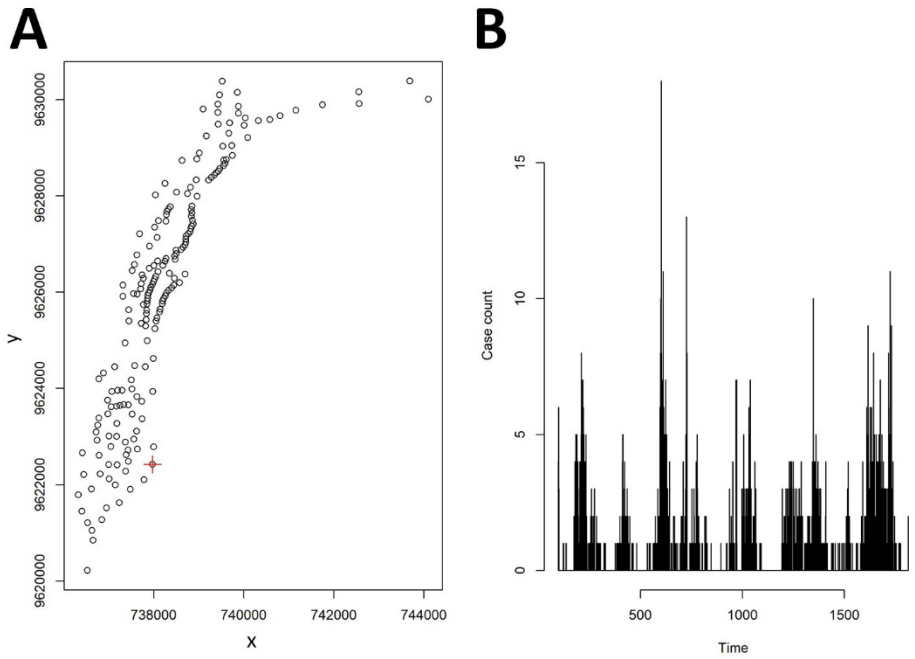
Appendix Table 2. Sensitivity analysis: statistically-significant spatiotemporal clusters of suspected cholera cases detected through annual scanning at the avenue level, Uvira, Democratic Republic of the Congo, 2016—2020

Year	Cluster No.	Cases observed: expected	Population at-risk	RR	Cluster radius (meters)	Cluster start date (mm/dd)	Cluster duration (d)
2016	1	57:5	177,122	10.8*	378	04/07	15
	2	51:4	187,076	12.1*	647	03/24	11
	3	45:6	183,225	7.2*	1,557	08/06	17
	4	27:3	120,498	8.4*	368	04/09	13
	5	40:9	147,424	4.6*	709	07/22	30
	6	18:2	29,390	7.8*	436	02/18	40
2017	1	130:13	148,014	10.8*	908	08/07	43
	2	91:16	150,104	5.9*	897	08/19	52
	3	39:6	88,959	6.6*	704	08/29	32
	4	23:2	134,147	10.6*	378	12/24	7
	5	26:5	143,948	5.2*	1,001	08/23	16
	6	9:1	42,275	17.3*	331	02/14	5
2018	1	50:3	130,673	15.3*	963	10/26	12
	2	24:2	134,311	15.1*	397	01/01	5
	3	61:15	132,515	4.2*	906	07/29	56
	4	44:10	128,631	4.5*	708	08/21	38
	5	18:3	70,142	5.9†	653	10/30	21
	6	9:1	52,203	14.4†	477	02/17	5
2019	1	50:4	93,453	14.3*	831	09/10	18
	2	30:2	21,965	13.9*	0	09/01	48
	3	47:7	105,035	7.1*	524	04/27	31
	4	48:10	115,699	5.0*	836	09/07	41
	5	36:8	120,197	4.7*	995	06/08	31
	6	14:2	40,341	7.4†	626	06/23	22
	7	6:0	45,292	32.2†	350	09/20	1
2020	1	105:17	159,204	6.7*	860	07/29	59
	2	59:11	141,671	5.8*	488	05/31	41
	3	38:5	106,256	8.6*	1,121	02/20	23
	4	57:13	155,765	4.6*	395	05/30	46
	5	49:13	120,618	3.9*	490	07/27	59
	6	39:10	159,261	4.0*	959	05/30	34
	7	15:2	44,366	10.1*	468	09/10	18

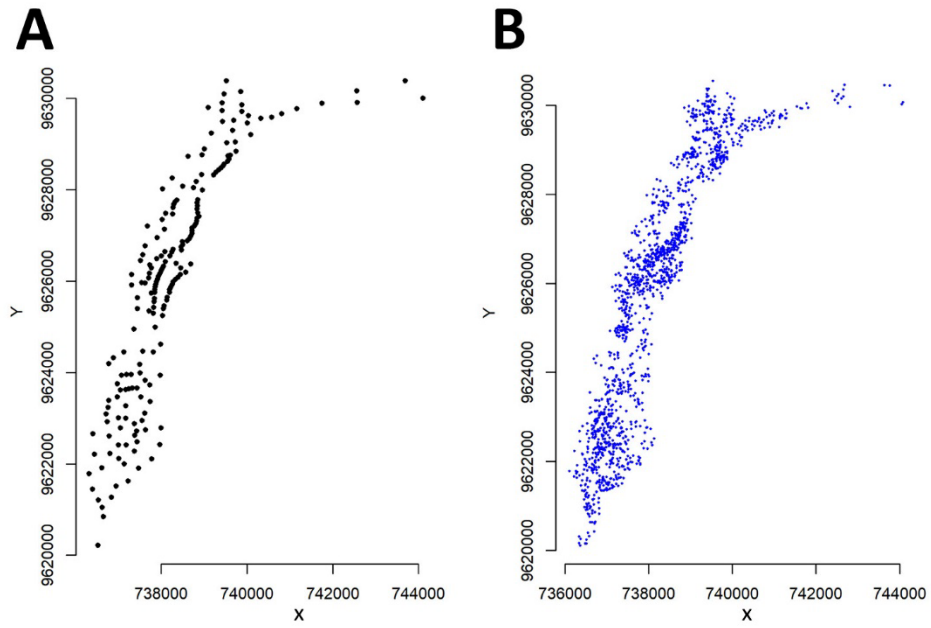
*p-value <0.001
† p-value <0.05
The p-value indicates the statistical significance of clusters derived from Monte Carlo simulations.
RR, relative risk.



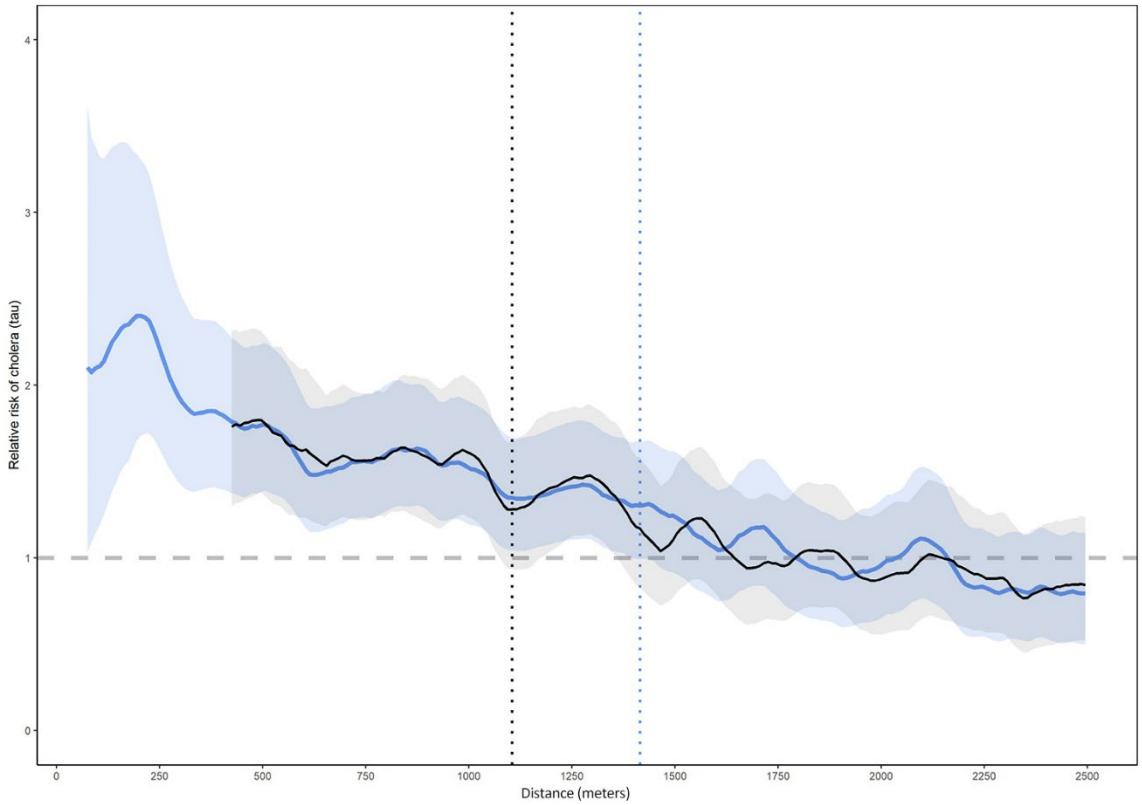
Appendix Figure 1. Information relevant to an analysis of spatiotemporal modeling of cholera, Uvira, Democratic Republic of the Congo, 2016–2020. Map of the centroid locations and borders of Uvira's 216 avenues.



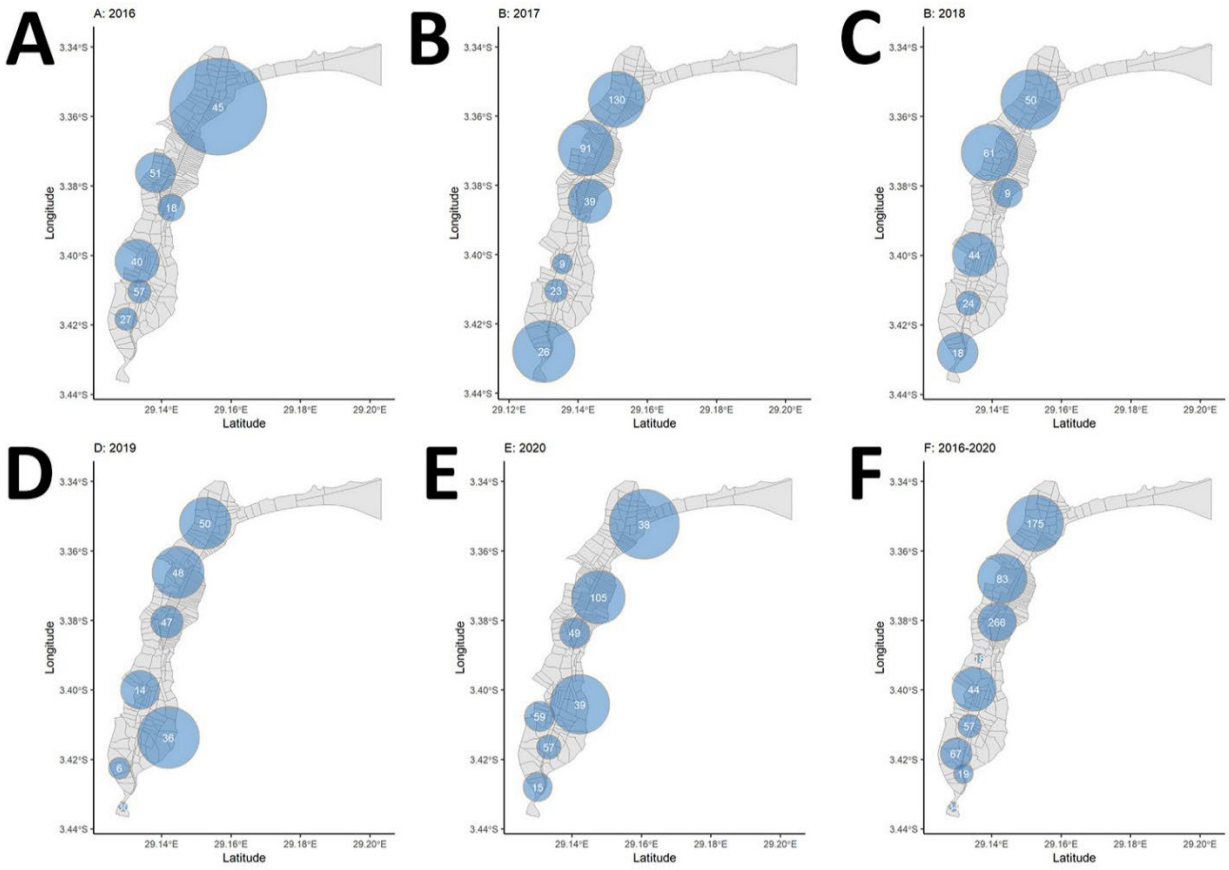
Appendix Figure 2. Information relevant to an analysis of spatiotemporal modeling of cholera, Uvira, Democratic Republic of the Congo, 2016–2020. Uvira 2016–2020 dataset of rapid diagnostic positive cases with avenue centroids of cases *(index case in red)*.



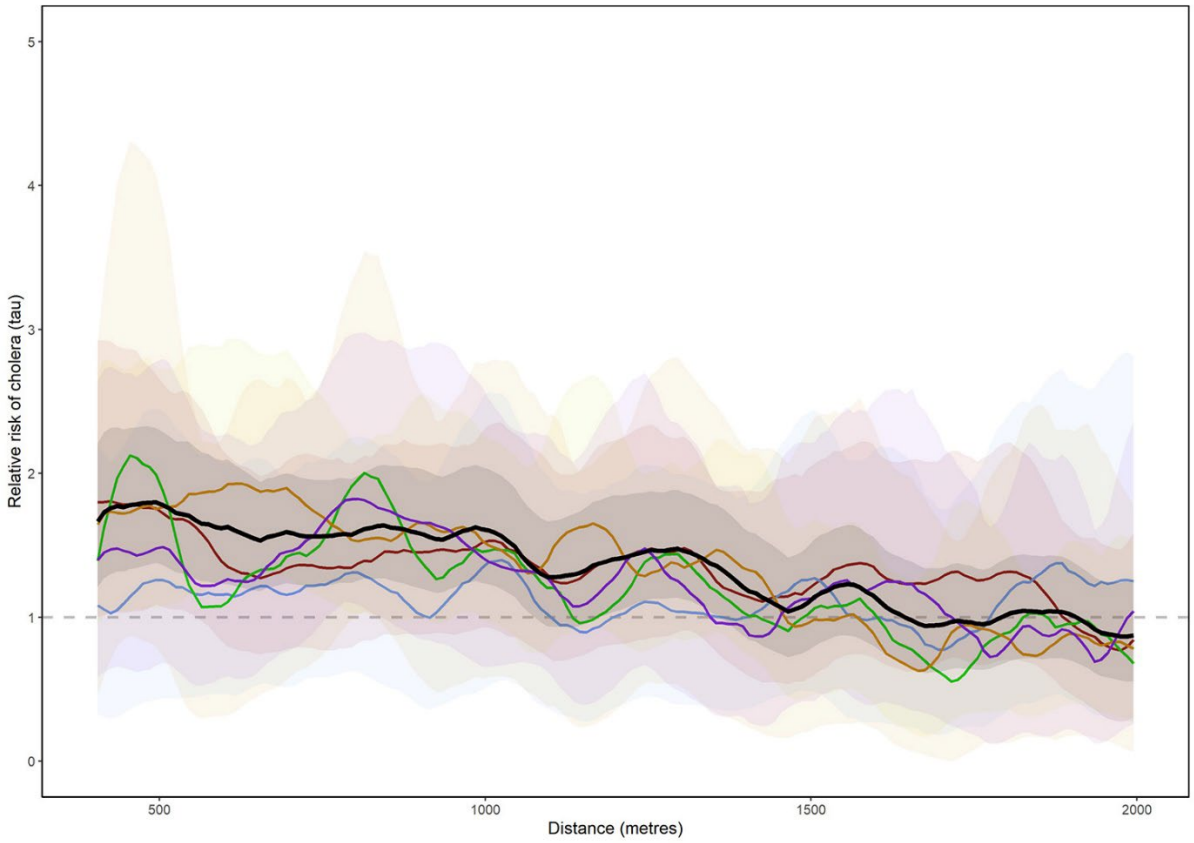
Appendix Figure 3. Information relevant to an analysis of spatiotemporal modeling of cholera, Uvira, Democratic Republic of the Congo, 2016–2020. Case centroid locations (black) and simulated household locations (blue).



Appendix Figure 4. Information relevant to an analysis of spatiotemporal modeling of cholera, Uvira, Democratic Republic of the Congo, 2016–2020. Moving average of point estimates and 95% confidence intervals for tau τ statistic for RDT-positive cholera cases (75–2500 m) of the centroids (black, starting at 420 m) and the household locations (blue, starting at 75 m). The dashed line is where the lower confidence interval for the moving average crosses 1.0 for ≥30 consecutive meters consecutively.



Appendix Figure 5. Information relevant to an analysis of spatiotemporal modeling of cholera, Uvira, Democratic Republic of the Congo, 2016–2020. Sensitivity analyses of prospectively detected spatiotemporal clusters of suspected cholera cases, 2016–2020. days. All scans had a maximum spatial window of 10% of the geographic area. The size of the orange circle depicts the radius with the number of suspected cases (in white). A–E depict scans with a temporal window of 7–60 days and F depicts a scan with a temporal window of 7–365. A) 2016; B) 2017; C) 2018; D) 2019; E) 2020; F) 2016–2020.



Appendix Figure 6. Information relevant to an analysis of spatiotemporal modeling of cholera, Uvira, Democratic Republic of the Congo, 2016–2020. Annual and aggregated moving average estimates of τ (relative risk) and 95% CIs (solid line and shading) for days 0–4. 2016–2020 in black, 2016 in purple, 2017 in orange, 2018 in green, 2019 in blue, 2020 in red.