Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: Jones FK, Wamala JF, Rumunu J, et al. Successive epidemic waves of cholera in South Sudan between 2014 and 2017: a descriptive epidemiological study. *Lancet Planet Health* 2020; 4: e577–87.
Appendix 2

Successive epidemic waves of cholera in South Sudan between 2014 and 2017: a descriptive epidemiological study

Forrest K Jones¹, Joseph F Wamala², John Rumunu³, Pinyi Nyimol Mawien³, Kol Mathew Tut³, Shirlee Wohl¹, Lul Deng³, Lorenzo Pezzoli⁴, Linda Haj Omar⁵, Justin Lessler¹, Marie-Laure Quilici⁶, Francisco J Luquero⁷, Andrew S Azman¹,⁸

¹ Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, USA
² World Health Organization, Juba, South Sudan
³ South Sudan Ministry of Health, Juba, South Sudan
⁴ World Health Organization, Geneva, Switzerland
⁵ World Health Organization, Brazzaville, Republic of Congo
⁶ Institut Pasteur, Paris, France
⁷ Epicentre, Geneva, Switzerland
⁸ MSF, Geneva, Switzerland
Table S1. Laboratory culture confirmation of Vibrio cholerae O1.

Number of samples testing positive for cholera culture from the National Public Health Laboratory in Juba. Three samples in 2017 tested had an unknown geographical state; all were negative.

| State                  | 2014       | 2015       | 2016       | 2017       |
|------------------------|------------|------------|------------|------------|
| Central Equatoria      | 133/312 (42.6%) | 39/104 (37.5%) | 71/200 (35.5%) | 133/238 (55.9%) |
| Eastern Equatoria      | 39/67 (58.2%)  | 0/6 (0%)   | 7/9 (77.8%) | 14/23 (60.9%) |
| Jonglei                | 1/3 (33.3%)   | 4/17 (23.5%) | 20/59 (33.9%) | 20/40 (50%) |
| Lakes                  | -           | 0/1 (0%)   | 12/46 (26.1%) | 33/57 (57.9%) |
| Unity                  | -           | 0/3 (0%)   | 29/76 (38.2%) | 20/60 (33.3%) |
| Upper Nile             | 16/43 (37.2%) | -           | 0/5 (0%)    | 7/24 (29.2%) |
| Warrap                 | -           | 0/1 (0%)   | 0/1 (0%)    | 4/8 (50%)   |
| Western Bahr el Ghazal | -           | -          | 0/6 (0%)    | -           |
| Western Equatoria      | 1/5 (20%)   | -          | -           | -           |
| South Sudan            | 190/430 (44.2%) | 43/132 (32.6%) | 139/402 (34.6%) | 231/453 (51%) |
Figure S1: Timing of the first round of vaccination campaigns.

Each campaign is represented by a dot. The length of the line is the amount of time between either the peak or last case of a county-level outbreak. Ten campaigns occurred in areas without cholera cases. One campaign in Juba 2015 had a missing date and was excluded from this analysis.
Figure S2. Maximum likelihood tree of publicly available VC whole core genome sequences.

T10 and T13 lineages, as defined in Weill et al. 2019, are indicated in black; sub-lineage of T10 containing South Sudan isolates (see Fig. 2) indicated in blue; 2014–2017 isolates from South Sudan indicated in orange.
Figure S3. MLVA and antibiotic sensitivity analyses of V. cholerae isolates from South Sudan

Each row of the heatmaps represents one isolate. The left heatmap shows MLVA data with each column being a single locus. Color represents the number of copies that differ from a reference sequence 10-7-3-7-10-18. The right heatmap shows sensitivity for each sample with each column representing a different antibiotic (sulfonamides, SUL; trimethoprim-sulfamethoxazole, SXT; chloramphenicol, CHL; nitrofurantoin, NIT; erythromycin, ERY; Polymyxin B, PMB; colistin, CST; tetracycline, TET; nalidixic acid, NAL; norflaxacin, NOR; vibriostatic agent, O129; streptomycin, STR). Based on Comité de l’Antibiogramme de la Société Française de Microbiologie 2013 standards for Enterobacteriaceae, nearly all samples were resistant to colistin, nitrofurantoin, polymyxin B, streptomycin, sulfonamides, trimethoprim-sulfamethoxazole and vibriostatic agent O/129 (associated with resistance to trimethoprim) and showed decreased susceptibility to chloramphenicol (Figure S3B); MICs determination with Etest (AB bioMérieux, Solna, Sweden) confirmed resistance of all strains to nalidixic acid as confirmed by detection of one mutation in the gyrA gene (substitution of serine by isoleucine at position 83). Resistance to streptomycin (strAB), sulfonamides (sul2), trimethoprim and O139 vibriostatic agent (dfrA1), and trimethoprim-sulfamethoxazole (sul2 and dfrA1), were reported to be associated with an integrating conjugative element called ICEVchInd5.
**Figure S4. Temporal association between the start of county-level outbreaks and rainy season, South Sudan, 2014 - 2017**

Each point represents a county-level outbreak during a specific year. Color indicates if the outbreak occurred in Juba County (red) or in another county (blue). Solid line represents $y=x$ while the dashed lines are offset by 5 weeks.
Equation S1. Distributed lag non-linear model equations

Our approach follows the statistical methods outlined in an analysis of cholera surveillance data from Yemen [1]. We assume that the number of cases reported in Juba, \( I(t) \), on a given day, \( t \), follows a Poisson process

\[
I(t) \sim \text{Poisson}(R(t) \sum_{s=0}^{10} I(t-s)w(s))
\]

where \( R(t) \) is the instantaneous reproductive number and \( w(s) \) is the probability mass function of the serial interval, assuming a shifted (by one day) gamma distributed serial interval with a mean of 3.7 days, a standard deviation of 1.5 days, and a maximum of 10 days [2].

To investigate whether rainfall may have modulated cholera transmission, we accounted for population-level immunity, by estimating the instantaneous basic reproductive number, \( R_0(t) \). Specifically, we estimated the proportion of the population that is susceptible, \( S(t) \), as follows:

\[
R(t) = R_0(t) \frac{S(t)}{N}
\]

\[
S(t) = N - \sum_{j=1}^{t-1} I(j)/p
\]

where \( p \) is the detection probability, assumed to be 10% [3]. Using this set up, we can write out a quasi-Poisson regression model equation:

\[
\log(\mathbb{E}[I(t)]) = \log(R_0(t)) + \log(\sum_{s=1}^{10} I(t-s)w(s)) + \log(N - \sum_{j=1}^{t-1} I(j)/p)
\]

By analyzing the serial interval and susceptibility terms as offset terms, we regressed the instantaneous basic reproductive number on covariates of interest.

\[
\log(R_0(t)) = \beta_0 + \sum d=1^6 \beta_{dow,a} \delta_d(t) + \sum k=1^3 \beta_{wave,k} \delta_k(t) + s(x(t-1),\ldots,x(t-10))
\]

\( \beta_0 + \sum d=1^6 \beta_{dow,a} \delta_d(t) \) Specifically, we controlled for the day of the week for potential reporting biases through the week.

\( \sum k=1^3 \beta_{wave,k} \delta_k(t) \) We also included a random effect for the wave (Wave 1: 2014, Wave 2: 2015 or Wave 3: 2016/2017) of the epidemic.

Lastly, we included a cross-basis function, \( s \), to examine the lagged effects of rainfall using a distributed lag non-linear model (DLNM)
\[ s(x(t-1), \ldots, x(t-10)) = \sum_{l=1}^{10} f \cdot \omega(x(t-l), l) \]

where \( x(t) \) represents the accumulated rainfall over the previous 7 days (AR7D). A maximum lag of 10 days was used, meaning that the AR7D on the previous 10 days was used in the analysis of the instantaneous basic reproductive number on a specific day \( t \). The cross-basis is a sum of bi-dimensional dose-lag-response functions, \( f \cdot \omega(x, l) \) each composed of two marginal functions; \( f(x) \), the standard rainfall response and \( \omega(l) \), the additional lag response function. The marginal functions were both modeled using penalized splines, each with five degrees of freedom. Lastly, we calculated the overall rainfall-response association by summing the risk during the 10-day lag period and centering it to a reference value (AR7D = 0). We fit a generalized additive mixed model (GAMM) using quasi-poisson regression.
Figure S5. Serial interval and instantaneous basic reproductive number estimates

Results from bayesian analysis to jointly estimate the serial interval and reproductive are illustrated. Discretized serial intervals used in the analysis are shown and have been normalized to been limited to be over a 10 day period. Estimates of the instantaneous basic reproductive number (estimated using the serial interval from the 2015 cases data) are shown over time.
Figure S6. Analysis of rainfall and cholera transmission using a distributed lag non-linear model

Histogram indicates the number of days with AR7D during which there was cholera was reported. Solid lines are the mean estimate of the relative instantaneous basic reproductive number while shading is the 95% confidence interval.
Table S2. Number of arrivals recorded by the International Office of Migration by country of origin

| Country of Origin | 2014 | 2015 | 2016 | 2017 | Total |
|-------------------|------|------|------|------|-------|
| DRC               | 0    | 1091 | 658  | 2904 | 4653  |
| Ethiopia          | 0    | 559  | 357  | 2109 | 3025  |
| Kenya             | 0    | 954  | 3291 | 12631| 16876 |
| South Sudan       | 164119 | 96291 | 121179 | 172100 | 553689 |
| Sudan             | 0    | 27785 | 23779 | 35611 | 87175 |
| Uganda            | 0    | 28160 | 17600 | 49610 | 95370 |
Figure S7. Heatmap of domestic population movement between states of South Sudan.

Color is log transformed. If no block exists, no movement was recorded for that particular origin-assessment pair.
Figure S8. International cross-border movement of individuals returning to South Sudan (A) and Incidence rates of cholera in neighboring countries (B)

Geographic location is not known below the country level. Incidence rates calculated using WHO reports and population number from (https://population.un.org/wpp/DataQuery/). Though Sudan is believed to be endemic for cholera, but has not historically reported it to the WHO. Grey areas were not assessed by the IOM for migration.
Figure S9. Analysis of cholera cases in the counties of South Sudan, Jan 2014 - Dec 2018

Incidence and the proportion of weeks with cholera cases are shown for each cholera affected county. Both axes are square-root transformed.
Figure S10. Monthly geographic distribution of suspected cholera cases in South Sudan, June 2016 - November 2017.

County borders are indicated in black. Light blue line represents the Nile river.
References

1. Camacho A, Bouhenia M, Alyusfi R, Alkohlani A, Naji MA, de Radiguès X, Abubakar AM, Almoalmi A, Seguin C, Sagrado MJ, Poncin M. Cholera epidemic in Yemen, 2016–18: an analysis of surveillance data. The Lancet Global Health. 2018 Jun 1;6(6):e680-90.

2. Phelps M, Perner ML, Pitzer VE, Andreasen V, Jensen PK, Simonsen L. Cholera epidemics of the past offer new insights into an old enemy. The Journal of infectious diseases. 2018 Jan 30;217(4):641-9.

3. King AA, Ionides EL, Pascual M, Bouma MJ. Inapparent infections and cholera dynamics. Nature. 2008 Aug;454(7206):877-80.