# Exploring relationships between drought and epidemic cholera in Africa using generalised linear models

Gina E. C. Charnley[1,2*], Ilan Kelman[3,4,5], Nathan Green[6], Wes Hinsley[1,2], Katy A. M. Gaythorpe[1,2] and Kris A. Murray[1,2,7]

## Abstract

**Background:** Temperature and precipitation are known to affect *Vibrio cholerae* outbreaks. Despite this, the impact of drought on outbreaks has been largely understudied. Africa is both drought and cholera prone and more research is needed in Africa to understand cholera dynamics in relation to drought.

**Methods:** Here, we analyse a range of environmental and socioeconomic covariates and fit generalised linear models to publicly available national data, to test for associations with several indices of drought and make cholera outbreak projections to 2070 under three scenarios of global change, reflecting varying trajectories of $CO_2$ emissions, socio-economic development, and population growth.

**Results:** The best-fit model implies that drought is a significant risk factor for African cholera outbreaks, alongside positive effects of population, temperature and poverty and a negative effect of freshwater withdrawal. The projections show that following stringent emissions pathways and expanding sustainable development may reduce cholera outbreak occurrence in Africa, although these changes were spatially heterogeneous.

**Conclusions:** Despite an effect of drought in explaining recent cholera outbreaks, future projections highlighted the potential for sustainable development gains to offset drought-related impacts on cholera risk. Future work should build on this research investigating the impacts of drought on cholera on a finer spatial scale and potential non-linear relationships, especially in high-burden countries which saw little cholera change in the scenario analysis.

**Keywords:** Public health, Epidemiology, Cholera, Disease outbreaks, *Vibrio cholerae*, Droughts, Climate change, Africa

## Background

*Vibrio cholerae* is a water-borne bacterial pathogen, causing profuse watery diarrhoea and rapid dehydration in symptomatic cases. This can lead to death within 2 h of symptom onset and case fatality ranging from 3 to 40% [1, 2]. The seventh and ongoing cholera pandemic began in 1961, spreading to Africa by the 1970s, where it now shows signs of endemicity in several countries [3, 4]. Despite over 94% of World Health Organization (WHO) reported cholera cases occurring in Africa and some of the highest mortality rates [5], previous research has heavily focused on South America, the Indian subcontinent and more recently Haiti.

Cholera outbreak frequency is closely related to environmental and climatic changes [6–8]. For instance, temperature and precipitation are considered important in cholera outbreak occurrence, with temperature driving epidemics and precipitation acting as a dispersal mechanism [9]. These relationships have

*Correspondence: g.charnley19@imperial.ac.uk
[1] School of Public Health, Imperial College London, Norfolk Place, London W2 1PG, UK
Full list of author information is available at the end of the article

Charnley *et al. BMC Infectious Diseases*     (2021) 21:1177

Page 2 of 12

implications for cholera outbreaks after natural hazards, such as droughts. Several links between drought and cholera outbreaks have been described [2, 10, 11] and it is hypothesised that increasing concentrations of infectious bacteria in more limited drinking water sources and increased risky drinking water behaviours are likely mechanisms for transmission [2, 12]. Despite this, drought and cholera in Africa are understudied in isolation and links have more commonly been made between flooding, despite droughts potentially posing a considerably greater risk than floods [11].

Cholera is considered a disease of inequity and several socio-economic risk factors have been implicated with cholera outbreaks, which may be further exacerbated by droughts. Some studies suggest that human-induced factors are more important for cholera dynamics than climate or environmental ones [13], including poverty [14], sanitation [15], drainage [16], water quality [17] and poor healthcare [9]. This supports the notion that outbreaks result from the breakdown of societal systems responses to a hazard, leading to a human–environment link and subsequent pathogen shedding [18]. Water, sanitation and hygiene (WASH) factors are considered particularly significant, as the importance of the water body reservoirs depends on the sanitary conditions of the community [19]. Eight

hundred and forty-four million people worldwide lack access to basic drinking water and a further 2.4 billion are without basic sanitation [20], putting many people at risk of water-borne disease outbreaks including cholera.

Here, we aim to address the research gap of drought-related health outcomes by investigating its implications on cholera. The work fills an important research gap as few studies have investigated the link between drought and cholera outbreaks in Africa, or projected outbreak changes into the future, and investigating mechanisms through which global change might yield health impacts. Research in this area is particularly important due to a significant number of people at risk of both cholera and drought and the negative implications that climate change may have for these communities.

## Methods

In this study, we aimed to understand the implications of drought for cholera outbreak occurrence at a continental scale across Africa, after accounting for important socio-economic factors. We aim to use these results to further understand the hypothesis that droughts lead to cholera outbreaks through elevated pathogen concentrations in limited water and an increase in risky drinking water behaviours, Fig. 1 shows a schematic to help visualise this
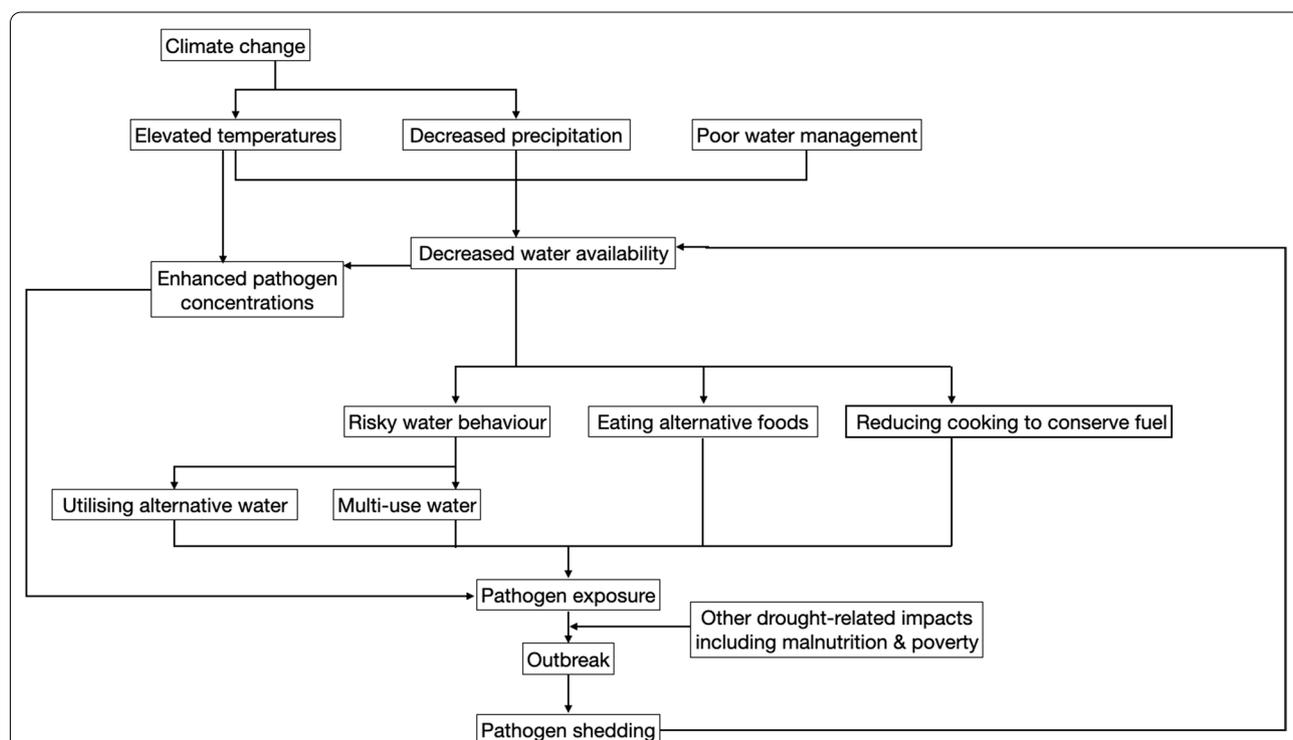


**Fig. 1** Pathways from water shortages to cholera outbreaks: suggested mechanism through which drought can lead to cholera outbreaks in Africa [2, 12]

Charnley *et al. BMC Infectious Diseases*    (2021) 21:1177

Page 3 of 12

hypothesis and potential pathways. In addition, we aimed to evaluate how future changes in drought area and risk due to climate change [21, 22], alongside other development factors, may impact future cholera outbreak occurrence. We thus developed several projection scenarios incorporating different greenhouse gas emissions and socio-economic development trajectories.

### Datasets and study period

We compiled data on cholera outbreaks and a range of social and environmental covariates over the period 1970–2019. Annual cholera cases were retrieved from the WHO's Global Health Observatory [23], which provides reported annual cholera case for each country, which were confirmed either clinically, epidemiologically, or by laboratory investigation. For analysis, these numbers were transformed into a binary outcome to reflect outbreak occurrence [i.e., set at 0 for no outbreak and 1 for an outbreak (>1 case/death)], which was then used as the outcome variable in the models. We opted not to analyse raw case data to minimise the effect of unmeasured observations and reporting biases among countries. For years with no outbreak data, the outcome was set to 0, assuming if cholera cases/deaths occurred within a country then they would have been identified and reported (a sensitivity analysis for this assumption is presented in Additional file 1: Information 1).

In total, 19 environmental and socio-economic covariates were selected for investigation based on prior hypotheses and previous results linking cholera outbreaks to risk factors (summarised in Additional file 1: Table S1). Environmental data were extracted from a variety of sources and included climate (temperature and precipitation) [24], meteorological drought (Palmer Drought Severity Index, PDSI) [25], agricultural drought (soil moisture and potential evapotranspiration (PET) [26, 27] and hydrological drought (runoff and freshwater withdrawal annually and per capita) [28, 29]. Where monthly or sub-national data were available, we calculated national yearly means. Climate data were missing for Côte d'Ivoire and drought data were missing for Rwanda, The Gambia, Guinea-Bissau, Djibouti, Burundi, Benin, Cabo Verde, São Tomé and Principe, Comoros, Mauritius and Seychelles. Environmental data for these countries were derived by taking the mean of their neighbouring countries, whereas islands were excluded.

Socio-economic data including annual indicators of poverty and development, WASH, malnourishment, and population (on a logarithmic scale), were taken from the WorldBank [30] and the United Nations Development Programme [31] datasets. Where a country's socio-economic data were missing for some years, a national average was taken from the available data points and used for all years. If national data were missing for the full instrumental period, these countries were removed from the analysis.

After examining data completeness across the full dataset, we designated the instrumental period for analysis to be 2000–2016 to limit omitting missing data and interpolation. Summary figures of the climate and cholera data over the instrumental period are shown in Additional file 1: Fig. S1. Summary figures of the drought indices and their definitions are shown in Additional file 1: Fig. S2 and Information 2.

### Model structure and fitting

Generalised linear models (GLM) were fitted to the dataset describing the cholera outbreak occurrence for the instrumental period (2000–2016), for all countries in mainland Africa and Madagascar, using maximum likelihood estimation. Due to the binary outcome variable for cholera outbreak occurrence, a binomial likelihood with a log–log link function was used in all models. Rows with missing values were removed from the data frame.

From this initial dataset, a reduced pool of potential covariates was selected for model fitting using a covariate selection process developed by Garske et al. [32] and Gaythorpe et al. [33]. In summary, univariate models for each potential variable were fitted to the binary outcome variable and any variables not significantly associated with the outcome at a 10% confidence limit ($p < 0.1$) were excluded. Of the remaining covariates, absolute pairwise correlations were calculated, and highly correlated variables ($r > 0.75$) were then clustered into groups, to prevent multilinearity. Parametric correlations were used here, as the data followed normal distributions. Parametric correlations have more assumptions and therefore have more statistical power, meaning they are more likely to detect a significant difference when one truly exists (a correlation matrix and correlation plots are shown in Additional file 1: Table S2 and Fig. S3). The covariates from each cluster most strongly correlated with the outcome variable was then selected for inclusion in the multivariate models, fit using the function glm. Model fit was evaluated using Bayesian Information Criterion (BIC) and a single best-fit model was found using the stepAIC function. In addition, area under the receiver operator characteristic curve (AUC) was used to quantify model performance. All statistical analyses were carried out in R Studio version 3.6.2 (packages: tidyr, MASS, ggplot2, dplyr, magrittr, corrplot, caret, nlme, MuMIn, car, boot).

### Testing for temporal and spatial effects

The inclusion of multiple years of data across multiple countries raises the possibility of spatial and temporal confounding (e.g., autocorrelation). To investigate the

Charnley *et al. BMC Infectious Diseases*    (2021) 21:1177

Page 4 of 12

potential influence on the covariate selection and subsequent model, separate analyses were run including year and ISO3 country code as predictor variables following the same step-wise covariate selection process and multivariate model approach as described above. Autocorrelation diagnostics were run on selected spatial and temporal covariates by testing the significance of the linear relationship with and without consideration of AR1 (autoregressive model of order 1) autocorrelation and assessing evidence of autocorrelation in the residuals. Leave-one-out (LOO) cross validation using Akaike Information Criterion (AIC) was used to assess model performance of both the original (without year/ISO) and the updated (with year/ISO) multivariate models selected through the covariate selection process.

### Projection scenarios

Three scenarios (S1, S2 and S3) were developed for 2020–2070 (at decadal increments) as summarised in Table 1. Each scenario represents an alternative possible future trajectory of the variables retained in the best fit model, parameterised to varying degrees of climate mitigation and socio-economic development. Here, S1 represents a "best-case" scenario, loosely aligning to highly ambitious climate change mitigation and strong progress towards the Sustainable Development Goals (SDG), S2 represents an intermediate scenario, and S3 a "worst-case" scenario with slower progress towards emissions reductions and the SDGs.

Detailed descriptions and justifications of the projected changes for each variable are provided in full in Additional file 1: Information 3. Briefly, projected temperature data (for 2050 and 2070) were taken from WorldClim [34], as this was also used for historical data. The data is Coupled Model Intercomparison Project 6 (CMIP6) downscaled future climate projections, processed for nine global climate models using three Representative Concentration Pathways (RCP). We used RCP4.5, 6.0 and 8.5 for scenarios S1, S2 and S3, respectively. This was projected for 2050 and 2070 and we used the instrumental period average (2000–2016) for 2020–2040 values. The average was used to account for interannual climate variability. Additional file 1: Fig. S4 summaries the data for each pathway and year.

**Table 1** Cholera projection scenarios for 2020–2070 at decadal intervals: Scenario 1 (S1), a "best-case" scenario; Scenario 2 (S2), an intermediate scenario and Scenario 3 (S3), a "worst-case" scenario. The scenarios were projected over 50 years from 2020 to 2070. HWC = high withdraw countries including MDG, LBY, SDN, MRT and MAR

| | Year | Drought | Temperature | Poverty | Water withdrawal |
|---|---|---|---|---|---|
| Scenario 1 | 2020 | 2000–2016 average | 2000–2016 average | 2016 | 2016 |
| | 2030 | | 2000–2016 average | Reduce 2016 by 50% | 2016 |
| | 2040 | | 2000–2016 average | Reduce 2016 by 50% | 2016 |
| | 2050 | | RCP4.5 2050 | Medium value between 2030 & 2070 | 20% increase and 20% decrease for HWC |
| | 2060 | | RCP4.5 2050 | Medium value between 2030 and 2070 | 20% increase and 20% decrease for HWC |
| | 2070 | | RCP4.5 2070 | Poverty elimination (0%) | 20% increase and 20% decrease for HWC |
| Scenario 2 | 2020 | Median value between S1 and S2 | 2000–2016 average | 2016 | 2016 |
| | 2030 | | 2000–2016 average | 2016 | 2016 |
| | 2040 | | 2000–2016 average | 2016 | 2016 |
| | 2050 | | RCP6.0 2050 | Reduce 2016 by 50% | 10% increase and 10% decrease for HWC |
| | 2060 | | RCP6.0 2050 | Medium value between 2050 and 2070 | 10% increase and 10% decrease for HWC |
| | 2070 | | RCP6.0 207 | Poverty elimination (0%) | 10% increase and 10% decrease for HWC |
| Scenario 3 | 2020 | [(Coefficient*4) + 2016 value] | 2000–2016 average | 2016 | 2016 |
| | 2030 | [(Coefficient*10) + 2020 value] | 2000–2016 average | 2016 | 2016 |
| | 2040 | [(Coefficient*10) + 2030 value] | 2000–2016 average | 2016 | 2016 |
| | 2050 | [(Coefficient*10) + 2040 value] | RCP8.5 2050 | 2016 | 5% increase and 5% decrease for HWC |
| | 2060 | [(Coefficient*10) + 2050 value] | RCP8.5 2050 | Medium value between 2050 and 2070 | 5% increase and 5% decrease for HWC |
| | 2070 | [(Coefficient*10) + 2060 value] | RCP8.5 2070 | Reduce 2016 value by 50% | 5% increase and 5% decrease for HWC |

Charnley *et al. BMC Infectious Diseases* (2021) 21:1177

Page 5 of 12

Projecting PDSI at a continental or national scale is contentious showing a range of projection outcomes, due to high spatial heterogeneity and between model uncertainty/disagreement [21, 35], as well as computational discrepancies depending on the PET algorithm used [36]. Several PDSI modelling studies [36, 37] and paleoclimatic studies [38, 39] found that drought severity and durations remained constant despite periods of extreme dryness, over a range of time scales. We also observed this in our dataset for both the full data range and the instrumental period and our data accurately captured past drought as its changes tracked with soil moisture, a good index of drought (see Additional file 1: Figs. S5 and S6) [40]. Given these disagreements and following other drought projection studies [41], we opted to estimate future drought conditions for each scenario as follows: For S1, we included no change relative to a current "baseline" by fixing drought values to the instrumental period average (2000–2016), the average was used to account for interannual climate variability. For S3 (representing "business-as-usual") we extrapolated the full historical data trends for each country (1850–2016) using univariate linear regression models (drought ~ year). The results of these models are available in Additional file 1: Table S3 and the coefficients then acted as a yearly multiplier (up until the extreme values of $+4$ for extreme wetness and $-4$ for extreme dryness). For S2, we took an intermediate value between S1 and S3. To account for uncertainty in the drought projections and to further examine how drought in isolation may alter future cholera outbreaks, a second sensitivity analysis was run, maintaining the other covariates at the 2016 levels and altering drought in six analyses $\pm 0.5, \pm 1$ and $\pm 2$ (or until the extreme values, $+4$ or $-4$). Full details and results of this sensitivity analysis are shown in Additional file 1: Table S4 and Fig. S7.

Poverty changes were based on SDG 1 [42], despite the limitations of the SDGs (e.g., ambiguous terms), they are a globally recognised standard for sustainable development. As such, S1 meets the goal of a 50% reduction in extreme ($< \$1.25$/day) poverty by 2030 and poverty eliminated by 2070. In S2, the 50% reduction goal is met by 2050 and by 2070 for S3. The poverty setting used in the SDGs is slightly lower ($\$1.25$) than the WorldBank data used in this analysis ($\$1.90$), and it is difficult to distinguish the level of poverty within the data; therefore, the projected scenarios mainly aligned with the second part of the goal, to halve the population in poverty by 2030.

Projected changes in freshwater withdrawal are largely dependent on future human behaviour and adaptation to changing water security, which are highly uncertain. Therefore, freshwater withdrawal projections were based on SDG6.4 and either increased or decreased based on each country's historical freshwater withdrawal relative

to available water resources, taken from the same source used in the model [28]. This indicator of freshwater security for each country is plotted in Additional file 1: Fig. S8. Expanded freshwater withdrawal would likely increase peoples' access but this must be done sustainably and in line with resources. Increased withdrawal may also be a sign of development as more people have access to wells, boreholes and piped water. As such, for S1 we increased sustainable freshwater availability by the middle of the projection period (2050) by 20% for countries with sufficient resources. For S2, we increased freshwater availability by 10% and for S3 by 5%.

For population projections, the United Nation's World Population Prospectus [43] median variant was used for all three scenarios. Although population growth is expected to be more restricted under high attainment of the SDGs, we opted to use a single medium population size to isolate the effects of the other environmental and socio-economic covariates.

## Results

### Model fitting and covariate selection

The univariate model results (p-values, coefficients, BIC and AUC of the 19 tested covariates against cholera outbreak occurrence) are shown in Table 2. Six of these were not significantly associated with the data at the threshold of $p < 0.1$. Of the remaining 13, one cluster was formed containing two highly correlated variables (soil moisture and drought), while all other covariates were considered uncorrelated at the given threshold and therefore could be included in the full model.

### Output from the best-fit model

After model fitting, five covariates were retained in the best-fit model. These include population, mean meteorological drought (in PDSI), average temperature, poverty headcount and per capita freshwater withdrawal. Goodness of fit measures and outputs for the best-fit model are shown below in Table 3. Higher population numbers and more people living in poverty were associated with increased cholera outbreaks. For the environmental covariates, per capita freshwater withdrawal was negatively associated with cholera, while higher temperatures and drier conditions (more negative PDSI) were both associated with increases in cholera outbreaks. These relationships are shown in the marginal effect plots in Fig. 2.

### Temporal and spatial effects

Re-running the covariate selection process with year, ISO3 country code and the 19 original predictor variables, selected year but not ISO3 at the significance threshold ($p = < 0.1$). It also selected the same covariates

Charnley *et al. BMC Infectious Diseases*     (2021) 21:1177

Page 6 of 12

**Table 2** Univariate model outputs and goodness-of-fit measures for the tested covariates against cholera outbreak occurrence, including p-values, coefficients, BIC and AUC

| Covariate | p value | Coefficient | BIC | AUC |
|---|---|---|---|---|
| Potential evapotranspiration (mm/day) | 0.961 | 0.011 | 785.323 | 0.5979 |
| Annual freshwater withdrawal (billion $m^3$) | 0.649 | − 0.029 | 784.570 | 0.5279 |
| Runoff (mm/year) | 0.373 | − 0.064 | 785.395 | 0.6068 |
| Health expenditure (% GDP) | 0.371 | 0.126 | 783.253 | 0.5389 |
| Prevalence of malnourishment (% population) | 0.139 | − 0.169 | 784.014 | 0.5892 |
| Gross domestic output (current $) | 0.126 | − 0.091 | 783.079 | 0.5148 |
| Population density (people/$km^2$) | 0.051* | − 0.145 | 781.802 | 0.5773 |
| Water withdrawal per capita ($m^3$/person/year) | 0.032* | 0.151 | 762.801 | 0.6184 |
| Average precipitation (mm) | 0.021* | − 0.263 | 780.742 | 0.6345 |
| People with basic handwashing facilities (% population) | 0.018* | 0.189 | 766.430 | 0.5882 |
| Percentage living in informal settlement (% urban population) | 0.013* | − 0.467 | 778.730 | 0.4903 |
| Mean drought | 0.003* | − 0.199 | 768.863 | 0.5827 |
| Human Development Index | 0.0002* | 1.014 | 767.927 | 0.6562 |
| People using at least basic sanitation (% population) | 0.0001* | 0.384 | 757.283 | 0.6347 |
| Poverty headcount (% population at < $1.90/day) | 0.0001* | − 0.583 | 768.649 | 0.7018 |
| Average temperature (°C) | 0.00005* | − 1.715 | 765.124 | 0.5349 |
| Soil moisture (%) | 0.00003* | − 0.706 | 768.044 | 0.6871 |
| People with basic drinking water (% population) | 0.00002* | 0.906 | 762.312 | 0.6521 |
| Population (log. population in thousands) | 0.0000000004* | − 3.064 | 741.192 | 0.6521 |

*p < 0.1

**Table 3** Output and goodness of fit measures for the best-fit model

| | Coefficient | Exp(coefficient) | p value |
|---|---|---|---|
| Mean national drought (PDSI) | − 0.0927813 | 0.9113928127 | 0.051172 |
| Population, total (log) | 1.3125412 | 3.7156036497 | $2.85 \times 10^{-13}$ |
| Average temperature (°C) | 0.0927423 | 1.0971789754 | 0.000113 |
| Poverty headcount (at < $1.90/day) | 0.0327487 | 1.0332908900 | $4.23 \times 10^{-16}$ |
| Per capita freshwater withdrawal ($m^3$/person/year) | − 0.0024225 | 0.9975804550 | $5.43 \times 10^{-7}$ |

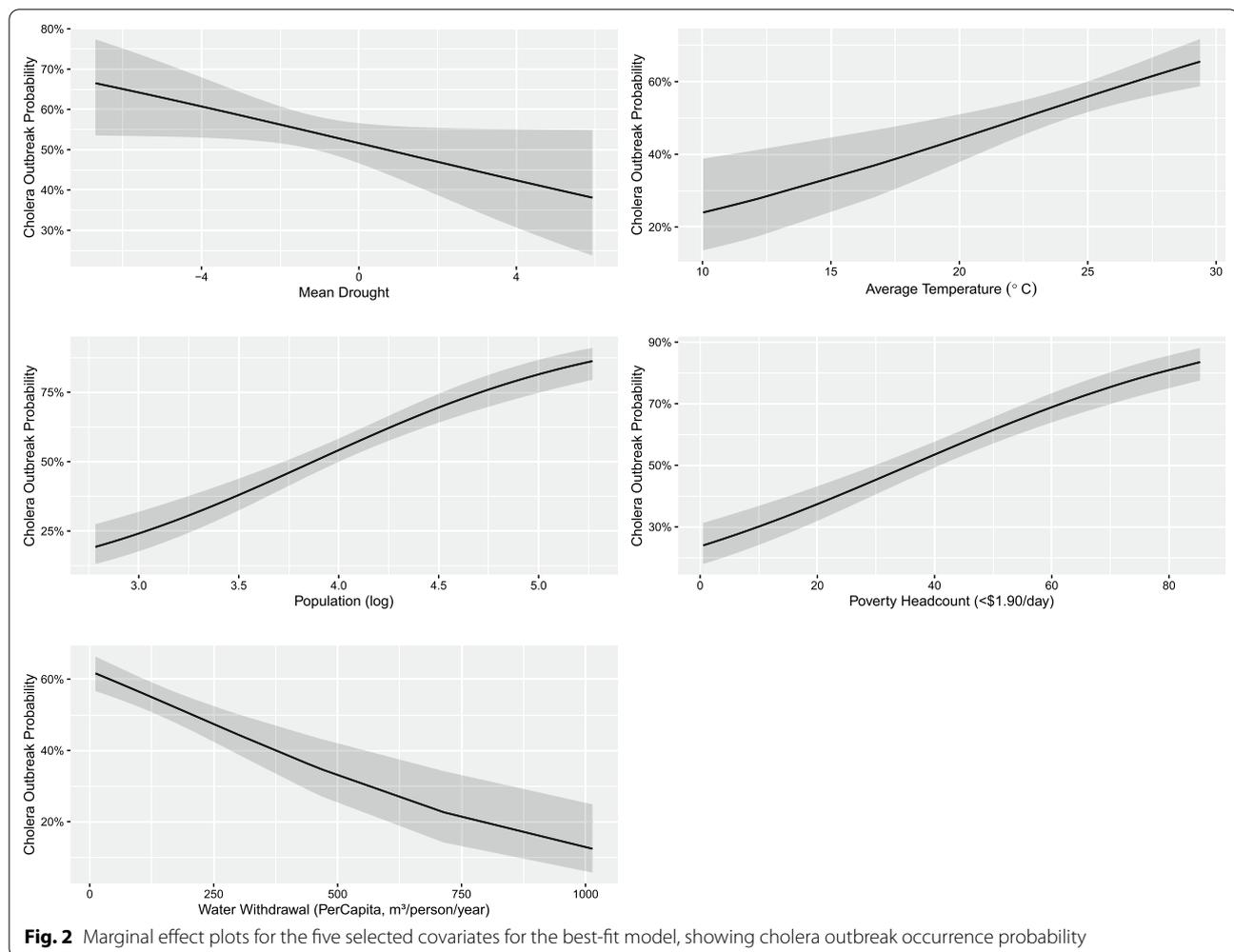| Residuals | Min | 1Q | Median | 3Q | Max |
|---|---|---|---|---|---|
| | − 2.0286 | − 0.7974 | 0.4069 | 0.8601 | 2.2564 |

$R^2$: 0.276254

AUC: 0.7784

as the original model and additionally basic handwashing and Human Development Index. The linear relationship between year and cholera was visualised using loess curves for each country (shown in Additional file 1: Fig. S9) and when accounted for AR1 autocorrelation year was found to no longer be significant (p = < 0.05).

Out-of-sample validation using AIC and LOO found no appreciable difference between the two selected best-fit models. Therefore, the model selected without the inclusion of year and country code in the selection process

was thus chosen as the best-fit model (diagnostic results are shown in Additional file 1: Information 4).

Cholera outbreak occurrence appears conditionally independent of year given the other covariates in the model, as time does not cause cholera but instead the changes in covariates over time, making them good predictors of cholera outbreak occurrence. It is also thought that some temporal increases in cholera are due to global improvements for detection of all-pathogen outbreaks from the mid 1990s onwards, especially in low- and

**Fig. 2** Marginal effect plots for the five selected covariates for the best-fit model, showing cholera outbreak occurrence probability

middle-income countries, improving countries' capacity for detection, response and therefore reporting [44, 45].

### Cholera projections to 2070

Cholera projections from the best-fit model according to the parameter values for each of the three scenarios are shown in Fig. 3. The cholera outbreak projections show several changes through to 2070 and spatial heterogeneity among countries over the continent. Most countries show a general decrease in cholera outbreaks in S1 and S2, with few exceptions e.g., Tunisia. Although countries with the highest cholera levels saw little change, remaining at a high outbreak occurrence level throughout, including the Democratic Republic of Congo (DRC) and Nigeria.

Figure 4 shows the decadal continental average for the projected cholera outbreak occurrence, to help understand the general trend across the continent. Overall, S3 shows a slight increase throughout the projected period, whereas S1 and S2 exhibit declines.

However, overlapping confidence intervals between S1 and S2 mean it is difficult to distinguish meaningful differences, although by 2070 S3 projects significantly more outbreaks than S1 and S2.

The drought sensitivity analysis showed modest changes through the six different analyses, with more negative values of PDSI seeing higher cholera outbreak occurrence (Additional file 1: Table S4 and Fig. S7). Despite this, these changes were not excessive with a 0.06 average increase in continental cholera outbreak occurrence from the 2000 to 2016 averages to sensitivity analysis 6 (2016 value – 2). This suggests that while future drought is likely to continue to affect cholera in Africa, improved socio-economic conditions may counteract this effect, by reducing pathogen exposure.

### Discussion

Cholera has well established environmental [6–8] and socio-economic links [9, 14, 15], such as poverty, poor WASH conditions, the Intertropical Convergence Zone
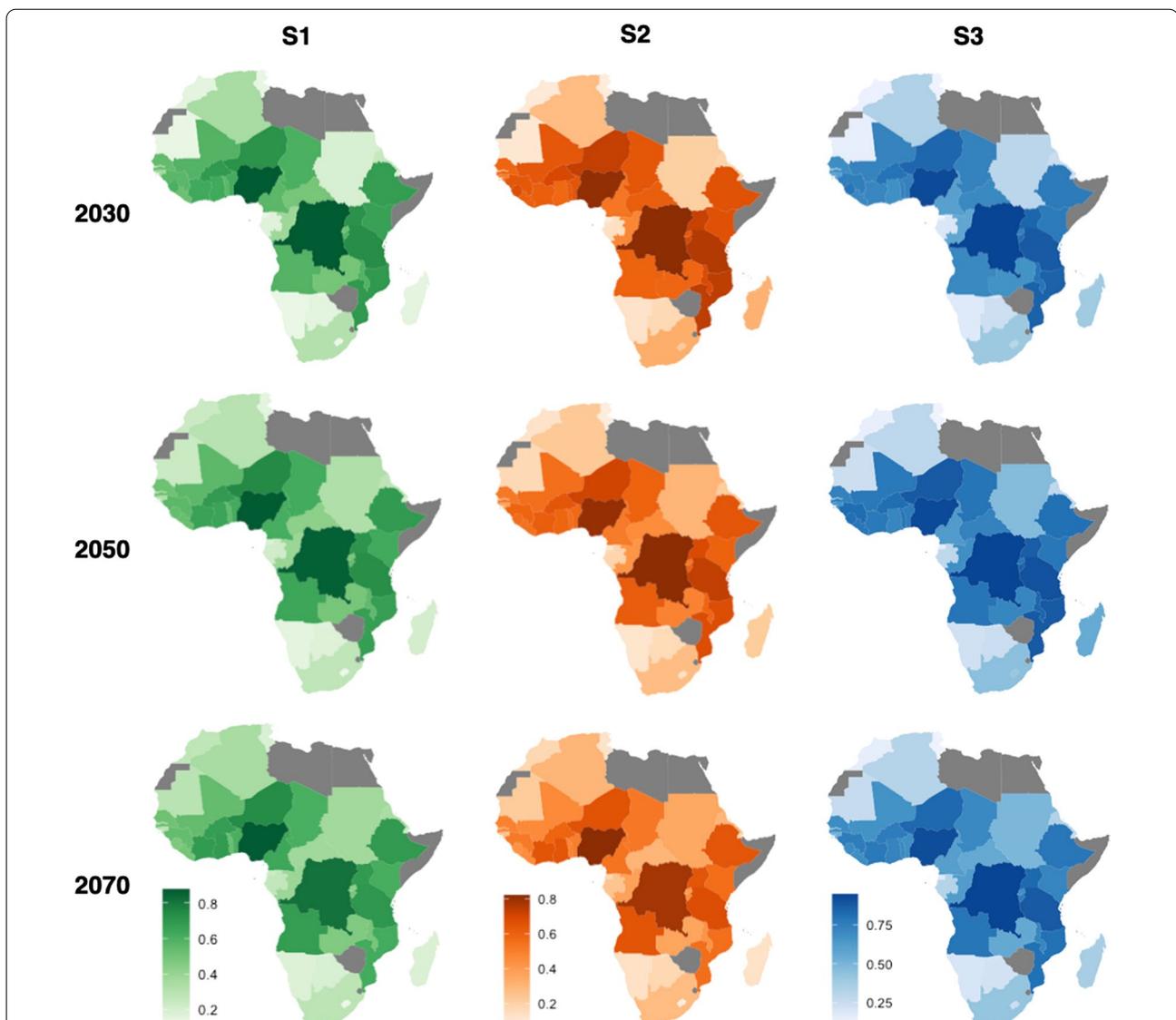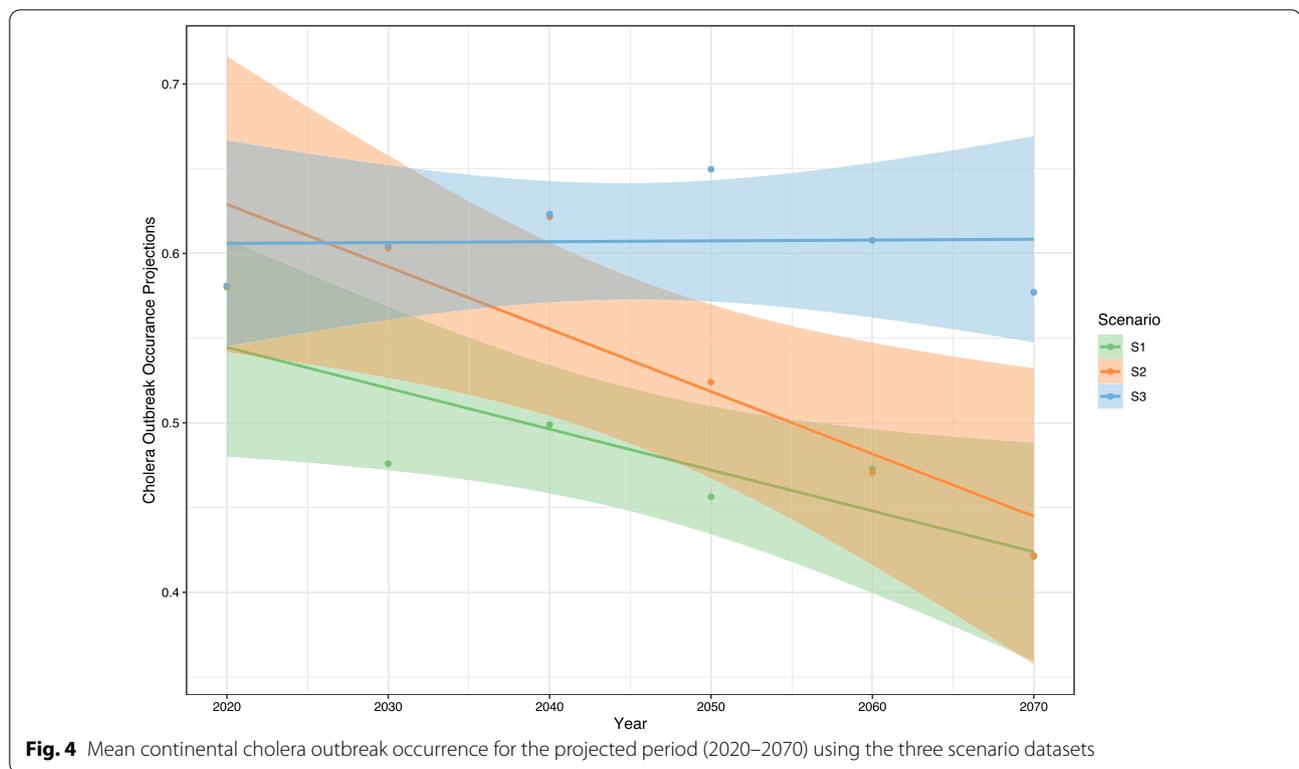
**Fig. 3** Projected cholera outbreak occurrence (0–1) for the three scenarios in 2030, 2050 and 2070. Grey represents countries where covariate data was missing (Botswana, Zimbabwe, Somalia, Egypt, Eswatini, Western Sahara, Algeria, Libya and Eritrea) and therefore could not be included in the model. The map is our own work and the shapefiles are taken from [46] under CC-BY SA, allowing them to be shared and adapted

and El Niño Southern Oscillation. Here, environmental variables were important covariates in the model. Meteorological drought (according to PDSI) was found to be a significant predictor of cholera outbreaks, with drier conditions seeing higher cholera outbreak occurrence. While previous studies have implicated drought in cholera outbreaks [2, 10, 11], our study models drought in isolation allowing a more in-depth investigation of its impacts, which have been largely understudied in comparison to flooding. In addition, we tested whether drought is likely to influence cholera outbreaks under scenarios of climatic change and socio-economic development (attainment of

the SDGs). While we found drought will continue to be an important hazard for cholera outbreaks in the future, our results suggest that gains in sustainable development (reduction of poverty, increased water security) may offset cholera risk in the future.

Temperature was identified as a significant predictor, providing another link between changing drought risk and increased cholera outbreak occurrence, as an increased temperature is important in both drought onset and duration. The positive relationship between temperature and cholera is expected, as cholera is considered a temperature-sensitive pathogen, with optimum

**Fig. 4** Mean continental cholera outbreak occurrence for the projected period (2020–2070) using the three scenario datasets

growth at elevated temperatures (up to a threshold) [47]. This may also represent an independent effect of temperature from drought and why both variables are independently selected in the model. For example, a 1 °C rise in temperature was associated with a twofold increase in cholera cases in Zanzibar [8]. Moreover, when run in the univariate models, precipitation had a slightly negative coefficient, again providing a potential link between drought, decreased water availability and cholera outbreaks. Precipitation, however, was not selected in the final model, potentially suggesting that precipitation effects for cholera in Africa, may be less important than temperature.

The inclusion of more than one type of drought index in the best-fit model (PDSI and water withdrawal) shows the importance of considering several drought definitions and measures when investigating its implications. Drought is a complex phenomenon involving climate, agriculture, water stress and societal response and therefore including additional drought variables can help capture the varying elements of the hazard, exposure and vulnerability. Water withdrawal per capita was a highly significant environmental variable in the model, linking to the original hypothesis that a reduction in water availability leads to riskier water practices. More water withdrawal suggests higher water availability for drinking and washing and a reduction in risky behaviour such as

with multi-use water. Better water management may help mitigate negative drought-related health outcomes, and when water is available, this should not be exploited to avoid times of scarcity.

Cholera is a disease of inequity and poverty and is often seen in combination with poor WASH facilities [14, 48]. Here, poverty was the most significant variable (according to the p-values) included in the model and may suggest that environmental determinants of cholera are only key drivers up to certain thresholds and then socioeconomic covariates are more appropriate predictors [13]. For example, droughts have been known to impact the US and Europe [49, 50], but large-scale cholera outbreaks do not occur due to generally high levels of sanitation and hygiene. Several socio-economic covariates were expected to be important here but only poverty was selected in the final model and all socioeconomic covariates were independently selected for model inclusion. A possible explanation is that other socio-economic covariates such as, sanitation, hygiene, drinking water and people living in informal settlements is captured within the effects of poverty and possibly enhancing its impact. Even with the ideal environment for cholera to proliferate, social conditions allow the link to be made for pathogen exposure and spread. Poor access to WASH facilities means that large groups of people are at risk, not just for cholera, but for several other diseases. For example,

Charnley *et al. BMC Infectious Diseases*     (2021) 21:1177

Page 10 of 12

nearly 90% of diarrhoeal disease has been attributed to sub-optimal WASH [51]. These findings highlight the need to meet or exceed the SDGs, lifting people out of poverty and providing basic sanitation and hygiene as a public health priority.

The scenario dataset and projections provide some insight into the future importance of climate and socio-economic development on cholera outbreak occurrence in Africa. Historical and projected changes are spatially heterogenous but projected continental trends under S3 slightly increased cholera outbreak occurrence to 2070. Whereas, under S2 and S1 cholera occurrence decreased to 2070, with S1 showing the lowest levels. The projected changes over the next 50 years show that reducing poverty, expanding sustainable freshwater availability and striving for greater emissions reductions will be important for achieving positive health outcomes. How societies will continue to respond and adapt to climate change and drought is difficult to determine in the future and therefore understanding future risks can be challenging. As with any projections and the creation of scenarios, uncertainty can be high, arising from theoretical, methodological and computational challenges in projecting future climate change and its consequences. There are also the realities of meeting or exceeding the SDGs. Several of the terms within the SDGs are ambiguous, consequently the aims and roadmap to achieve them are not clearly defined. Finally, we did not consider that in a "worst-case" scenario poverty and water withdrawal may regress or the introduction of new strains and changing immunity could complicate cholera eradication efforts. Despite this, with decreasing poverty and the expansion of freshwater availability, even the introduction of new cholera cases and strains could be offset.

Climate, drought and socio-economic data were missing for several countries and years, meaning that data had to be averaged or omitted. This meant that data were then missing from the model, or assumptions had to be made both spatially and temporally about conditions in certain countries, potentially introducing error. Using annual national data also meant that changes on a finer spatial and temporal scale cannot be determined from the work presented here such as seasonal changes in cholera and the presence of waterbodies within countries facilitating transmission [5]. Cholera is largely underreported, and many people never seek formal medical assistance. The WHO's most optimistic estimate suggests only 5–10% of cases are reported [52], possibly due to a spectrum of transmission dynamics and lulls in cases meaning focus on tracking the diseases can be lost [53]. Countries can be disincentivized to report outbreaks due to potential impact on tourism and trade [54]. Considering this underreporting, issues may have arisen from assigning the outcome variable to zero for missing years, as this could have led to the underrepresentation of cholera outbreaks. Given the results of the sensitivity analysis in Additional file 1: Information 1, however, we believe this is the best interpretation of missing values, as removing values created issues when trying to select covariates from small numbers of data points. Furthermore, the cholera data lack age and sex-disaggregation, meaning that demographic differences were not captured. GLMs assume a monotonic relationship and therefore non-linear effects of several covariates might not be captured and evaluating these non-linear effects are a potential area of future work. This issue may also have been present for the S3 drought projections as some countries fit the linear trend better than others.

## Conclusions

In conclusion, the relationships between temperature, drought and water withdrawal help add further evidence to the original hypothesis that hotter and drier conditions and a lack of freshwater availability increases cholera outbreak occurrence, potentially through risky water behaviours. Although elevated pathogen concentrations are difficult to distinguish from these results, the importance of elevated temperatures and its effect on cholera may be related to increases in pathogen concentrations. Socio-economic variables came out highly significant in the best-fit model, showing the impact of vulnerability in times of water shortage and the importance of lifting people out of poverty to improve health and reduce mortality.

The work presented here offers additional insight into how climate change may yield health impacts in the future and work should build on these results, to understand these relationships on a finer spatial scale. High burden countries such as the DRC and Nigeria saw very few changes in cholera over the projected period and scenarios, showing potential areas for further work to understand outbreak drivers and mitigators in the most at-risk countries.

## Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12879-021-06856-4.

**Additional file 1.** Additional details about the data and models used here.

Charnley *et al. BMC Infectious Diseases*　(2021) 21:1177

Page 11 of 12

## Declarations

### Ethics approval and consent to participate

This manuscript does include human data, but no ethical approval or consent was needed, as this data is freely available through WHO, the WorldBank and the UN Development Programme and completely anonymised. No animal data was used in this research.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

### Author details

[1]School of Public Health, Imperial College London, Norfolk Place, London W2 1PG, UK. [2]MRC Centre for Global Infectious Disease Analysis, School of Public Health, Imperial College London, London W2 1PG, UK. [3]Institute for Risk and Disaster Reduction, University College London, London WC1E 6BT, UK. [4]Institute for Globa Health, University College London, London, UK. [5]University of Agder, Kristiansand, Norway. [6]Department of Statistical Science, University College London, London WC1E 6BT, UK. [7]MRC Unit The Gambia at London School of Hygiene and Tropical Medicine, Atlantic Boulevard, Fajara, The Gambia.

## References

1. Mendelsohn J, Dawson T. Climate and cholera in KwaZulu-Natal, South Africa: the role of environmental factors and implications for epidemic preparedness. Int J Hyg Environ Health. 2008;211(1–2):156–62.
2. Tauxe RV, Holmberg SD, Dodin A, Wells JV, Blake PA. Epidemic cholera in Mali: high mortality and multiple routes of transmission in a famine area. Epidemiol Infect. 1988;100(2):279–89.
3. Germani Y, Quilici ML, Glaziou P, Mattera D, Morvan J, Fournier JM. Emergence of cholera in the Central African Republic. Eur J Clin Microbiol Infect Dis. 1998;17(12):888.
4. World Health Organization. Cholera. 2020. https://www.who.int/health-topics/cholera#tab=tab_1. Accessed 20 Jul 2020.
5. Nkoko DB, Giraudoux P, Plisnier PD, Tinda AM, Piarroux M, Sudre B, et al. Dynamics of cholera outbreaks in Great Lakes region of Africa, 1978–2008. Emerg Infect Dis. 2011;17(11):2026.
6. De Magny GC, Thiaw W, Kumar V, Manga NM, Diop BM, Gueye L, et al. Cholera outbreak in Senegal in 2005: was climate a factor? PLoS ONE. 2012;7(8):e44577.
7. Rebaudet S, Sudre B, Faucher B, Piarroux R. Environmental determinants of cholera outbreaks in inland Africa: a systematic review of main transmission foci and propagation routes. J Infect Dis. 2013;208(Suppl 1):S46–54.
8. Reyburn R, Kim DR, Emch M, Khatib A, Von Seidlein L, Ali M. Climate variability and the outbreaks of cholera in Zanzibar, East Africa: a time series analysis. Am J Trop Med Hyg. 2011;84(6):862–9.
9. Olago D, Marshall M, Wandiga SO, Opondo M, Yanda PZ, Kangalawe R, et al. Climatic, socio-economic, and health factors affecting human vulnerability to cholera in the Lake Victoria basin, East Africa. AMBIO. 2007;36(4):350–8.
10. Abdussalam AF. Modelling the climatic drivers of cholera dynamics in northern Nigeria using generalised additive models. Int J Geogr Environ Manag. 2016;2(1):84–97.
11. Rieckmann A, Tamason CC, Gurley ES, Rod NH, Jensen PKM. Exploring droughts and floods and their association with cholera outbreaks in sub-Saharan Africa: a register-based ecological study from 1990 to 2010. Am J Trop Med Hyg. 2018;98(5):1269–74.
12. Mark O, Jørgensen C, Hammond M, Khan D, Tjener R, Erichsen A, Helwigh B. A new methodology for modelling of health risk from urban flooding exemplified by cholera—case Dhaka, Bangladesh. J Flood Risk Manag. 2018;11:S28–42.
13. Weill FX, Domman D, Njamkepo E, Tarr C, Rauzier J, Fawal N, et al. Genomic history of the seventh pandemic of cholera in Africa. Science. 2017;358(6364):785–9.
14. Talavera A, Perez EM. Is cholera disease associated with poverty? J Infect Dev Ctries. 2009;3(06):408–11.
15. Mari L, Bertuzzo E, Righetto L, Casagrandi R, Gatto M, Rodriguez-Iturbe I, et al. Modelling cholera epidemics: the role of waterways, human mobility and sanitation. J R Soc Interface. 2012;9(67):376–88.
16. Sasaki S, Suzuki H, Fujino Y, Kimura Y, Cheelo M. Impact of drainage networks on cholera outbreaks in Lusaka, Zambia. Am J Public Health. 2009;99(11):1982–7.
17. Ranjbar R, Rahbar M, Naghoni A, Farshad S, Davari A, Shahcheraghi F. A cholera outbreak associated with drinking contaminated well water. Arch Iran Med. 2011;14(5):339–40.
18. Jutla A, Khan R, Colwell R. Natural disasters and cholera outbreaks: current understanding and future outlook. Curr Environ Health Rep. 2017;4(1):99–107.
19. Codeço CT. Endemic and epidemic dynamics of cholera: the role of the aquatic reservoir. BMC Infect Dis. 2001;1(1):1.
20. Global Task Force on Cholera Control. Ending cholera a global roadmap to 2030. World Health Organization; 2017 Oct 3.
21. Haile GG, Tang Q, Hosseini-Moghari SM, Liu X, Gebremicael TG, Leng G, et al. Projected impacts of climate change on drought patterns over East Africa. Earth's Future. 2020;8(7):e2020EF001502.
22. Ahmadalipour A, Moradkhani H, Castelletti A, Magliocca N. Future drought risk in Africa: integrating vulnerability, climate change, and population growth. Sci Total Environ. 2019;662:672–86.
23. World Health Organization. The global health observatory. 2020. https://www.who.int/data/gho. Accessed 21 Jan 2021.
24. ECMWF. ERA5. 2020. https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5. Accessed 21 Jan 2021.
25. NCAR. Dai Global Palmer Drought Severity Index (PDSI). 2020. https://rda.ucar.edu/datasets/ds299.0/index.html#!sfol-wl-/data/ds299.0. Accessed 21 Jan 2021.
26. Copernicus. Soil moisture gridded data from 1978 to present. 2018. https://cds.climate.copernicus.eu/cdsapp#!/dataset/satellite-soil-moisture?tab=form. Accessed 21 Jan 2021.
27. NCAR. CRU TS gridded precipitation and other meteorological variables SINCE 1901. 2015. https://climatedataguide.ucar.edu/climate-data/cru-ts-gridded-precipitation-and-other-meteorological-variables-1901. Accessed 21 Jan 2021.

28. Ritchie H. Water use and stress. 2017. https://ourworldindata.org/water-use-stress. Accessed 21 Jan 2021.
29. UNH/GRDC. UNH/GRCD composite runoff fields V1.0. 2000. https://www.compositerunoff.sr.unh.edu. Accessed 21 Jan 2021.
30. WorldBank. World Bank open data. 2017. https://data.worldbank.org. Accessed 21 Jan 2021.
31. United Nations Development Programme. Human development data (1990–2018). 2018. http://hdr.undp.org/en/data#. Accessed 21 Jan 2021.
32. Garske T, Van Kerkhov, MD, Yactayo S, Ronveaux O, Lewis RF, Staples JE, et al. Yellow fever in Africa: estimating the burden of disease and impact of mass vaccination from outbreak and serological data. PLoS Med. 2014;11(5):e1001638.
33. Gaythorpe KA, Hamlet A, Jean K, Ramos DG, Cibrelus L, Garske T, et al. The global burden of yellow fever. eLife. 2021;10:e64670.
34. WorldClim. Future climate data. 2020. https://worldclim.org/data/cmip6/cmip6climate.html. Accessed 21 Jan 2021.
35. Shanahan TM, Overpeck JT, Anchukaitis KJ, Beck JW, Cole JE, Dettman DL, et al. Atlantic forcing of persistent drought in West Africa. Science. 2009;324(5925):377–80.
36. Yang Y, Zhang S, Roderick ML, McVicar TR, Yang D, Liu W, et al. Little change in Palmer Drought Severity Index across global land under warming in climate projections. 2020. Hydrol Earth Syst Sci Discuss. https://doi.org/10.5194/hess-2019-701.
37. Sheffield J, Wood EF, Roderick ML. Little change in global drought over the past 60 years. Nature. 2012;491(7424):435–8.
38. Touchan R, Anchukaitis KJ, Meko DM, Attalah S, Baisan C, Aloui A. Long term context for recent drought in northwestern Africa. Geophys Res Lett. 2008. https://doi.org/10.1029/2008GL034264.
39. Verschuren D, Laird KR, Cumming BF. Rainfall and drought in equatorial east Africa during the past 1,100 years. Nature. 2000;403(6768):410–4.
40. Tian-Jun Z, Tao H. Projected changes of palmer drought severity index under an RCP8.5 scenario. AOSL. 2013;6(5):273–8.
41. Ahmadalipour A, Moradkhani H. Multi-dimensional assessment of drought vulnerability in Africa: 1960–2100. Sci Total Environ. 2018;644:520–35.
42. United Nations. The 17 goals. 2015 [cited 2021 Jan 21]. https://sdgs.un.org/goals.
43. United Nations Department for Economic and Social Affairs. Population dynamics. World population prospectus 2019. 2019. https://population.un.org/wpp/. Accessed 21 Jan 2021.
44. Kluberg SA, Mekaru SR, McIver DJ, Madoff LC, Crawley AW, Smolinski MS, et al. Global capacity for emerging infectious disease detection, 1996–2014. Emerg Infect Dis. 2016;22(10):e151956.
45. Ratnayake R, Finger F, Edmunds WJ, Checchi F. Early detection of cholera epidemics to support control in fragile states: estimation of delays and potential epidemic sizes. BMC Med. 2020;18(1):1–16.
46. Thematicmapping.org. Thematic Mapping API. 2009. https://thematicmapping.org/downloads/world_borders.php. Accessed 21 Jan 2021.
47. Borroto RJ. Ecology of *Vibrio cholerae* serogroup 01 in aquatic environments. Pan Am J. 1997;2:328–33.
48. Penrose K, de Castro MC, Werema J, Ryan ET. Informal urban settlements and cholera risk in Dar es Salaam, Tanzania. PLoS Negl Trop Dis. 2010;4(3):e631.
49. Easterling DR, Wallis TW, Lawrimore JH, Heim RR Jr. Effects of temperature and precipitation trends on US drought. Geophys Res Lett. 2007. https://doi.org/10.1029/2007GL031541.
50. Moravec V, Markonis Y, Rakovec O, Kumar R, Hanel M. A 250-year European drought inventory derived from ensemble hydrologic modelling. Geophys Res Lett. 2019;46(11):5909–17.
51. Ramesh A, Blanchet K, Ensink JH, Roberts B. Evidence on the effectiveness of water, sanitation, and hygiene (WASH) interventions on health outcomes in humanitarian crises: a systematic review. PLoS ONE. 2015;10(9):e0124688.
52. Ali M, Lopez AL, You Y, Kim YE, Sah B, Maskery B. The global burden of cholera. Bull World Health Organ. 2012;90:209–18.
53. Azman AS, Moore SM, Lessler J. Surveillance and the global fight against cholera: setting priorities and tracking progress. Vaccine. 2020;38(Suppl 1):A28.
54. Legros D. Global cholera epidemiology: opportunities to reduce the burden of cholera by 2030. J Infect Dis. 2018;218(Suppl 3):S137–40.

## Publisher's Note