Imperial College London
Department of Infectious Disease Epidemiology
School of Public Health

Exploring disaster-related disease outbreaks and the impact of pre-existing vulnerabilities: Focusing on cholera and Africa

Georgina Eva Ceres Charnley

Thesis submitted for PhD Examination
2022

Statement of Originality

I declare that the work presented in this thesis is my own work, produced under the supervision of Dr Kris Murray, Dr Katy Gaythorpe and Professor Ilan Kelman.

All sources used including figures, code and literature are cited appropriately in the thesis.

Georgina Eva Ceres Charnley

November 2022

Copyright Declaration

Acknowledgements

I would like to express gratitude and thanks for the help, support and encouragement of my supervisory team; Dr Kris Murray, Dr Katy Gaythorpe and Professor Ilan Kelman. The work in this thesis would not have been possible without their knowledge and insight, along with their endless time and patience in helping me make the most of my PhD experience.

I would also like to thank the Nigeria Centre for Disease Control, namely Mr Sebastian Yennan, Dr Chinwe Ochu and Dr Chikwe Ihekweazu, who kindly provided data for some of the analyses completed here, along with their support and guidance on cholera in Nigeria.

I am very grateful of the time, patience, and encouragement of fellow researchers at several institutions, including Mr Myles Harris, Dr Heather Whitaker, Dr Yonas Weldeselassie, Dr Nathan Green, Dr Wes Hinsley, Dr Kévin Jean, Dr Laura Peters and Dr Jamon Van Den Hoek. Not only did these researchers improve the work here but they also helped me develop skills I can take forward in my career.

I would like to thank my assessors, Dr Isobel Blake, Dr Virginie Le Masson, Dr Lucy Okell, Dr Anne Cori and Professor Rachel Lowe, and how accommodating and flexible they have been as we switched from in-person to remote working. I appreciate their time in assessing my PhD and the feedback they have provided, which has greatly strengthened the work presented here.

All the PhD students in the School of Public Health and the Grantham Institute at Imperial College London and UCL's Institute for Risk and Disaster Reduction, who have been very helpful and supportive, creating a community even in the challenging times we have been working in.

Finally, I would like to thank my family, both chosen and biological. The support and encouragement they provide has been fundamental to everything I have accomplished.

Abstract

Research into the links between infectious disease outbreaks and disasters has increased over time, with continued discussion regarding the rate and mechanisms for these links. Yet, significant discrepancies and research gaps remain globally in terms of the frequency, geography and characteristics of post-disaster disease outbreaks. Here, the aim was to address these gaps and discrepancies by exploring several disasters and their associated risk factors and further quantifying their impact. By increasing this understanding, this improves the extent to which disaster-related disease outbreaks can be prepared for, to prevent outbreaks from exacerbating and prolonging disaster recovery.

To further understand these risk factors, disease outbreaks in drought and conflict settings were analysed and an overview of the research area was gained through the first global systematic review of the literature. Cholera was selected as a focus for the research here, due to it being a disease of global public health importance and potentially linked to drought and conflict. The area of study was predominately Africa (due to the relatively high cholera, conflict and drought burden) and was studied at several spatial scales, from national to administrative level 1. The research identified many important risk factors for cholera outbreaks in a disaster context.

To help further identify areas for prioritisation both nationally and sub-nationally, drought-related cholera outbreaks were investigated, as droughts are a relatively understudied natural hazard. Generalised linear models were used to identify a potential relationship and the flexibility of the modelling approach allowed for multiple covariates to be tested. The lack of available water during a drought exacerbates risk factors relating to cholera transmission. However, increasing freshwater availability, improving access to sanitation, poverty elimination and emissions reductions could help to offset cholera risk in the future. Using random forest models, specific targets for these risk factors were further quantified for Nigeria. States with Multi Dimensional Poverty Index values over 0.38 and sanitation access below 54% were particularly at risk for cholera transmission, which includes most northern states in Nigeria.

Conflict-related cholera outbreaks were analysed by applying the self-controlled case series in a new application. The modelling approach uses conditional logistic regression to understand the impact of an exposure (conflict) on an event (outbreak). Conflict had the most significant

impact on cholera in the first week after the conflict and increased the risk of cholera outbreaks by as much as 3.6 times. The recently developed percentage attributable fraction equations were applied to these results and identified <20% of cholera outbreaks being attributable to conflict in Nigeria and the Democratic Republic of Congo (DRC). The research helped quantify a specific health effect of both the Boko Haram conflict in Nigeria and the civil unrest in eastern DRC.

Disease, disasters and global change were brought together here beyond what has previously been done and the knowledge gained was applied to policy throughout. Furthermore, a variety of projections and scenarios were used to identify how potential future conditions could alter cholera transmission. The Global Task for Cholera Control have ambitious 2030 targets, that will be essential for global cholera eradication and are important to make sure governments are committed to overcoming challenges. Both disasters and cholera outbreaks are not new phenomenon and societies have always had to respond and adapt but marked differences in global inequity can prevent this. Giving people agency and empowerment to react to sudden changes and make informed decisions to protect their health will require long-term investment in sustainable development. Enhancing development has far-reaching impacts and is essential for controlling disease, both regionally and globally.

Contents

List of Figures

List of Tables

List of Abbreviations

ACLED	Armed Conflct Location & Event Data Project
ACF	AutoCorrelation Function
AIC	Akaike Information Criterion
AUC	Area Under the Receiver Operator Characteristic Curve
ARIMA	AutoRegressive Integrated Moving Average
BIC	Bayesian Information Criterion
CMIP6	Coupled Model Inter-Comparison Project 6
CTC	Cholera Treatment Centre
CV	Coefficient of Variation
DALY	Disability-Adjusted Life Year
DRC	Democratic Republic of Congo
GHDx	Global Health Data Exchange
GLM	Generalised Linear Models
GTFCC	Global Task Force on Cholera Control
HWC	High Water Withdrawal Countries
IDP	Internally Displaced Person
IRR	Incidence Rate Ratio
LAC	Latin America & the Caribbean
LOO	Leave-One-Out Cross Validation
MAE	Mean Absolute Error
MSE	Mean Squared Error
MeSH	Medical Subject Heading
MLE	Maximum Likelihood Estimation
MPI	Multidimensional Poverty Index
NCDC	Nigeria Centre for Disease Control
NCP	National Cholera Plans
NIH	National Institute of Health
NGO	Non-Governmental Organisation
OCV	Oral Cholera Vaccination
OOB	Out-of-Bag
ORS	Oral Re-hydration Solution
PAF	Percentage Attributable Fraction
PDSI	Palmers Drought Severity Index

PET	Potential Evapotranspiration
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analysis
r	Pearson Correlation Coefficient
R_t	Time-Varying Reproductive Number
RF	Random Forest
RCP	Representative Concentration Pathways
RMSE	Root-Mean-Square Error
scPDSI	Self-Calibrated Palmers Drought Severity Index
SCCS	Self-Controlled Case Series
SI	Serial Interval
SPEI	Standardised Precipitation Index
SDG	Sustainable Development Goals
WASH	Water, Sanitation & Hygiene
WHO	World Health Organization

Chapter 1

Introduction

Global health has made steady improvements through recent decades, especially in terms of communicable diseases due to improvements in sanitation and hygiene, along with vaccine discovery and mass immunisation schedules. As standard of living increases, increasing life expectancy, non-communicable diseases have replaced the communicable disease burden in many countries. A decrease in infectious diseases has meant less premature deaths and with an ageing population the incidence of ischemic heart disease, stroke and cancer has increased [1].

However, after accounting for population growth and ageing, the absolute number of Disability-Adjusted Life Years (DALYs) has remained stable [2]. Four communicable diseases remain in the top six causes of global DALYs (2010), lower respiratory tract infection, diarrhoea, HIV/AIDS and malaria [3]. One of the main reasons for this continued burden is the high levels of inequity that remains worldwide, with marked regional and national divergences from the global trend (of decreasing infectious disease mortality). This is especially concerning as children under 5 years old represent 57% of the global communicable disease burden (2019) [3, 4], particularly in sub-Saharan Africa (Figure 1.1), where the growing population will put more people at risk in the future [2].

Several diseases have epidemic and pandemic potential to cause public health emergencies, incurring large economic and social costs such as HIV/AIDS, cholera and Zika virus [5, 6]. For example, extreme poverty has increased for the first time in 20 years, with the COVID-19 pandemic considered the cause [7]. The global impact of infectious disease outbreaks remains an ever-present threat, with globalisation meaning no country is immune to these risks. The far-reaching repercussions of outbreaks make understanding the complexities of outbreak risk fundamental.

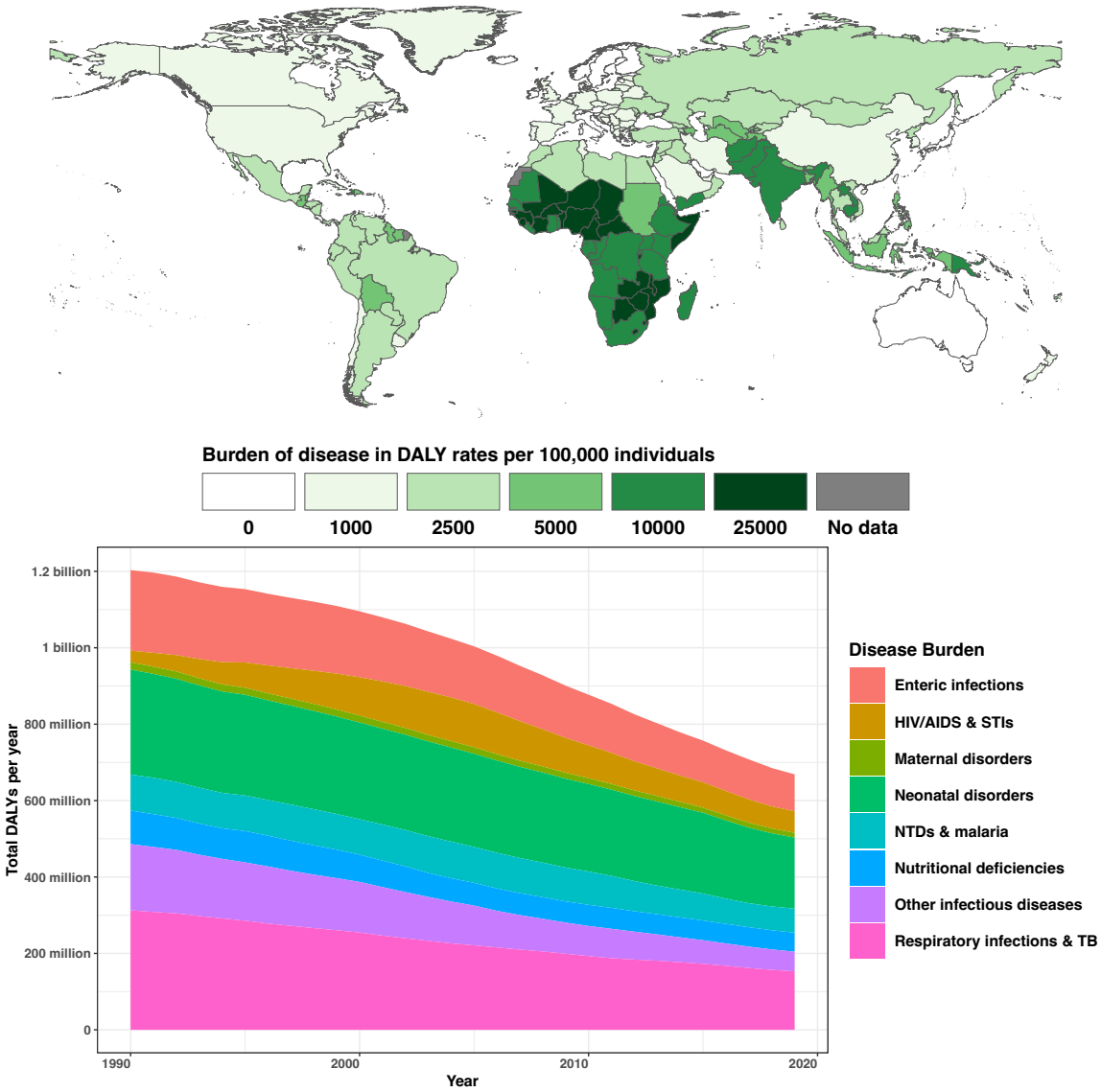


Figure 1.1: Global burden of communicable, neonatal, maternal and nutritional diseases in **a**, national DALY rates per 100,000 individuals for 2019 and **b**, total global DALYs per year for 1990 to 2019. . STIs - Sexually Transmitted Infections. NTDs - Neglected Tropical Diseases. TB - Tuberculosis. License: CC-BY [4, 8].

1.1 Mechanisms for Disaster-related Disease Outbreaks

1.1.1 Disease and Disaster Terminology

Across diseases, outbreaks can be initiated or exacerbated by a variety of influences, e.g., food, water, sanitation and health systems [9]. Other than the failure of governments to provide and protect such services, specific events can exacerbate several of these influences. An example of such event is a disaster, with their consequences resulting in a regression of many of the disease burden gains mentioned [10, 11].

The exacerbation of disease outbreaks is caused by a change in population vulnerability. The term vulnerability will be used throughout the thesis and both pre-existing vulnerability (present before the disaster) and vulnerability changes during or after the disaster, will be discussed. Vulnerabilities are conditions and long-term processes that increase the susceptibility to harm caused by the adverse event and include a wide range of factors e.g., poverty, education and housing [12].

Here, in accordance with disaster research terminology, the term disaster is a threatening event within a given time period and encompasses a natural hazard (e.g., earthquakes, floods, droughts) or an armed conflict (e.g., terrorism, civil war) [13, 14, 15]. The terms disaster, hazard and extremes will be used when discussing these events and locations which are experiencing them will be termed fragile settings or complex emergencies.

Hazards in relation to climate change will be discussed throughout the thesis. Climate is long-term patterns and trends in the weather, the classical period for averaging these trends is 30 years. Extreme weather or climate (natural hazards) are changes in the weather above or below the upper or lower ranges of observed climate values. Additionally, climate change is changes in the mean and/or the variability of the climate's properties, that persists for an extended period of time [16].

According to the World Health Organization (WHO), an infectious disease outbreak is an occurrence of a disease above normal expectancy [5]. The number of cases may vary according

to the aetiological agent as well as the size and type of previous and existing exposure, while the geographic occurrence of some outbreaks may be further shaped by whether a pathogen is endemic or epidemic.

Risk factors are specific vulnerabilities that interact with the hazard, whether that a natural hazard, conflict or an infectious disease outbreak, and increase the probability of the population being exposed to harm. The term risk factor cascade, also used in a disaster-disease context by Hammer *et al.* [17], will be used here and relates to risk factors which are linked or lead to other risk factors.

Disaster research terminology is contentious, with much disagreement over the correct terms to accurately describe complex hazard parameters. It is acknowledged that at the current time, no term may be correct or fully encompass all aspects of the hazard or vulnerability. However, for the purpose of this work, and inline with the discussion above, brief descriptions of all the terms used in this thesis are below:

- Vulnerability/Pre-existing vulnerability - Any condition or long-term process which increases susceptibility to the hazard
- Disaster/Hazard/Extreme - A threatening event within a given time period
- Climate/Climate Change - Long-term weather patterns/long-term changes in the climate
- Fragile settings/Complex emergencies - Locations experiencing a hazard
- Outbreak - An infectious disease occurrence above normal expectancy
- Risk factor - Specific vulnerabilities that increase the susceptibility to harm/exposure
- Risk factor cascade - Risk factors which are thought to be linked to or leads to other risk factors

1.1.2 Natural Hazard-related Disease Outbreaks

Natural hazards can be hydrological, meteorological, climatological and geophysical and include floods, droughts, earthquakes and storms (cyclones, hurricanes and typhoons). These events can cause widespread disruption to the affected population, especially without effective disaster mitigation and adaptation. The disruption caused by these events largely focuses on their direct effects such as destruction of infrastructure [18, 19] and fatalities [20, 21] but long-term and indirect impacts of the disaster are also important and often more impactful.

An example of an indirect impact of a natural hazard is an infectious disease outbreak, as the disruption caused by the hazard can act as a catalyst for outbreaks by initiating or worsening risk factors within the population [17]. Natural hazard-related outbreaks are a global issue with low- (post-earthquake rotavirus outbreak in India [22]), middle- (flood-related leptospirosis outbreak in Brazil [23]) and high- (norovirus outbreak after Hurricane Katrina in the US [24]) income countries being affected. Causative agents of these outbreaks tend to be endemic (constant presence and/or usual prevalence of a disease in a location [25]), as the exacerbated vulnerabilities increase exposure to the circulating pathogen reservoirs. Exceptions of where a non-endemic disease has caused a post-disaster outbreak has mainly been due to introductions by troops/peacekeepers and humanitarians, e.g., the 2010 Haitian cholera outbreak [26].

1.1.3 Conflict-related Disease Outbreaks

Similar to natural hazards, the impacts of conflicts often focus on direct infrastructure damage and fatalities in both the media and research but there are also several examples of conflict-related disease outbreaks, such as an outbreak of cutaneous leishmaniasis during the Syrian conflict [27] and an ongoing cholera outbreak during the Yemani civil war [28]. The effects of conflicts in terms of disease risk are complex, with fear and a lack of trust playing a more pivotal role. A lack of trust is especially problematic in conflicts where atrocities are committed on multiple sides and causes difficulties when providing healthcare and support [29].

Conflict-affected populations may not perceive accessing care as safe or people may not want to

leave their homes in fear of violence and roadblocks. For example, willingness to access Ebola treatment centres was low in the DRC due to healthcare attacks [30]. Furthermore, disease control efforts such as vaccination, can have poor uptake due to a lack of trust between the population and the government. Mistrust can then create an ideal environment for the spread of misinformation and vaccine hesitancy [31, 32]. Disease outbreak control requires a community effort and cooperation from everyone to follow public health guidelines. Unfortunately, disasters can sometimes result in an individualistic mentality, especially when satisfaction with the government and health officials is low [33, 34].

1.1.4 Historically Identified Risk Factors

Several risk factors have been suggested for causing post-disaster disease outbreaks. Identifying these risks and understanding their complexities is essential in preventing post-disaster outbreaks. Commonly cited risk factors include:

- Poor access to water, sanitation and hygiene (WASH) [35]
- Changes in diet and available foodstuff [36]
- Alterations in vector behaviour and control [37]
- Issues with housing and shelter, especially overcrowding [38]
- Problems obtaining healthcare [39, 40]
- Breakdown in preventative disease programs e.g., vaccinations, bed nets [39, 40]
- Unmanaged population displacement [41, 10, 42]

Displacement has the potential to worsen many of the risk factors mentioned, especially when people are displaced to overcrowded camps without proper facilities. The increased numbers of displaced persons (and troops/soldiers or humanitarians) can increase the number of people susceptible and naive to a disease, therefore propagating outbreaks [43]. These factors influence

population risk, commonly expressed as hazard x vulnerability, with exposure sometimes added (Crichton's risk triangle) [44]. Understanding these factors are important as over-emphasis on a single hazard (such as a natural hazard or armed conflict), reduces the insight into population exposure and vulnerability [45].

To add further complexity, few risks act solely to cause an outbreak and can typically be multi-faceted. Therefore, understanding how risk factors are potentially linked in risk factor cascades and their impact on pre-existing population vulnerability is vitally important [17, 46]. Both natural hazards and conflicts relate to and exacerbate pre-existing vulnerability and can increase the probability of post-disaster outbreaks. For example, following a disaster there can be a loss of income generation and disruption to education, exacerbating poverty [47].

1.2 Implications for Climatic and Social Changes

1.2.1 Climate Change and Natural Hazards

Intersections between disasters and disease provides an opportunity to explore the mechanisms through which global change (such as climate change and sustainable development) could yield health impacts [48, 49]. Climate change has the potential to alter some hazard parameters (e.g., intensity or frequency) [50, 51]. For example, sea level rise and warming temperatures are projected to change hurricane and Asian monsoon frequency and intensity [50, 52]. Furthermore, fewer cold and frost days and an average increase in global temperatures are very likely to play a role in heatwaves and droughts [53].

As climate models improve and uncertainty decreases, certainty around how these hazard parameters may alter has increased [54]. This thesis aims to capitalise on this new understanding to explore an aspect of future disease risk. These relationships should not be over-simplified though and projected climatic changes are complex, with their effects being spatially and temporally heterogeneous. Exposure and vulnerability are dynamic and inequities are not expressed in a uniform way, depending on economic, social, geographic, demographic, cultural, institu-

tional, governmental, and environmental factors [55]. It is therefore important for disaster risk reduction to analyse and estimate how communities may be impacted across different areas of the world.

1.2.2 Climate Change and Conflict

Previous research has suggested several links between climate change and conflict, in terms of frequency, intensity and duration. Examples of how climate change may lead to conflicts include a loss of income generation and livelihoods [56], marginalisation of communities [57], food and water insecurity [56, 58] and migration [59]. Many of these risk factors may increase the risk of conflict over resources, as they become more scarce or unpredictable. For example, altered drought frequency and intensity in the Fertile Crescent was suggested as a reason for armed conflict escalation in Syria [60]. Additionally, during El Niño years (compared to La Niña), new civil conflicts nearly double in the tropics from 1950 to 2004 [61].

People who live in conflict-affected or post-conflict areas are particularly vulnerable to climate change, due to a decreased ability to adapt [62]. Akin to the link between natural hazards and climate change, these relationships are complex, and vary according to location and vulnerability. Additionally, several studies contest the climate's influence on conflicts and there appears to be little consensus in the scientific community [63, 64, 65]. Some have even suggested that the climate change/conflict narrative is dangerous, shifting attention away from government action and conflict resolution [66]. Understanding how any potential changes in climate and conflict will alter related disease outbreaks is important and relatively understudied. Exploring this research gap is vital in protecting vulnerable populations and resource allocation.

1.2.3 Social Change and Sustainable Development

One way to mitigate the effects of climate change on both natural hazards and conflicts is through social changes in terms of sustainable development. Despite their ambiguous nature and no clear pathway of how they will be achieved [67], the Sustainable Development Goals

(SDG) are a relatively universally accepted measure of social development [68] (Figure 1.2). Several of the risk factors stated above would be reduced through pursuit of the SDGs including WASH (SDG6), education (SDG4), poverty (SDG1) and access to healthcare (SDG3). Alleviating these risk factors would give communities a greater ability to adapt to changes and more options to ensure their health and well-being.

Figure 1.2: The United Nations Sustainable Development Goals [68].

Societal changes and a reduction in vulnerability will play a significant role in how future climate-related changes in health and disease will be experienced [69]. The most vulnerable populations are often cited as those most at risk of climate change and this includes natural hazards and conflicts, as vulnerable populations have less opportunities and available resources [70]. For example, both low levels of education and higher poverty increased the risk to the detrimental consequences of flooding in rural Bangladesh [71]. Furthermore, early warning systems are considered fundamental in mitigating the impacts of disasters on health, but inequities in gender and wealth mean they disproportionately under-serve parts of the population [72].

The uncertainty in projected socio-economic conditions is greater than environmental ones, as much of this will depend on human behaviour, which is challenging to predict. It is therefore important to incorporate a wide range of potential changes to social parameters in future

scenario projections. Statistical and mathematical modelling studies are pivotal in exploring future disease changes, as the flexibility of several modelling techniques allow many of these parameters to be considered at once, termed multi-parameter modelling. This type of modelling is important for real-world application and for understanding interactions among social and environmental factors. In the absence of these studies and regardless of their publication, continued progression towards and beyond the SDGs will be essential to improving the health and well-being of the global population.

1.3 Cholera

1.3.1 History and Global Burden of Cholera

Cholera is an ancient disease, beginning to cause outbreaks during the transition away from nomadic lifestyles and into settlements. When humans were predominately hunter-gathers, the constant movement meant ever changing water sources, making it unlikely that people would contaminate their own water. As settlements became larger and more densely populated, pollution became more likely. There is no definitive date for the first appearance of cholera but the disease seems likely to have been endemic in certain areas from as early as the 5th century, with several historical accounts of symptoms that could now be attributed to cholera [73].

As previously stated, diarrhoeal diseases are a major contributor to the global disease burden, especially in children in low- and middle-income countries [6]. Cholera contributes significantly to this burden and is a disease of global importance (Figure 1.3). Since the beginning of the seventh and ongoing pandemic in 1961, the disease is now endemic in 51 countries [74]. In 2020, over 320,000 cholera cases and 857 deaths were reported to WHO [75], while in 2022, 20 countries reported cases, exclusively in Africa and Asia [76]. The true impact of cholera though is difficult to decipher due to reporting heterogeneity, such as the impact on trade and tourism causing reporting hesitancy. Previous estimations have suggested annual cholera cases of 2.86 million and between 21,000-143,000 deaths [77, 74].

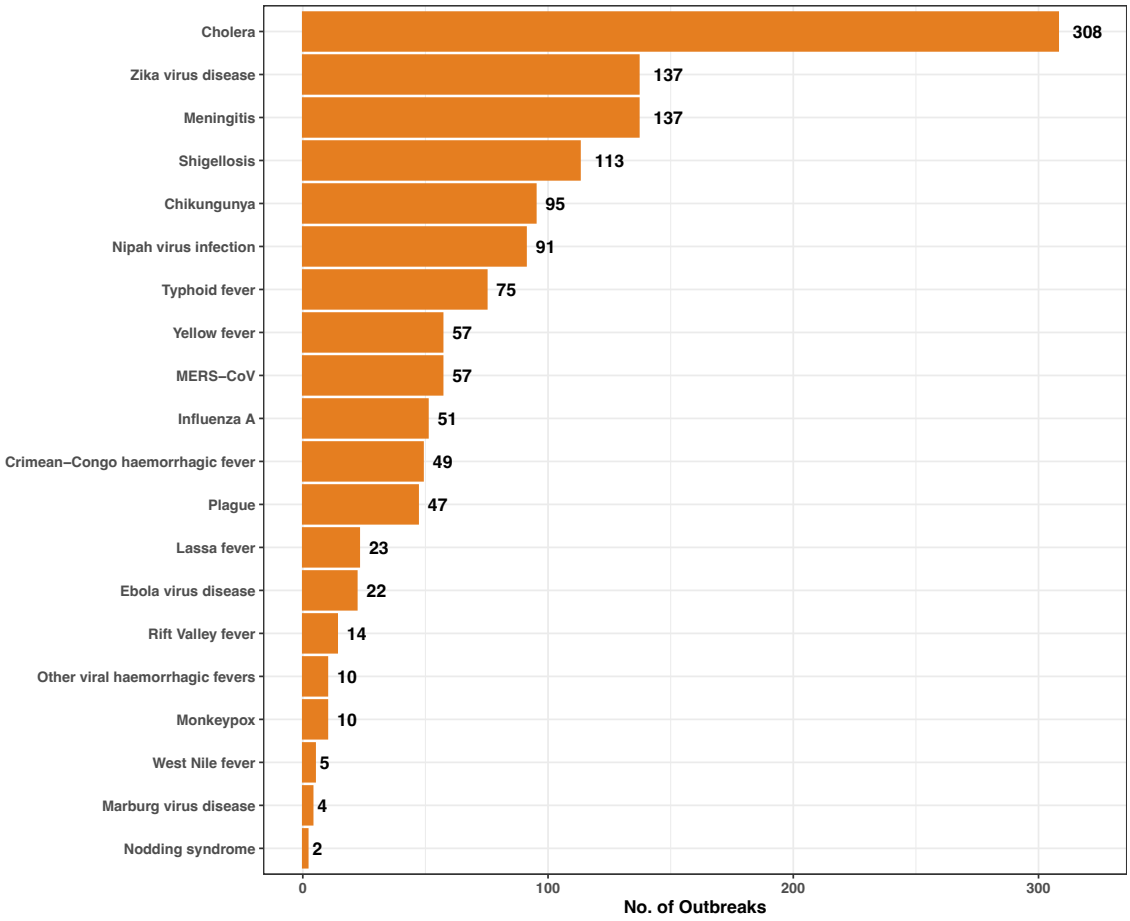


Figure 1.3: Total number of epidemics reported between 2011 and 2017 by disease (total = 1,307). License CC-BY [7].

The causative agent, *Vibrio cholerae*, is an extremely virulent gram-negative bacteria, having two outbreak causing strains, O1 and O139. The water-borne bacterial pathogen causes profuse watery diarrhoea outbreaks and in some cases vomiting [75, 78]. Persistence of the bacteria in aquatic reservoirs, the formation of biofilms and asymptomatic cases, which help sustain transmission by bacterial shedding, means those living in areas with poor access to water and hygiene are highly likely to be exposed [79]. Explosive cholera outbreaks are common due to the short incubation period (2 hours to 5 days) and the risk of rapid dehydration, especially among young children, meaning deaths can occur quickly and case fatality can reach up to 40% [80].

Effective symptom management is considered the cornerstone to cholera treatment including

oral rehydration solution (ORS) and in some instances (particularly severe cases) antibiotic use, which helps to keep outbreak mortality to the global goal of <1% [81, 36]. Cholera can be vaccinated against with the oral cholera vaccine (OCV). There are currently three licensed OCVs, all administered as two-doses taken orally 7 days to 6 weeks apart. The vaccine can be given to children from 2 years of age and more than 20 million doses have been used in mass vaccination campaigns [82]. However, vaccination is not considered the primary method for cholera eradication, only providing around 2-3 years of protection with the current vaccines available. Instead vaccination is mainly used to curb transmission in outbreak settings, along with other interventions [83, 84].

1.3.2 Disaster-related Cholera Outbreaks

Cholera outbreak frequency is linked to environmental and climatic changes and has been implicated in several post-disaster outbreaks, including floods, droughts and cyclones [85, 86, 87, 88]. Environmental mechanisms through which natural hazards cause cholera transmission are related to temperature and precipitation. Temperature helps to drive epidemics, by creating an ideal environment for the pathogen to grow in the environment and precipitation and storm water can then act as a dispersal mechanism [89]. Cholera outbreaks often occur when the disaster also results in the breakdown and damage of sanitation, hygiene and municipal waste systems.

Some studies suggest that human-induced factors are more important for cholera dynamics than climate or environmental ones [90], possibly due to the need for poor socio-economic conditions for pathogen exposure. Cholera is considered a disease of inequity, predominantly affecting the poorest and most vulnerable [91, 92], making those in conflict-affected cholera endemic countries particularly vulnerable. Prominent examples of conflict-related cholera outbreaks include those in Yemen [93] and the DRC [94].

Several of the socio-economic risk factors involved in cholera outbreaks also align with the historically identified risk factors for post-disaster outbreaks delineated above. These include poverty [95], sanitation and hygiene [96], drainage [97], water quality [98] and poor healthcare

[89]. The links between these risk factors have made cholera a common aetiological agent and significant risk for post-disaster outbreaks.

1.3.3 Cholera Outbreak Policy

Actions have been taken to reduce the global burden of cholera, both indirectly through sanitation and hygiene programmes and more specifically through the development of the Global Task Force on Cholera Control (GTFCC). The GTFCC is a global network coordinating the fight against cholera. The aim of the organisation is to significantly reduce global cholera burden and work towards eradication in many countries through the Global Roadmap. In 2018, at the 71st World Health Assembly, WHO member states passed a resolution committing to the Global Roadmap and 47 African countries adopted a regional framework in alignment with the Roadmap at the WHO Regional Committee for Africa [99].

A serious barrier to reaching the GTFCC goals is the stark reminders of the gains needed in terms of WASH development. In 2020, 2.2 billion people lacked access to safe drinking water, 3 billion people were without access to handwashing facilities and more than half of the population live without access to safe sanitation [100]. Much greater international and national commitments towards sustainable development are required to prevent erosion of progress towards the GTFCC goals and several other disease and health targets. Sustainable development in terms of poverty alleviation and the provision of WASH services would undoubtedly be cost-effective [101, 102, 103], due to the far reaching implications this would have on health and quality of life.

1.4 Previous Methodological Approaches Used to Explore Disaster-related Outbreaks

1.4.1 Outbreak Investigation

There is an absence of quantitative research in terms of natural hazards/conflicts and disease outbreaks, especially in terms of statistical and mathematical modelling. Alternatively, there is abundance of quantitative outbreak investigations (retrospective case control and cohort studies) [38, 104, 43] and serological surveys (cross-sectional and longitudinal) [105, 106, 107, 108], which are helpful in terms of deciphering immediate risk factors and disease burden. These studies give an indication of cases, deaths, case fatality ratio and risk factors, all of which are useful in understanding cholera outbreaks and provide crucial knowledge that is used throughout this thesis. One limitation of outbreak investigations is that they only provide a snapshot of a single outbreak and are less helpful in understanding outbreaks on larger temporal and spatial scales.

1.4.2 Reviews and Qualitative Research

There are several published reviews which are particularly helpful in collating the vast number of serological surveys and outbreak investigations. In particular, there are a few broader reviews evaluating natural hazards and conflicts on a global scale, which are helpful in trying to understand the frequency of specific risk factors [41, 42, 109]. Other reviews tend to focus more specifically on either a disaster type such as droughts [110] or tsunamis [111] or a specific timescale (2000-2011) [46]. These more focused reviews can be helpful in understanding a specific research question but often fall short in providing more generalisable results. However, there is a general consensus among the available reviews that infectious disease outbreak risk is heightened following disasters and several risk factors are commonly cited, as discussed above (1.1.4 Historically Identified Risk Factors).

Qualitative studies are often outbreak investigations, exploring a specific disaster and disease

outbreak, typically through a humanitarian lens [93, 112]. Previous examples of these include investigations of a hepatitis E outbreak in Nepal following an earthquake [113] and cholera outbreaks in Nigeria and Yemen during ongoing conflicts [93, 112]. These qualitative studies have the same advantages as the quantitative outbreak investigations while also providing the perspectives of those in the affected area, both in terms of government and non-governmental organisations (NGO) and the local population. A limitation of much of the qualitative research and reviews are a lack of definitions [93, 111]. What constitutes to a specific disaster or disease outbreak is key in understanding the study results and recommendations and for making comparisons between studies. Clear definitions and terminology are particularly important in disaster research, as there is much contention and disagreement within the research community [114, 115].

1.4.3 Quantitative and Modelling Research

Despite the advantages identified above in the previously used methodological approaches, research needs to move beyond this type of analysis to plan for disasters and outbreaks in the future. Collating and making use of large datasets is essential in evaluating these outbreaks on a greater temporal and spatial scale. Previous examples of quantitative modelling studies investigating natural hazards, conflicts and cholera are limited and specific examples include:

- Spatio-Temporal Clustering:
    - Flooding and cholera in Bangladesh [116]
    - Conflict and cholera in the DRC [94]
- Machine Learning:
    - Conflict and cholera in Yemen [117]
- General Additive Models:
    - Drought, floods and heatwaves and cholera in Nigeria [118]

- Poisson Regression:
    - Floods, droughts and cholera in sub-Saharan Africa [119]

Several of the previous modelling approaches either do not account for socio-economic conditions [116, 94, 118, 117] or attempt to offset one factor (e.g., Human Development Index [119]). Few try to account for the large number of potentially influential factors in post-disaster outbreaks [109]. Issues also arise when using an index as a sole measure of socio-economic vulnerability, as these metrics include several variables and determining which are the most important is not possible from the single value they provide. Very few studies also try to project cholera risk and outbreaks [117], instead taking a qualitative approach to discuss possible future changes [11].

The limited quantitative modelling work on natural hazards/conflict and cholera completed to date gives abundant opportunity to use a combination of novel methodological techniques and previously established methodology in new applications. The work presented here will consider social changes central to the research and a wide range of parameters and future conditions will be evaluated. The modelling will further inform how disasters impact infectious disease outbreaks historically, currently and projected into the future, encompassing as many potentially influential factors in terms of disease and disasters as possible, to gain a greater understanding of the full scope of potential risk factors (Figure 1.4).

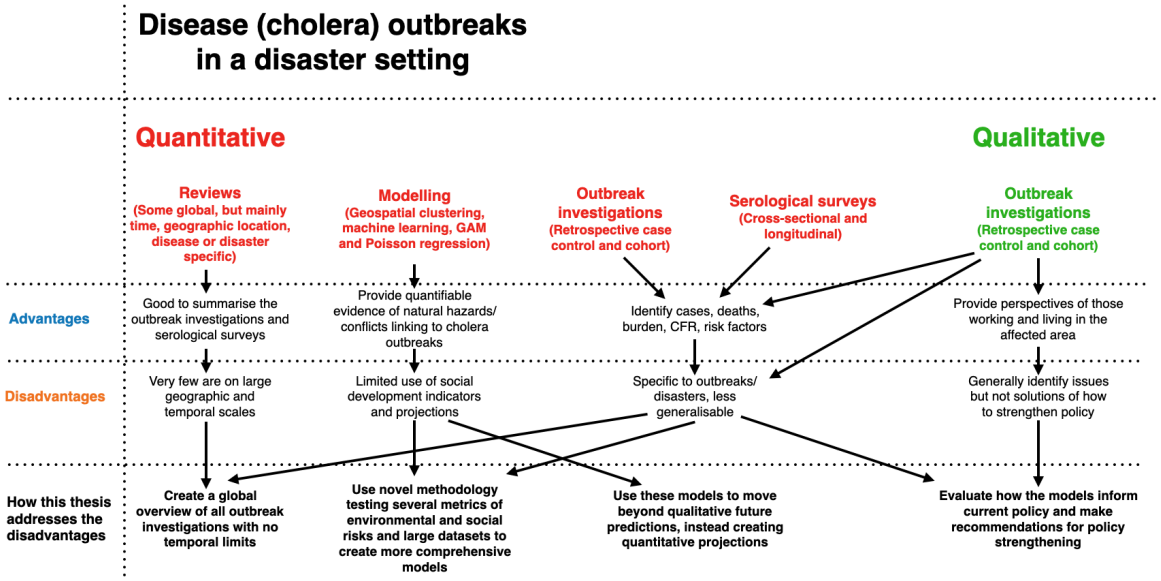


Figure 1.4: Schematic diagram of the current research landscape for post-disaster disease outbreaks and how this thesis addresses the current limitations and research gaps. GAM - General Additive Models, CFR - Case Fatality Ratio.

1.5 Motivation and Objectives

1.5.1 Primary Motivation

The primary motivation of this thesis is to understand why disaster-related outbreaks occur despite longstanding experience of disaster mitigation and adaptation. Populations experience significant mortality and morbidity in fragile settings, which society can reduce, but too often does not. The devastating impacts of disasters are still repeatably witnessed, even though natural hazards and conflicts have occurred throughout history. Post-disaster outbreaks are a global issue, and no community, regardless of economic and political stability, appear unaffected. The aim is to use this enhanced knowledge of potential risk factors and thresholds for these outbreaks, to understand where current disaster adaptation fails in preventing outbreaks and what more can be done.

1.5.2 Secondary Motivation

The secondary motivation is to use a disease of global public health importance, such as cholera, as a case study in understanding these outbreaks. Cholera outbreaks, like disasters, are not a new phenomenon, with documented outbreaks of the disease dating back to 1817. The pathogen emerged out of the Ganges Delta and has since caused seven pandemics, with the current and most persistent pandemic (7th) showing no signs of regressing. The disease causes serious ethical issues, due to its large burden and mortality on young children in poor communities, despite having simple preventative interventions e.g., sanitation and handwashing. The drivers of cholera are complex and more research and focus is needed to understand them better, this thesis will aim to address these drivers in the context of post-disaster outbreaks.

1.5.3 Tertiary Motivation

A tertiary motivation will be to evaluate how sustainable development can yield co-benefits for disease control and prevention, particularly for cholera and disaster settings. It could be argued that the answers to post-disaster disease outbreaks are already available, and the continuation of outbreaks is instead a failure to act. Whether this is the case or not, the longevity of the issues of post-disaster outbreaks and cholera may have resulted in action fatigue and it is important that attention is not completely shifted away from these issues. Development will empower communities and give people the control to adapt to disasters and climate change and reduce the need for humanitarian aid. The more scientific evidence showing the need to address specific inequities, the less justification there will be for inaction.

1.5.4 Objectives

The extensive and original research presented here will highlight the continued support and attention needed for post-disaster disease outbreaks and cholera. As problems persist, attention can be lost, but with such high levels of mortality and morbidity this cannot be the case. A

goal of this research is to re-focus this attention by increasing awareness and understanding of the risk factors through the following objectives:

1. Create a comprehensive review of post-disaster disease outbreaks, collating all previous quantitative research to identify commonly reports geographic areas, disasters, aetiological agents and risk factors (Chapter 2 and publications 1-3 & 7).

2. Use novel methodological approaches and datasets to gain a greater understanding of cholera outbreak risk factors and thresholds in a disaster setting, focusing on conflicts and droughts, to create more generalisable results than previous research (Chapter 3-5, publications 4-6 & 8).

3. Evaluate a range of social and development indicators in the fitting of all models used, to create a more comprehensive model of cholera outbreak risk in a disaster setting (Chapter 3 & 5, publications 4 & 8).

4. Apply the models to make quantitative predictions and projections of future cholera outbreak risk in the context of natural hazards and conflicts and establish ideal future scenarios to reduce this risk (Chapter 3, 5 & 6, publications 4 & 8).

5. Evaluate the results and conclusions from the modelling work completed here to evaluate the achievability of global cholera targets (Chapter 6, publication 9).

1.6 Publications

Several publications have been produced through the research presented in this thesis and are listed below and at the start of each chapter:

1. Charnley GEC, Kelman I, Gaythorpe KAM, Murray KA. Understanding the risks for post-disaster infectious disease outbreaks: a systematic review protocol. *BMJ Open* 2020;10:e039608.

2. Charnley GEC, Kelman I, Gaythorpe KAM, Murray KA. Traits and risk factors of post-disaster infectious disease outbreaks: a systematic review. *Scientific Reports* 2021;11:5616.

3. Charnley GEC, Kelman I, Murray KA. Drought-related cholera outbreaks in Africa and the implications for climate change: a narrative review. *Pathogens and Global Health* 2022;116(1):3-12.

4. Charnley GEC, Kelman I, Green N, Hinsley W, Gaythorpe KAM, Murray KAM. Exploring relationships between drought and epidemic cholera in Africa using generalised linear models. *BMC Infectious Diseases* 2021;21:1177.

5. Charnley GEC, Kelman I, Gaythorpe KAM, Murray KA. Accessing sub-national cholera epidemiological data for Nigeria and the Democratic Republic of Congo during the seventh pandemic. *BMC Infectious Diseases* 2022;22:288.

6. Charnley GEC, Jean K, Kelman I, Gaythorpe KAM, Murray KA. Using self-controlled case series to understand the relationship between conflict and cholera in Nigeria and the Democratic Republic of Congo. *Emerging Infectious Diseases* 2022;28:2472-2481.

7. Harris M, Charnley GEC. Disaster Risk Management: A Resilient Health System. In: Eslamian, S., Eslamian, F. (eds) Disaster Risk Reduction for Resilience. Springer, Cham, 2022.

8. Charnley GEC, Yennan S, Ochu C, Kelman I, Gaythorpe KAM, Murray KA. Investigating the impact of social and environmental extremes on cholera time varying reproduction number in Nigeria. *PLoS Global Public Health* 2022;2(12):e0000869

9. Charnley GEC, Yennan S, Ochu C, Kelman I, Gaythorpe KAM, Murray KA. Cholera past and future in Nigeria: are the Global Task Force on Cholera Control's 2030 targets achievable? *medRxiv* 2022;https://doi.org/10.1101/2022.12.06.22283154 [pre-print].

1.7 Data & Code Availability

Information on both the public and private datasets used here are detailed throughout the thesis, including how they were obtained, curated and how to access them (or request access if not publicly available).

All code used here for the analyses and data visualisation are available in a Github repository: https://github.com/GinaCharnley/Thesis. Details of the license are available in the repository (MIT License). All figures were created in R package "ggplot" [120], unless stated otherwise in the figure caption.

The shapefiles used to create any data visualisations involving maps were all taken from freely available data sources and the sources and licensing agreements are detailed below. All licensing was under Creative Commons, allowing them to be shared and adapted.

- World & Africa - API, Thematicmappingorg Thematic Mapping. 2009. https://thematicmapping.org/downloads/world_borders.php. CC-BY SA
- Nigeria - HDX, Nigeria - Subnational Administrative Boundaries. 2020. https://data.humdata.org/dataset/cod-ab-nga. CC-BY IGO
- DRC - HDX, Democratic Republic of the Congo - Subnational Administrative Boundaries. 2019. https://data.humdata.org/dataset/cod-ab-cod. CC-BY IGO

References

[1] H. Richie and R. Roser. *Cause of death*. en. 2021. URL: https://ourworldindata.org/causes-of-death.

[2] T. Vos et al. "Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019". In: *Lancet* 396.10258 (2020), pp. 1204–1222.

[3] C.J.L. Murray and A.D. Lopez. “Measuring the global burden of disease”. In: *New Eng. J. Med.* 369.5 (2013), pp. 448–457.

[4] R. Roser and H. Richie. *Burden of Disease*. en. 2021. URL: https://ourworldindata.org/burden-of-disease.

[5] World Health Organization. *Disease outbreaks*. en. 2020. URL: https://www.who.int/.

[6] World Health Organization. *The top 10 causes of death*. en. 2018. URL: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death.

[7] Wellcome Trust. *The cost of not preparing for infectious diseases*. en. 2018. URL: https://wellcome.org/news/cost-of-not-preparing-for-infectious-diseases.

[8] Institute for Health Metrics and Evaluation. *Global Burden of Disease (GBD)*. en. 2019. URL: https://www.healthdata.org/gbd/2019.

[9] M.L. Cohen. “Changing patterns of infectious disease”. en. In: *Nature* 406 (2000), pp. 762–767.

[10] J. Leaning and D. Guha-Sapir. “Natural disasters, armed conflict, and public health”. en. In: *New Engl. J. Med.* 369 (2013), pp. 1836–1842.

[11] A.J. McMichael and R. Beaglehole. “The changing global context of public health”. In: *Lancet* 356.9228 (2000), pp. 495–499.

[12] B. Wisner, J.C. Gaillard, and I. Kelman. “Framing disaster: Theories and stories seeking to understand hazards, vulnerability and risk”. In: *The Routledge handbook of hazards and disaster risk reduction*. Routledge, 2012, pp. 18–33.

[13] I. Kelman. “Lost for words amongst disaster risk science vocabulary?” en. In: *Int. J. Disast. Risk Sc.* 9 (2018), pp. 281–291.

[14] K. Eshghi and R.C. Larson. “Disasters: lessons from the past 105 years”. en. In: *Disaster Prev. Manag.* (2008).

[15] M.I. Shaluf. “An overview on disasters”. nl. In: *Disaster Prev. Manag.* 16 (2007), pp. 687–703.

[16] Intergovernmental Panel on Climate Change. *Climate Change 2014 Synthesis Report Fifth Assessment Report Future Climate Changes, Risks and Impacts*. 2014. URL: https://www.ipcc.ch/site/assets/uploads/2018/05/SYR_AR5_FINAL_full_wcover.pdf.

[17] C.C. Hammer, J. Brainard, and P.R. Hunter. "Risk factors and risk factor cascades for communicable disease outbreaks in complex humanitarian emergencies: A qualitative systematic review". en. In: *BMJ Glob. Health* 3 (2018), p. 000647.

[18] E. Helderop and T.H. Grubesic. "Hurricane storm surge in Volusia County, Florida: evidence of a tipping point for infrastructure damage". In: *Disasters* 43.1 (2019), pp. 157–180.

[19] R.E. Berggren and T.J. Curiel. "After the storm—health care infrastructure in post-Katrina New Orleans". In: *New Engl. J. Med.* 354.15 (2006), pp. 1549–1552.

[20] J.M. Nichols and J.E. Beavers. "World earthquake fatalities from the past: implications for the present and future". In: *Nat. Hazards Rev.* 9.4 (2008), pp. 179–189.

[21] O.W. Morgan et al. "Mass fatality management following the South Asian tsunami disaster: case studies in Thailand, Indonesia, and Sri Lanka". In: *PLoS Med.* 3.6 (2006), e195.

[22] S. Karmakar et al. "Post-earthquake outbreak of rotavirus gastroenteritis in Kashmir (India): An epidemiological analysis". In: *Public Health* 122.10 (2008), pp. 981–989.

[23] L.D. Londe et al. "Flood-related leptospirosis outbreaks in Brazil: perspectives for a joint monitoring by health services and disaster monitoring centers". In: *Nat. Hazards* 84.2 (2016), pp. 1419–1435.

[24] E.L. Yee. "Widespread outbreak of norovirus gastroenteritis among evacuees of Hurricane Katrina residing in a large "megashelter" in Houston, Texas: lessons learned for prevention". en. In: *Clin. Infect. Dis.* 44 (2007), pp. 1032–1039.

[25] Centres for Disease Control and Prevention. *Lesson 1: Introduction to Epidemiology*. en. 2012. URL: https://www.cdc.gov/csels/dsepd/ss1978/lesson1/section11.html.

[26] F.D. Orata, P.S. Keim, and Y. Boucher. "The 2010 Cholera Outbreak in Haiti: How Science Solved a Controversy". en. In: *PLoS Pathog.* 10.4 (2014), e1003967. URL: https://doi.org/10.1371/journal.ppat.1003967.

[27] W.S. Al-Salem. "Cutaneous leishmaniasis and conflict in Syria". en. In: *Emerg. Infect. Dis.* 22 (2016), p. 931.

[28] F.A. Dureab et al. "Yemen: cholera outbreak and the ongoing armed conflict". en. In: *J. Infect. Dev. Countr.* 12 (2018), pp. 397–403.

[29] A.J. Bellamy. "Mass Atrocities and Armed Conflict: Links, Distinctions, and Implications for the Responsibility to Prevent". In: *Policy Analysis Brief* (2011).

[30] C.R. Wells. "The exacerbation of Ebola outbreaks by conflict in the Democratic Republic of the Congo". en. In: *PNAS* 116 (2019), pp. 24366–24372.

[31] D. Peprah et al. "Perceptions of oral cholera vaccine and reasons for full, partial and non-acceptance during a humanitarian crisis in South Sudan". In: *Vaccine* 34.33 (2016), pp. 3823–3827.

[32] P. Vinck et al. "Institutional trust and misinformation in the response to the 2018–19 Ebola outbreak in North Kivu, DR Congo: a population-based survey". In: *Lancet Infect. Dis.* 19.5 (2019), pp. 529–536.

[33] L. Barsky, J. Trainor, and M. Torres. *Disaster realities in the aftermath of Hurricane Katrina: Revisiting the looting myth.* Natural Hazards Research and Applications Information Center, University of Colorado, 2006.

[34] Dwight T.C.K. et al. "Prosociality and hoarding amid the COVID-19 pandemic: A tale of four countries". In: *J. Appl. Soc. Psychol.* 32.3 (2022), pp. 507–520.

[35] D.A. Walton and L.C. Ivers. "Responding to cholera in post-earthquake Haiti". en. In: *New Engl. J. Med.* 364 (2011), pp. 3–5.

[36] R.V. Tauxe et al. "Epidemic cholera in Mali: high mortality and multiple routes of transmission in a famine area". en. In: *Epidemiol. Infect.* 100.2 (1988), pp. 279–289.

[37] M. Baqir. “Infectious diseases in the aftermath of monsoon flooding in Pakistan”. en. In: *Asian Pac. J. Trop. Biomed.* 2 (2012), pp. 76–79.

[38] F. Fiasca. “Bacterial Meningitis Hospitalizations after the 2009 L’Aquila Earthquake: A Retrospective Observational Study”. en. In: *Asian J. Epidemiol.* 11 (2018), pp. 46–51.

[39] F. Valente. “Massive outbreak of poliomyelitis caused by type-3 wild poliovirus in Angola in 1999”. en. In: *Bull. World Health Organ.* 78 (2000), pp. 339–346.

[40] B. McPake et al. “Ebola in the context of conflict affected states and health systems: case studies of Northern Uganda and Sierra Leone”. en. In: *Confl. Health* 9 (2015), p. 23.

[41] M. Gayer et al. “Conflict and emerging infectious diseases”. en. In: *Emerg. Infect. Dis.* 13 (2007), p. 1625.

[42] J.T. Watson, M. Gayer, and M.A. Connolly. “Epidemics after natural disasters”. en. In: *Emerg. Infect. Dis.* 13, 1 (2007).

[43] K.A. Alghazali et al. “Dengue outbreak during ongoing civil war, Taiz, Yemen”. In: *Emerg. Infect. Dis.* 25.7 (2019), p. 1397.

[44] D. Crichton. *The Risk Triangle*. London: Tudor Rose, 1999, pp. 102–103.

[45] K. Hewitt. *Interpretations of calamity: From the viewpoint of human ecology*. Routledge, 2019.

[46] I.K. Kouadio et al. “Infectious diseases following natural disasters: prevention and control measures”. en. In: *Expert Rev. Anti-Infect.* 10 (2012), pp. 95–104.

[47] O.C. Okunlola and I.G. Okafor. “Conflict–Poverty Relationship in Africa: A Disaggregated Approach”. en. In: *J. Interdiscip. Econ.* (2020), pp. 1–26.

[48] P. Loebach and K. Korinek. “Disaster vulnerability, displacement, and infectious disease: Nicaragua and Hurricane Mitch”. en. In: *Popul. Environ.* 40 (2019), pp. 434–455.

[49] N. Watts. “The 2019 report of The Lancet Countdown on health and climate change: ensuring that the health of a child born today is not defined by a changing climate”. en. In: *Lancet* 394 (2019), pp. 1836–1878.

[50] D. Coumou and S. Rahmstorf. “A decade of weather extremes”. en. In: *Nat. Clim. Change* 2 (2012), pp. 491–496.

[51] I. Kelman. “Climate change and the Sendai framework for disaster risk reduction”. en. In: *Int. J. Disast. Risk Sci.* 6 (2015), pp. 117–127.

[52] T.F. Stocker. *Climate change 2013: The physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change*. en. 2013.

[53] V. Aalst and K. Maarten. “The impacts of climate change on the risk of natural disasters”. In: *Disasters* 30.1 (2006), pp. 5–18.

[54] Z. Hausfather et al. “Evaluating the performance of past climate model projections”. In: *Geophys. Res. Lett.* 47.1 (2020), e2019GL085378.

[55] C.B. Field et al. *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*. Cambridge: Cambridge University Press, 2012, p. 582. URL: https://www.ipcc.ch/report/managing-the-risks-of-extreme-events-and-disasters-to-advance-climate-change-adaptation/.

[56] M.B. Burke et al. “Warming increases the risk of civil war in Africa”. In: *PNAS* 106.49 (2009), pp. 20670–20674.

[57] D.C. Bowles, C.D. Butler, and N. Morisetti. “Climate change, conflict and health”. In: *J. R. Soc. Med.* 108.10 (2015), pp. 390–395.

[58] N.P. Gleditsch. “Whither the weather? Climate change and conflict”. In: *J. Peace Res.* 49.1 (2012), pp. 3–9.

[59] G.E.C. Charnley, I. Kelman, and K.A. Murray. “Drought-related cholera outbreaks in Africa and the implications for climate change: a narrative review”. en. In: *Pathog. Glob. Health* (2021), pp. 1–10.

[60] C.P. Kelley et al. “Climate change in the Fertile Crescent and implications of the recent Syrian drought”. en. In: *PNAS* 112 (2015), pp. 3241–3246.

[61] S.M. Hsiang, K.C. Meng, and M.A. Cane. “Civil conflicts are associated with the global climate”. en. In: *Nature* 476 (2011), pp. 438–441.

[62] W.N. Adger et al. *Human Security*. Cambridge: Cambridge University Press, 2014, pp. 755–791. URL: https://www.ipcc.ch/report/ar5/wg2/human-security/.

[63] H. Buhaug. “Climate not to blame for African civil wars”. en. In: *PNAS* 107 (2010), pp. 16477–16482.

[64] J. Selby et al. “Climate change and the Syrian civil war revisited”. en. In: *Polit. Geogr.* 60 (2017), pp. 232–244.

[65] S.M. Hsiang and M. Burke. “Climate, conflict, and social stability: what does the evidence say?” In: *Clim. Change* 123.1 (2014), pp. 39–55.

[66] R.T. Slettebak. “Don’t blame the weather! Climate-related natural disasters and civil conflict”. In: *J. Peace Res.* 49.1 (2012), pp. 163–176.

[67] F. Biermann et al. “Scientific evidence on the political impact of the Sustainable Development Goals”. In: *Nature Sustain.* (2022), pp. 1–6.

[68] United Nations. *The 17 Goals*. en. https://sdgs.un.org/goals. 2015.

[69] N. Watts et al. “Health and climate change: policy responses to protect public health”. In: *Lancet* 386.10006 (2015), pp. 1861–1914.

[70] Notre Dame Global Adaptation Initiative. *ND-GAIN Country Index*. en. 2019. URL: https://gain.nd.edu/our-work/country-index/.

[71] M.I. Rayhan. “Assessing poverty, risk and vulnerability: a study on flooded households in rural Bangladesh”. In: *J. Flood Risk Manag.* 3.1 (2010), pp. 18–24.

[72] L. Clarke and V. Le Masson. “Shocks, stresses and universal health coverage”. In: *ODI Working Paper* (2017).

[73] S.L. Kotar and J.E. Gessler. *Cholera: a Worldwide history*. McFarland, 2014.

[74] M. Ali et al. “Updated global burden of cholera in endemic countries”. en. In: *PLoS Neglect. Trop. Dis.* 9 (2015), p. 0003832.

[75] World Health Organization. *Cholera*. en. 2020. URL: https://www.who.int/news-room/fact-sheets/detail/cholera.

[76] European Centre for Disase Control and Prevention. *Cholera worldwide overview*. en. 2022. URL: https://www.ecdc.europa.eu/en/all-topics-z/cholera/surveillance-and-disease-data/cholera-monthly.

[77] M. Ali et al. "The global burden of cholera". en. In: *Bull. World. Health Organ.* 90 (2012), pp. 209–218.

[78] Y. Germani et al. "Emergence of cholera in the Central African Republic". en. In: *Eur. J. Clin. Microbiol. Infect. Dis.* 17.12 (1998).

[79] E.J. Nelson et al. "Cholera transmission: the host, pathogen and bacteriophage dynamic". In: *Nat. Rev. Microbiol.* 7.10 (2009), pp. 693–702.

[80] A.A. King et al. "Inapparent infections and cholera dynamics". en. In: *Nature* 454 (2008), pp. 877–80.

[81] J. Mendelsohn and T. Dawson. "Climate and cholera in KwaZulu-Natal, South Africa: The role of environmental factors and implications for epidemic preparedness". en. In: *Int. J. Hyg. Environ. Health* 211.1-2 (2008), pp. 156–162.

[82] World Health Organization. *Immunization, Vaccines and Biologicals. Cholera vaccine*. en. 2022. URL: https://www.who.int/teams/immunization-vaccines-and-biologicals/diseases/cholera.

[83] P. Calain et al. "Can oral cholera vaccination play a role in controlling a cholera outbreak?" In: *Vaccine* 22.19 (2004), pp. 2444–2451.

[84] F. Qadri et al. "Efficacy of a single-dose, inactivated oral cholera vaccine in Bangladesh". In: *N. Engl. J. Med.* 374.18 (2016), pp. 1723–1732.

[85] G.C. De Magny et al. "Cholera outbreak in Senegal in 2005: was climate a factor?" en. In: *PLoS ONE* 7.8 (2012).

[86] S. Rebaudet et al. “Environmental determinants of cholera outbreaks in inland Africa: a systematic review of main transmission foci and propagation routes”. en. In: *J. Infect. Dis.* 208.Suppl 1 (2013).

[87] R. Reyburn et al. “Climate variability and the outbreaks of cholera in Zanzibar, East Africa: a time series analysis”. en. In: *Am. J. Trop. Med. Hyg.* 84.6 (2011), pp. 862–869.

[88] A. Palit and P. Batabyal. “Toxigenic Vibrio cholerae from environmental sources associated with the cholera outbreak after ‘AILA’cyclone in West Bengal, India”. In: *Lett. Appl. Microbiol.* 51.2 (2010), pp. 241–243.

[89] D. Olago et al. “Climatic, socio-economic, and health factors affecting human vulnerability to cholera in the Lake Victoria basin, East Africa”. en. In: *Ambio* 36.4 (2007), pp. 350–358.

[90] F.X. Weill et al. “Genomic history of the seventh pandemic of cholera in Africa”. en. In: *Science* 358.6364 (2017), pp. 785–789.

[91] G.C. Leckebusch and A.F. Abdussalam. “Climate and socioeconomic influences on interannual variability of cholera in Nigeria”. it. In: *Health Place* 34 (2015), pp. 107–17.

[92] N. Anbarci, M. Escaleras, and C.A. Register. “From cholera outbreaks to pandemics: the role of poverty and inequality”. en. In: *Am. Econ.* 57 (2012), pp. 21–31.

[93] P. Spiegel et al. “Responding to epidemics in large-scale humanitarian crises: a case study of the cholera response in Yemen, 2016–2018”. In: *BMJ Glob. Health* 4.4 (2019), e001709.

[94] H.C.N. Kayembe et al. “Modalities and preferred routes of geographic spread of cholera from endemic areas in eastern Democratic Republic of the Congo”. In: *PloS ONE* 17.2 (2022), e0263160.

[95] A. Talavera and E.M. Perez. “Is cholera disease associated with poverty?” en. In: *J. Infect. Dev. Ctries.* 3.06 (2009), pp. 408–411.

[96] L. Mari et al. “Modelling cholera epidemics: the role of waterways, human mobility and sanitation”. en. In: *J. R. Soc. Interface.* 9.67 (2012), pp. 376–388.

[97] S. Sasaki et al. “Impact of drainage networks on cholera outbreaks in Lusaka, Zambia”. en. In: *Am. J. Public Health* 99.11 (2009), pp. 1982–1987.

[98] R. Ranjbar et al. “A cholera outbreak associated with drinking contaminated well water”. en. In: *Arch. Iran. Med.* 14.5 (2011), pp. 339–340.

[99] Global Task Force on Cholera Control. *Roadmap 2030*. en. 2020. URL: https://www.gtfcc.org/about-gtfcc/roadmap-2030/.

[100] U.N.I.C.E.F. *Water, Sanitation and Hygiene (WASH)*. en. 2020. URL: https://www.unicef.org/wash.

[101] T.O. Tengs et al. “Five-hundred life-saving interventions and their cost-effectiveness”. In: *Risk Anal.* 15.3 (1995), pp. 369–390.

[102] P.J. McEwan. “Cost-effectiveness analysis of education and health interventions in developing countries”. In: *J. Dev. Eff.* 4.2 (2012), pp. 189–213.

[103] T. Yates et al. “Efficacy and effectiveness of water, sanitation, and hygiene interventions in emergencies in low-and middle-income countries: a systematic review”. In: *Waterlines* (2018), pp. 31–65.

[104] T. Fredrick et al. “Cholera outbreak linked with lack of safe water supply following a tropical cyclone in Pondicherry”. en. In: *J. Health Popul. Nutr.* 33 (2012), p. 31.

[105] S. Pal et al. “An outbreak of hepatitis A virus among children in a flood rescue camp: a post-disaster catastrophe”. In: *Indian J. Med. Microbiol.* 34.2 (2016), pp. 233–236.

[106] S. Pal et al. “Post-disaster outbreak of scrub typhus in Sub-Himalayan region of Uttarakhand”. In: *J. Acad. Clin. Microbiol.* 18.2 (2016), p. 95.

[107] C.Y. Lin et al. “Serological investigation to identify risk factors for post-flood infectious diseases: a longitudinal survey among people displaced by Typhoon Morakot in Taiwan”. In: *BMJ Open* 5.5 (2015), e007008.

[108] World Health Organization. “Leptospirosis, India: report of the investigation of a post-cyclone outbreak in Orissa, November 1999”. In: *Weekly Epidemiological Record = Relevé épidémiologique hebdomadaire* 75.27 (2000), pp. 217–223.

[109] T. Alcayna et al. “Climate-sensitive disease outbreaks in the aftermath of extreme climatic events: A scoping review”. In: *One Earth* 5 (2022), pp. 336–350.

[110] T. Asmall et al. “The adverse health effects associated with drought in Africa”. In: *Sci. Total Environ.* 793 (2021), p. 148500.

[111] A. Wilder-Smith. “Tsunami in South Asia: what is the risk of post-disaster infectious disease outbreaks?” en. In: *Ann. Acad. Med. Singap.* 34 (2005), p. 625.

[112] M.C. Ngwa. “The multi-sectorial emergency response to a cholera outbreak in internally displaced persons camps in Borno state”. en. In: *Health* 5 (2020), p. 002000.

[113] B. Basnyat et al. “Nepali earthquakes and the risk of an epidemic of hepatitis E”. In: *Lancet* 385.9987 (2015), pp. 2572–2573.

[114] R. Staupe-Delgado. “Analysing changes in disaster terminology over the last decade”. In: *Int. J. Disaster Risk Red.* 40 (2019), p. 101161.

[115] L. Mayner and P. Arbon. “Defining disaster: The need for harmonisation of terminology”. In: *Australas. J. Disaster Trauma Stud.* 19 (2015).

[116] M. Carrel et al. “Spatio-temporal clustering of cholera: The impact of flood control in Matlab, Bangladesh, 1983–2003”. In: *Health Place* 15.3 (2009), pp. 771–782.

[117] R. Badkundri et al. “Forecasting the 2017-2018 Yemen cholera outbreak with machine learning”. In: *arXiv preprint: 1902.06739* (2019).

[118] A.F. Abdussalam. “Modelling the Climatic Drivers of Cholera Dynamics in Northern Nigeria Using Generalised Additive Models”. en. In: *Int. J. Geog. Environ. Manag.* 2.1 (2016), pp. 84–97.

[119] A. Rieckmann et al. “Exploring droughts and floods and their association with cholera outbreaks in sub-Saharan Africa: a register-based ecological study from 1990 to 2010”. en. In: *Am. J. Trop. Med. Hyg.* 98.5 (2018), pp. 1269–1274.

[120] Hadley W. *ggplot2: Elegant Graphics for Data Analysis*. en. 2016. URL: https://ggplot2.tidyverse.org.

Chapter 2

A Systematic Review of Traits and Risk Factors of Post-disaster Infectious Disease Outbreaks

Dissemination

An extended version of the methods for this chapter is published at:

Charnley GEC, Kelman I, Gaythorpe KAM, Murray KA. Understanding the risks for post-disaster infectious disease outbreaks: a systematic review protocol. *BMJ Open* 2020;10:e039608.

A modified version of the full chapter is published at:

Charnley GEC, Kelman I, Gaythorpe KAM, Murray KA. Traits and risk factors of post-disaster infectious disease outbreaks: a systematic review. *Scientific Reports* 2021;11:5616.

Abstract

Infectious disease outbreaks are increasingly recognised as events that exacerbate impacts or prolong recovery following disasters. Yet, our understanding of the frequency, geography, characteristics and risk factors of post-disaster disease outbreaks globally is lacking. This limits the extent to which disease outbreak risks can be prepared for, monitored and responded to following disasters. Here, a global systematic review of post-disaster outbreaks was conducted and found that outbreaks linked to conflicts and hydrological events were most frequently reported, and most often caused by bacterial and water-borne agents. Lack of adequate WASH facilities and poor housing were commonly reported risk factors. Additionally, displacement through infrastructure damage, can lead to risk cascades for disease outbreaks; however, displacement can also be an opportunity to remove people from danger and ultimately protect health. The results shed new light on post-disaster disease outbreaks and their risks. Understanding these risk factors and cascades, could help improve future region-specific disaster risk reduction.

2.1 Introduction

Despite reports of disaster-related disease outbreaks, few studies have systematically reviewed or quantified such events or their associated risk factors on a global scale. The knowledge these reviews could provide would be helpful in resource allocation in disaster risk reduction activities and to guide new areas of research. Previous research on post-disaster disease outbreaks has for the most part resulted in the collation of individual examples over specific time scales [1], geographic areas [2] or focused on a certain disaster [3], resulting in limited generalisable results. In Chapter 2, this research gap is addressed by creating the first unified and comprehensive review and the results used to identify potential hypotheses for future quantitative analysis.

Disaster-related outbreaks are a product of risk factors created or exacerbated by the disaster, particularly if these risk factors are not prepared for or managed effectively. Here, a risk factor is defined as a clear mechanism that contributed to the disease outbreak and previously reported risks were discussed in Chapter 1 (1.1.4 Historically Identified Risk Factors), e.g., WASH [4],

disease vector changes [5], housing and shelter [6] and healthcare [7]. These risk factors are often linked and made worse by population displacement, especially when adequate facilities are not provided [8, 9]. Displacement can also increase the number of risk factors involved and the likelihood of risk factor cascades, making the cause of the outbreak difficult to ascertain and therefore control [10].

The aim of Chapter 2 is to gain a global overview of post-disaster disease outbreaks and their reported risk factors with no temporal or geographic limitations. The approach will enable the identification of links, if any, between certain hazards, vulnerabilities, disasters, geographic regions and aetiological agents. The specific objectives of the review are to:

1. Provide a global overview of infectious disease outbreaks that occurred in a post-disaster (disasters involving either natural hazard or armed conflict) setting, to show disaster types, geographic areas affected and outbreak aetiologies.
2. Examine the risk factors that lead to these outbreaks and how they may link to form cascades.
3. Use these links to identify areas of future research.

2.2 Methods

Systematic reviews sit at the top of the evidence hierarchy for medical research, considered both highly filtered and having low levels of bias [11]. Systematic reviews must first formulate questions, then appraise relevant studies on their quality and finally summarise the evidence found [12]. The review here most closely aligns to an aetiology or risk systematic review, which are used to determine to what degree a relationship exists between an exposure and a health outcome. In order to achieve this, the review aimed to outline the exposure, disease and health outcome of interest, the population and its location, and the study period where relevant [13].

The review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) 2015 checklist [14] (Supplementary Table 2.1) and were guided by the methodolog-

ical approach delineated by Khan *et al.* [15]. The framework was set out to follow five stages: (1) framing the questions, (2) identifying relevant work, (3) assessing study quality, (4) summarising the evidence and (5) interpreting the findings. The full methodological protocol used in this review underwent peer review prior to commencement of this work [16], with the key components summarised here.

2.2.1 Stage 1: Framing the Research Questions

After preliminary research on natural hazards and armed conflicts and their risk factors for communicable disease outbreaks, it became apparent that quantification of these contextual outbreaks and their risks was insufficient to gain a clear global understanding of the issue. Due to this deficiency, the review questions were defined as follows:

1. Which pathogens, disasters, global changes and geographic areas are commonly implicated in outbreaks in a post-disaster setting?

2. Which risk factors are important in causing post-disaster disease outbreaks and how are they potentially linked to form cascades?

2.2.2 Stage 2: Identifying Relevant Work

The following electronic databases were searched; MEDLINE, Embase and Global Health, but grey literature was not included. Reference lists of selected papers and reviews were screened for relevant research (snowballing) and subjected to the same screening process. Both key and medical subject heading (MeSH) terms varied depending on the database and were related to; (1) natural hazards, (2) armed conflict and (3) infectious disease outbreaks (Supplementary Table 2.2). No standard definitions for natural hazards, armed conflicts and disease outbreaks were set, as this may have excluded important studies, along with any specific risk factors. No temporal or geographic limits were set and no specific risk factors searched to avoid bias in the search results. Electronic database searching ceased in June 2020, so any relevant literature

retrieved after this date were excluded. The results and references were imported and managed in a Microsoft Excel spreadsheet [16].

Along with broad terms for outbreaks, specific diseases as identified by WHO [17] as common infectious disease outbreaks following disasters were also searched (diarrhoeal disease, hepatitis A & E, leptospirosis, measles, meningitis, acute respiratory infection, malaria, dengue, tetanus and coccidiomycosis), along with commonly reported diseases identified from preliminary scoping searches (e.g., viral heamorrhagic fevers, polio, leishmaniasis and causal agents of diarrhoeal disease including cholera, typhoid and dysentery). Despite evidence for contextual increases [18, 19], HIV, hepatitis B, hepatitis C and tuberculosis were not searched/included, as they were considered as not capable of causing acute outbreaks but instead more chronic disease and have a wide range of social implications beyond the scope of this review. Soft tissue injuries, wound infections, inhaled fungal spores and aspiration pneumonia (tsunami lung) were also not included. Such infections would only impact those that had open wounds and/or exposure to the pathogen in the environment, and as such the patient could not transmit the pathogen through environmental contamination or direct contact making it an unlikely pathway to a widespread outbreak.

2.2.3 Stage 3: Assessing Study Quality

After the removal of duplicates, search results were screened to assess the study quality and decide on selection against an eligibility criterion (Table 2.1), developed through the PICOS method [20]. After consideration of published tools, the National Institute of Health (NIH) quality assessment tool (Supplementary Table 2.3) was used for study appraisal and thresholds set for exclusion [21]. The NIH tool was chosen as it best captures the types of studies reviewed here, while accounting for bias and several methodological flaws. Studies score either 'good', 'fair' or 'poor', depending on the how many questions are answered 'yes'. All studies rated 'poor' were removed and any study which rated 'fair' was assessed to decide if the questions answered 'no' would lead to major biases in answering the research questions. The papers were screened by one reviewer (G. Charnley) and ineligible papers eliminated. All titles and

abstracts that met the criteria were subjected to full-text reading.

Table 2.1: Eligibility criteria for the included literature in the systematic review. Developed by the PICOS method: Population, Intervention, Comparison, Outcome and Study type

Inclusion Criteria	
Population	Any local population/community impacted by a post-disaster disease outbreak
Intervention	Any investigation carried out to quantify a disease outbreak and understand the risk factors
Comparator	Members of the disaster-affected population who did not acquire an infection during the outbreak
Outcomes	The primary outcome is to understand post-disaster disease outbreaks on a global scale. The secondary outcome consists of identifying the risk factors that result in these outbreaks
Study type	Retrospective observational studies, namely, cross-sectional, case-control and cohort studies. Full-text or abstracts in English
Exclusion Criteria	
Papers without an explicit link between a disaster and an outbreak	
Outbreaks in refugees/refugee camps, foreign armed forces, aid workers and international travellers, as this review aimed to look at local outbreaks in regional populations	
Non-English abstract and full-texts, due to linguistic constraints	
Review papers, as only primary sources were desired for this review	
Single case reports, as these were often not seen as representative of an outbreak in this context	

Publications discussing general risk factors, public health, mental health and non-communicable diseases, pathogen genetics or economic costs in a post-disaster setting were excluded, as they are beyond the scope of this review and its objectives. Disease outbreaks in international refugee camps were also removed, due to most refugees being housed in the camps from multiple countries, involving multiple disasters. Therefore, linking these outbreaks to a specific disaster was challenging and the only outbreaks in a camp setting included were national relief camps [16].

2.2.4 Stage 4: Summarising the Evidence

A predetermined data charting form was used based on preliminary reading and the objectives of the review. Extracted data included information on the publication (title, authors, date, journal), disaster type, disease, case numbers, study area, study period, identified risk factors, methodological details (study design, sample sizes, laboratory tests, statistical analysis) along with any other relevant information/data. Risk factors were recorded regardless of whether the author ran statistically analyses. To ensure all relevant data were collected, the form was reviewed by other members of the research team before implementation and the data were extracted independently by the sole reviewer (G. Charnley) [16].

To ensure that distinctions could be made between risk factors and there was no overlap in grouping, risk factor recording was a dynamic process. The exact wording of the reported risk factor was first entered into the data charting form and then reviewed and streamlined into categories after all the studies had been read, re-evaluating studies as needed. As this process is open for interpretation, Supplementary Table 2.4 shows all the individual risk factors and how they were clustered to improve transparency.

2.2.5 Stage 5: Interpreting the Findings

Descriptive Analysis

Categorisation was used only for ease in interpreting and presenting the results and were altered retrospectively as needed (e.g., categories removed if no outbreak was found). To improve transparency, how each outbreak was categorised is presented in tables throughout Chapter 2 and the Supplementary Material. Regions were categories based upon how the results were clustered (Africa, South & South East Asia, East Asia, Europe, Latin America and the Caribbean (LAC), North America, the Middle East, Oceania and Europe). It is acknowledged that the chosen regions were somewhat non-comparable due to differences in population sizes, environments and proximity to causative factors for hazards (e.g., fault lines for geophysical hazards).

Slow-onset and sudden-onset disasters were considered in this review and were categorised into the five groups outlined below. The classification followed those provided by EM-DAT, the Emergency events Database curated by the Centre for Research on the Epidemiology of Disasters [22]. Any flooding caused by a tsunami or storm was listed under the hazard causing the flooding, not hydrological, as this was considered the primary cause of the outbreak, while being aware that the vulnerabilities are still the cause of the disaster:

- Conflict - any form of reported armed conflict or violence
- Hydrological - flooding caused by high precipitation (not by hurricanes, cyclones, typhoons, tropical storms or tsunamis)
- Geophysical - earthquakes, volcanic eruption and tsunamis
- Meteorological - hurricanes, cyclones, typhoons and tropical storms
- Climatological - droughts

Diseases were also categorised into disease type (bacterial, viral, parasitic and mixed pathogen) and transmission type (water-borne, vector-borne, air-borne, direct contact and rodent-borne). Risk factors were identified by any study that specifically named them as risks or was suggested to have been involved in facilitating the outbreak (either statistically or not). Risk factors were divided manually into mutually exclusive clusters identified by similarities in how they resulted in an outbreak and preliminary reading. This formed fourteen clusters, which are delineated below, along with how they were defined:

- Displacement - A report of national population movement due to the disaster
- WASH - Any issues with access to or quality of water, sanitation and hygiene provisions, separate from disruptions to specific municipal services
- Housing - Reports of inadequate living conditions or the location of either, habitual residence or temporary housing provided after displacement

- Vector/animal - Changes in animals and vectors that increased contact with the population and subsequent disease spread
- Age - Reported age-related risk factors, either a demographic group e.g., children, or a specific age category e.g., <5 years
- Healthcare - Any issue that prevented people seeking formal healthcare
- Gender - Males or females being more at risk
- Behaviour - Any report of human behaviour which heightened the risk of contact with the pathogen, except human displacement
- Environment - Alterations in the natural environment that exacerbated the risk of contracting the disease
- Municipal services - Disruption to municipal services e.g., street cleaning, removal of waste
- Nutrition - Issues with insufficient diet or eating specific foods
- Occupation - An occupation which was associated with increased pathogen exposure and therefore disease
- Socio-economic - Further socio-economic conditions that increased cases but did not fit into any other cluster. These mainly included education and poverty
- Co-morbidities - Significant numbers of infected individuals also presented with another morbidity

Following the data extraction and to help illustrate how the information collected answered the aims and objectives, the results were presented both; (1) numerically, with outbreaks broken down and quantified by disaster, geographic region and pathogen, along with the importance of risk factors and (2) narratively, by synthesising the methods used, the importance of global change and the links between risk factors and possible cascades.

Statistical Analysis

A range of statistical analyses were used including Pearson's chi-squared (X^2), at a significance level of p <0.05 (equation below). The test identifies if the differences in categories e.g., regions, disasters, diseases and risk factors, were significant or due to chance. The X^2 value is the sum of the square of the residuals, where O_i is the observed values and E_i is the expected, if there was no significant difference between groups:

$$X^2 = \sum \frac{(O_i - E_i)^2}{E_i}.$$

The Pearson's chi-squared residuals indicate how far the observed value was from the expected value. The standardised residuals (to 3 decimal places) were plotted in a correlation matrix, with a residual closer to zero indicating less of an association between the corresponding row and column (and similar observed and expected values). A stronger positive association signifies the observed was higher than expected and a negative association indicates it was lower than expected (if there was no significant difference between groups). To extract the standardised residuals, the following equation is used:

$$Standardised\ residuals = \frac{(O_i - E_i)}{\sqrt{E_i}}.$$

Confidence intervals for multinomial proportions are a measure of uncertainty and work by taking a proportion of the sample and adjusting for sampling error, in this instance at a 95% level of confidence. Confidence intervals were used here to help illustrate the potential impact and limitation of sample size in some categories. Sample size was likely to vary widely here and a smaller sample size reduces the certainty that the sample reflects the population.

A hierarchical cluster analysis was utilised to show the similarity between risk factor cluster reporting. The aim of this analysis was to lend evidence to the hypothesis that some risk factors may be linked. A distance matrix was first computed using complete linkage clustering algorithm. Next, hierarchical clustering was used to analyse the set of dissimilarities, at each

stage joining the two most similar clusters by computing the maximum distance, continuing until there was a single cluster and plotting the results on a dendrogram [16]. All statistical analysis was completed in R using R version 3.6.273 (packages: "corrplot" [23], "DescTools" [24], functions: "corr()", "corrplot()", "MultinomCI()", "dist()" and "hclust()").

2.3 Results

2.3.1 Search Results

After screening the search results, 132 studies were selected for inclusion in the analysis (Supplementary Information 2.1) and a PRISMA flow diagram illustrates the selection process below (Figure 2.1). Electronic database searching ceased in June 2020 but no studies after 2019 met the inclusion criteria; the studies therefore spanned from 1940 to 2019 and included ten different types of disaster and 39 different diseases across six continents. The types of studies included were retrospective and mainly involved observational studies, namely cross-sectional, case-control, case-crossover, cohort studies and epidemiological and environmental field investigations.

Several studies were either multi-disaster or multi-disease events. These were split to allow full quantification of diseases and disasters and resulted in 140 separate disease outbreaks and 137 separate disasters. Eight studies had only an abstract in English available, therefore the full text was not reviewed. Given the caution that the searches were conducted in English only, expansion to other languages at this stage would have yielded inconsistent results. A further 25 studies were excluded because they focused on internationally displaced populations in refugee camps. Four studies with methods involving serological surveys for disease prevalence were removed, as confirmed cases of a current infection/outbreak related to a disaster could not be identified by serology. Relatively few studies were excluded due to being categorised as 'poor' by the NIH tool, and mainly included those published before 1950.

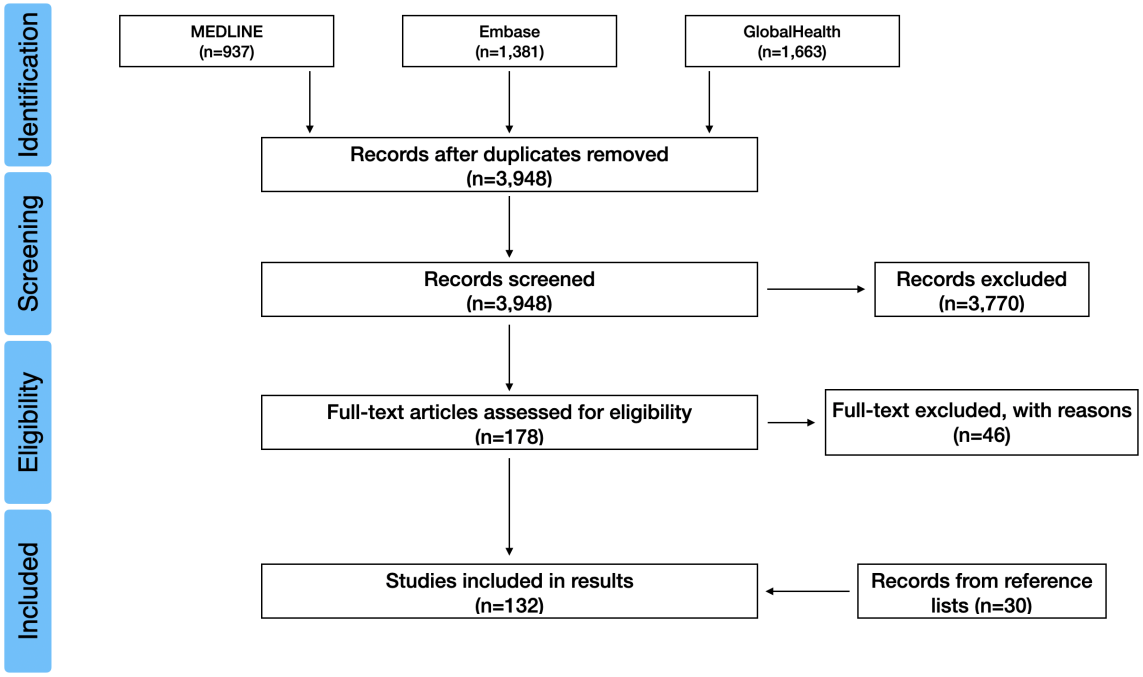


Figure 2.1: PRISMA diagram for the selected 132 studies on post-disaster disease outbreaks.

2.3.2 Disaster, Region & Disease

Conflicts, hydrological and geophysical events were the most commonly reported disasters associated with disease outbreaks, with fewer outbreaks associated with climatological and meteorological events. A full list of reported disaster frequencies is shown below in Table 2.2. It is worth noting that although conflicts appear frequent (n = 45), they were not sub-categorised (mainly due to the large proportion of civil wars) and are less frequently reported than all natural hazards grouped together (n = 92).

Table 2.2: Full list of reported disasters included in the review and their frequencies.

Disaster Type	Disaster	Frequency
Hydrological	Flood	43
Geophysical	Tsunami	9
	Earthquake	17
	Volcanic eruption	2
Meteorological	Cyclone	4
	Typhoon	4
	Tropical Storm	1
	Hurricane	6
Climatological	Drought	6
Conflict	Armed conflict	45

Africa, S & SE Asia and the Middle East were strongly over-represented compared to Oceania, the Americas and Europe in post-disaster disease outbreaks (Figure 2.2 and 2.3a). Within the regions, India (n = 12), the USA (n = 10) and China (n = 9) were predominant. A full list of reported region frequencies is shown in Supplementary Table 2.5.

The over-represented regions are mainly accounted for by the large number of conflict-related disease outbreaks in Africa and the Middle East (Figure 2.3a & 4a), especially in Sudan and South Sudan (9/45). In contrast, there were relatively few reports of African geophysical events and S & SE Asian conflicts. Africa also experienced a high proportion of climatological-related events, reporting 5 out of 6 drought-related outbreaks (Figure 2.4a). S & SE Asia mainly reported hydrological and geophysical-related outbreaks (Figure 2.3a), commonly in India (12/71), along with Bangladesh and Sri Lanka (7/71).

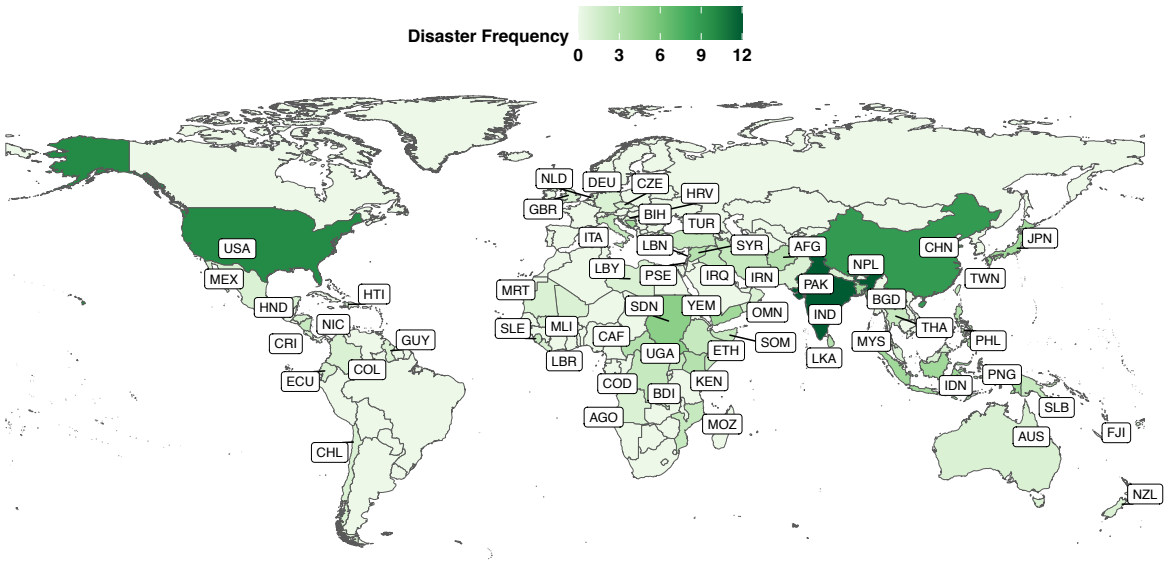


Figure 2.2: Frequency of reported post-disaster disease outbreaks by country for the 137 separate disaster events found in the literature search. Labels are ISO3 country code and only countries which reported outbreaks are labelled.

With respect to causal agent and transmission mode of disaster-related disease outbreaks, bacterial and water-borne diseases were predominant groups (Figure 2.3b and c), compared to mixed pathogen, direct contact and rodent-borne pathogens. A full list of reported aetiologies and transmission modes are shown in Supplementary Table 2.6.

Reported outbreaks were often disaster specific, and therefore diseases associated with hydrological events and conflicts were frequently reported. There were strong positive correlations between bacterial or water-borne diseases and hydrological events and viral or parasitic disease and conflicts (Figure 2.4b and c). This was mainly due to the number of post-flood leptospirosis (n = 18), cholera and dysentery outbreaks (n = 8). In addition, geophysical events and air-borne pathogens showed positive associations, whereas strong negative correlations were seen between conflicts and bacterial pathogens and vector-borne disease and hydrological events (Figure 2.4b and c). Additional Pearson's chi squared analysis for pair-wise comparisons is shown in Supplementary Table 2.7.

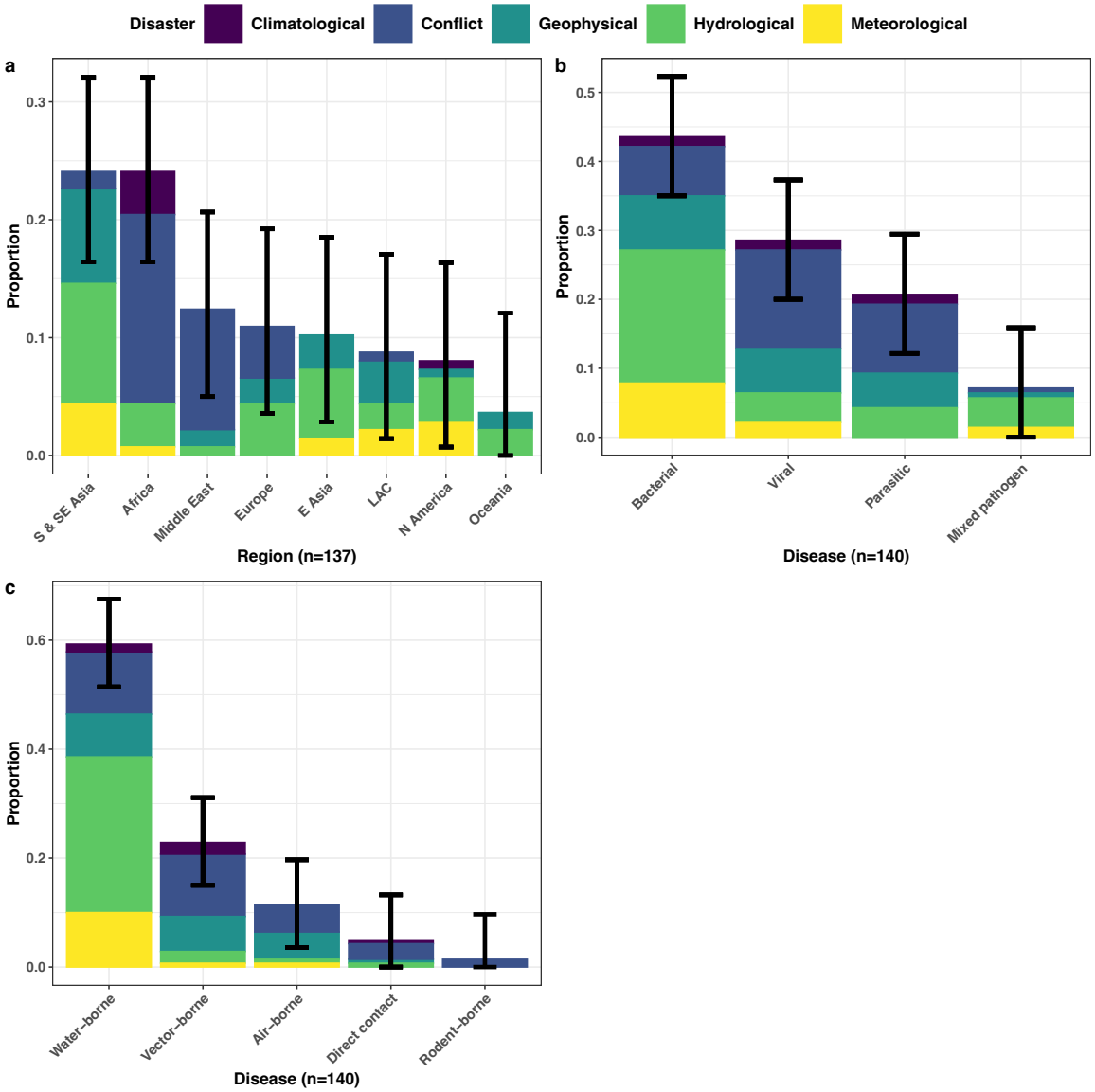


Figure 2.3: Proportion of reported post-disaster outbreaks by **a**, region against the 137 separate disasters, **b**, the 140 separate disease outbreaks by pathogen type against disaster and **c**, the 140 separate disease outbreaks by transmission against disaster with multinomial confidence intervals (95%). LAC – Latin America and the Caribbean.

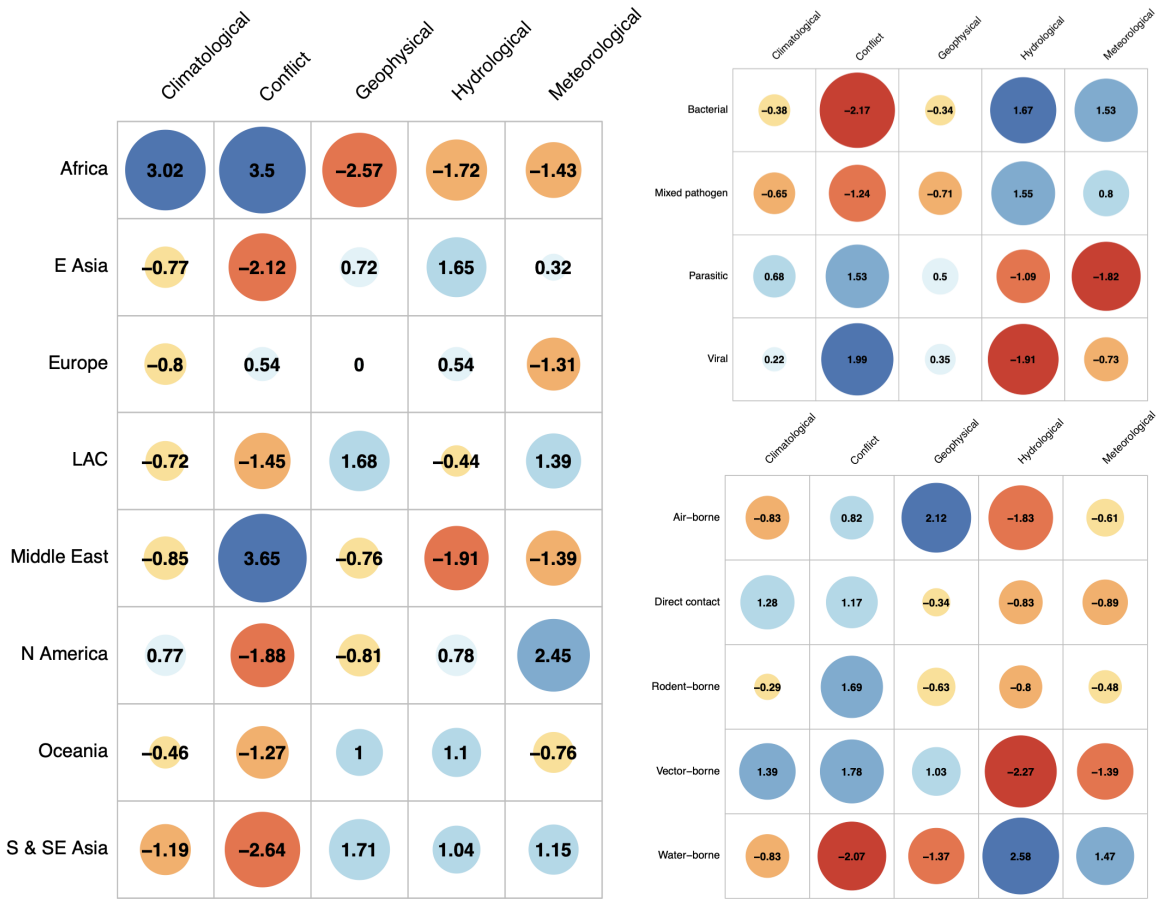


Figure 2.4: Correlation matrix for the Pearson's standardised chi-squared residuals of the categories in **a**, region against disaster ($chi^2 = 101.81$, p-value = <0.05), **b**, disease against disaster ($chi^2 = 31.49$, p-value = <0.05) and **c**, disease transmission against disaster ($chi^2 = 47.31$, p-value = <0.05). Positive residuals are in blue and signify a positive association and higher observed value than expected between the corresponding row and column. Negative residuals are in red and signify a negative association and lower observed value than expected. Created using R package "corrplot" [23].

2.3.3 Risk Factors

Across the 132 post-disaster disease outbreaks, 418 risk factors were reported in the studies reviewed. Individual risk factors had varying frequencies within the fourteen main clusters (Figure 2.5) and how they were grouped are shown in Supplementary Table 2.4. Pearson's chi-squared analysis found that risk factors were significantly different (at p <0.05) among post-disaster disease outbreaks. Additional figures for the fourteen main risk factor clusters against

disease are shown in Supplementary Figure 2.1 and additional statistics in Supplementary Table 2.7.

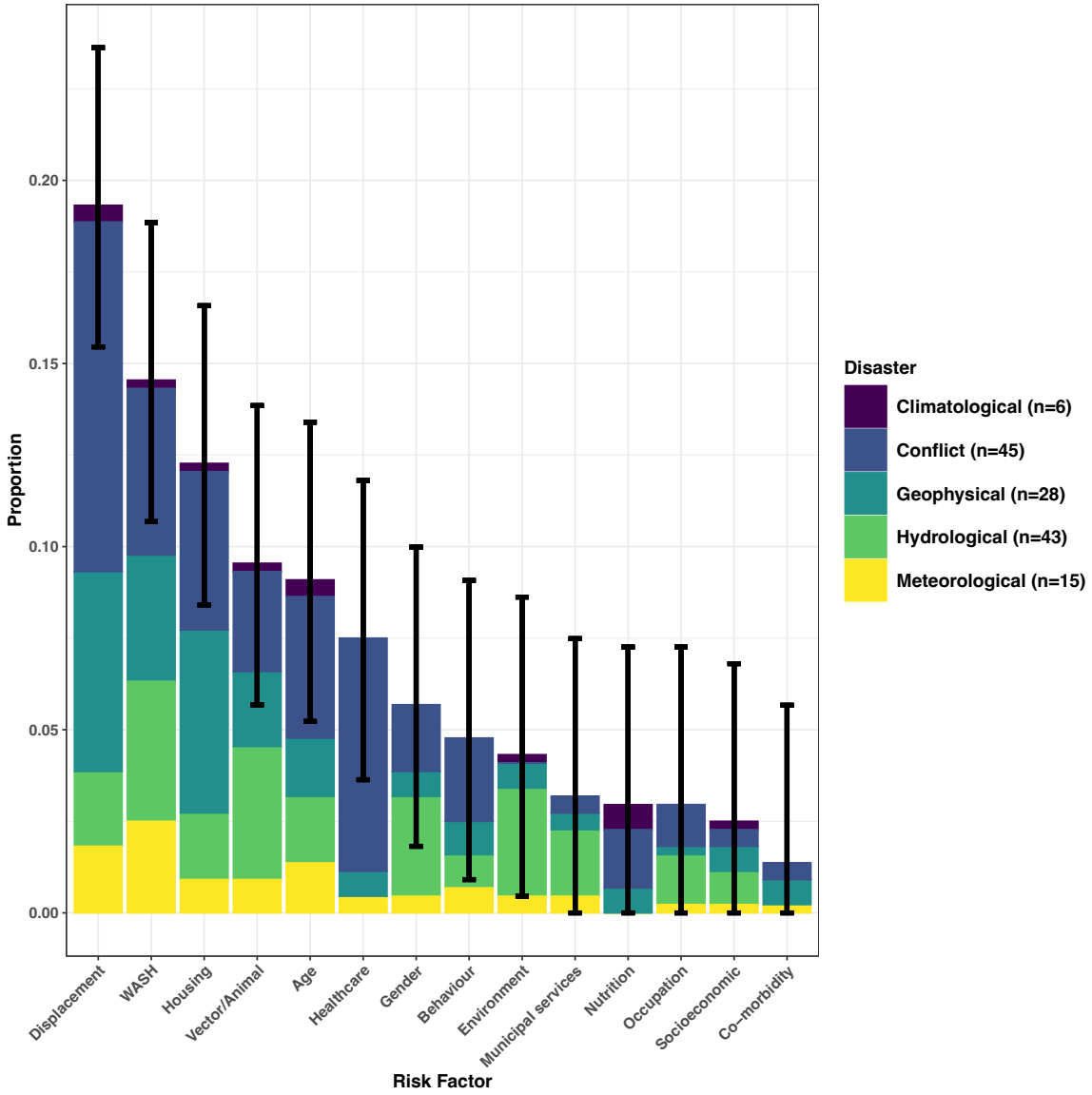


Figure 2.5: Proportions of the fourteen main risk factor clusters out of the 418 risk factors reported in the search results, against disaster, with multinomial confidence intervals (95%). WASH - Water, sanitation & hygiene.

The most frequently reported risk factor was displacement, being reported 81 times, especially in relation to conflict ($chi^2 = 4.29$, $p < 0.05$) and geophysical events ($chi^2 = 1.51$, $p = 0.22$). It was most frequently reported as a general risk factor, with no details given or to national relief camps and temporary housing. In two studies, displacement was expanded upon with details

on the initial and final destination e.g., rural to urban.

WASH was the second most commonly reported cluster (n = 59), due to poor sanitation, access to clean drinking water and poor hygiene (Figure 2.6a). WASH risk factors were high in all disaster types, other than climatological, potentially due to its small sample size (n = 6). The highest frequencies of WASH risk factors were seen in hydrological events (chi^2 = 0.3, p = 0.58) and conflicts (chi^2 = 2.16, p = 0.14), although chi-squared analysis showed they were not significant. Instead, WASH risk factors were particularly prominent among water-borne disease outbreaks (chi^2 = 13.64, p <0.05), such as leptospirosis, cholera and dysentery, and mainly attributed to the increase in standing floodwater and damage/overflow of sanitation systems.

Poor housing was the third most commonly reported cluster (n = 48), often associated with geophysical events (chi^2 = 10.66, p <0.05), such as earthquakes and tsunamis. The resultant extensive infrastructure damage following these events lead to displacement in conjunction with housing risk factors, presenting through the high incidence of overcrowding (19/48), poor or temporary shelter and camp settings (13/48) (Figure 2.6b).

Changes in vector (mosquito) or animal (domestic, livestock, wildlife) exposure were frequently linked with hydrological (chi^2 = 5.17, p <0.05) and conflict (chi^2 = 2.34, p 0.13) events, through alterations in vector breeding ground (14/39) and vector control (4/39) (Figure 2.6c), leading to parasitic diseases (chi^2 = 8.46, p <0.05), such as malaria.

Of the 40 reported age-related risk factors, a quarter were in children under five years, with people under 20 years increasing that proportion to 75% (Figure 2.6d). This was region and disease-specific, with several water-borne (chi^2 = 2.13, p = 0.14) diarrhoea outbreaks in conflict events (chi^2 = 0.16, p = 0.68) reporting children under 5 years as a risk factor.

Poor healthcare services resulting in disease outbreaks (n = 35) were particularly common in conflicts (chi^2 = 30.6, p <0.05) compared to natural hazards. Poor access and vaccination coverage were the most common risk factors in this cluster (Figure 2.6e), therefore high levels of vaccine-preventable viral diseases were reported (chi^2 = 14.4, p <0.05).

Gender was reported 25 times, 20 of these stated that being male was a risk factor (vs 5 reports of being female as a risk factor). A common narrative was that men assisted in post-hydrological event clean-up activities ($chi^2 = 8.5$, $p < 0.05$), increasing their exposure to floodwater, the most common risk factor reported in the environment cluster (13/19). The exposure to floodwater resulted in an enhanced likelihood of contracting water-borne diseases ($chi^2 = 2.08$, $p = 0.15$), especially leptospirosis (7/18).

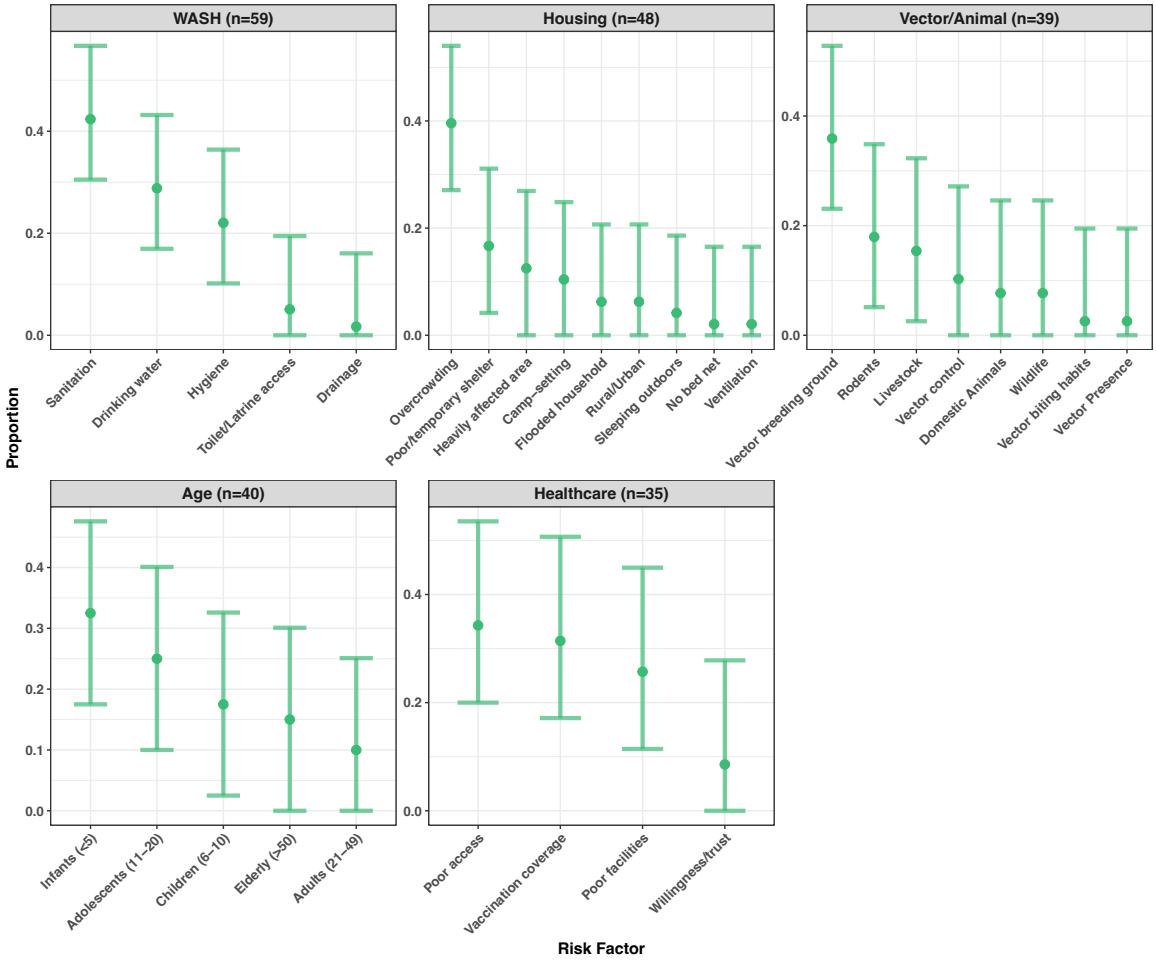


Figure 2.6: The five most commonly reported risk factor clusters (Water, sanitation and hygiene (WASH), Housing, Vectors/Animals, Age and Healthcare), split into the proportion of individual reported risk factors, with multinomial confidence intervals (95%). Although displacement was the most frequently reported risk factor, it was not included as it had few elements within the cluster.

2.3.4 Multi-Risk Factor Reporting and Clustering

Most of the reviewed disease outbreaks were associated with multiple risk factor clusters; almost half of the included studies cited two (29/132) or three (32/132) risk factors (Figure 2.7a). This is also underestimated, as multiple risk factors were often reported within each cluster, for each outbreak (Supplementary Table 2.4). Of the comparatively few studies that reported zero (n = 8) or one (n = 22) risk factors, several (3/7 and 4/20, respectively) were in studies where only an abstract was available and therefore risk factors may have been discussed in the full text. In the studies that reported at least 1 risk factor cluster (n = 124), conflicts were most common (n = 46) and India and China were the most common countries reporting multiple risk factors with 8 and 7 multi-risk factor outbreaks, respectively. Unspecified or multi-pathogen diarrhoeal disease and cholera were the most frequent multi-risk factor diseases, but the commonality of these groups (conflicts and water-borne diseases) may represent the comparatively large number of reported outbreaks.

The hierarchical clustering analysis (Figure 2.7b and Supplementary Figure 2.2) helps to illustrate and understand the relationships between risk factors and how they were reported together. It is clear that displacement, WASH and housing were the most related risk factors here. Thirteen studies reported WASH and housing risk factor clusters together, mainly through overcrowding (n = 10), hygiene (n = 7) and sanitation (n = 6). WASH risk factors were also commonly reported with displacement, being reported together 12/13 times. Healthcare and age were reported together eleven times, eight of which were in children <15 years old (and commonly male) and mainly reported issues with vaccination coverage or poor access. Of the twelve occurrences that age and gender were reported together, seven were in males <20 years old. The similarity of gender, the environment and behaviour were predominantly through male exposure to floodwater and assisting in post-disaster clean-up, as previously discussed, with gender and hydrological events also showing statistical significance, as shown on the previous page.

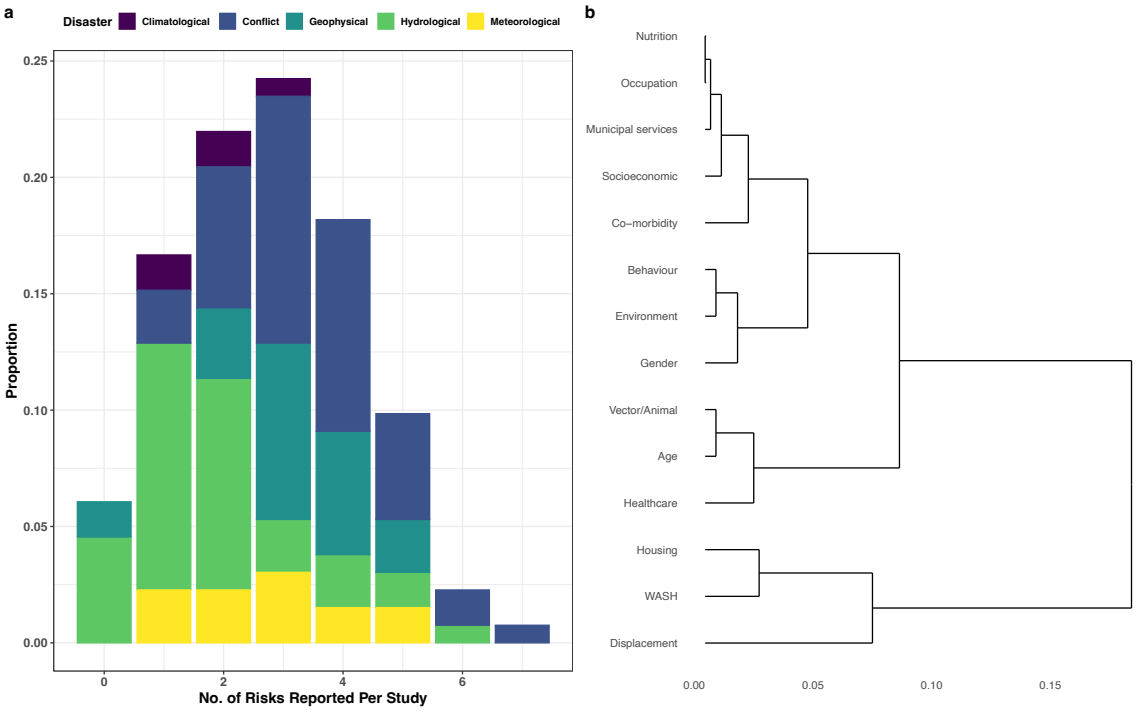


Figure 2.7: Multi-risk reporting and hierarchical clustering. **a**, Proportions of studies (n = 132) which reported either 0 to 7 different risk factors, within the fourteen main clusters. **b**, cluster dendrogram from hierarchical cluster analysis for the fourteen main risk factor clusters. Individual segments (leaves) on the lower part of the tree are more related to each other, as indicated by distances between the branches. The scale bar showing the dissimilarity distance between the proportions of each risk cluster. 8b was created using base R function "hclust()".

2.4 Discussion

The results shed new light on post-disaster disease outbreaks globally, including their frequency, geography and characteristics. The most striking results identified here include the large numbers of bacterial and water-borne disease due to hydrological events in South Asia and viral diseases in African conflicts. Diseases and their associated risk factors were often disaster-specific, as certain disasters created ideal conditions for specific pathogens. The hierarchical clustering showed further evidence for the multifaceted nature of these outbreaks and the idea of risk factor cascades contributing to these outbreaks.

Displacement was involved with many other risk factors, resulting in poor health outcomes and also involved in spreading diseases to new areas, as seen with Lassa fever in Sierra Leone [25]. Loss of infrastructure and the resultant displacement appears to be important in both

armed conflicts and natural hazards, leading to damage to habitual residence, healthcare and services. Examples include destruction of healthcare and housing after an earthquake in Japan, leading to a pneumonia outbreak [26] and difficulties in accessing health care in Yemen during the ongoing civil war and cholera outbreak [27]. Despite these conditions being potentially important in both natural hazards and armed conflicts, how they yield negative health impacts may be different and only conflict and displacement proved to be statistically significant from the chi-squared analysis.

2.4.1 Risk Factor Cascades

Natural Hazards

Natural hazards may result in risk factor cascades driven by displacement (Figure 2.8a), due to infrastructure damage. Damage can occur through flooding involving meteorological or hydrological events. Alternatively, it occurs through direct damage in geophysical events, with geophysical events and displacement showing a statistically significant relationship. Infrastructure damage and floodwater generally led to an increase in poor living conditions and an inability to maintain hygiene standards and access clean water, explained through the clustering of displacement, WASH and housing in the hierarchical clustering. Evidence for this is presented in over half of the reported WASH risk factors, occurring in post-hydrological or post-meteorological events and the statistically significant relationship between poor WASH conditions and water-borne diseases.

Flooding leads to increased exposure to groundwater and overflowing sewage systems. These conditions can expand vector breeding grounds, increasing the contact between populations and vectors and the resultant increase in disease cases [28, 29]. In contrast to this, Figure 2.4c showed a strong negative association between hydrological events and vector-borne disease. Vector breeding can be more complex than space to breed (e.g., standing water), and other factors (e.g., temperature, salinity) may prevent vectors breeding in floodwater. Floodwater has also been known to destroy breeding grounds, instead of creating new ones [30].

Conflicts

In armed conflict events, cascades may result from loss of healthcare infrastructure, limiting access to and the quality of health services (Figure 2.8b), especially for children. Statistical analysis adds further evidence to this statement, finding a significant relationship between healthcare risk factors and conflict and similarity between healthcare and age in the cluster analysis. Vaccination coverage was a commonly reported risk factor in these events, potentially accounting for the significant relationship between healthcare risk factors and viral disease. Fourteen out of twenty conflict-related viral outbreaks were vaccine-preventable diseases, including hepatitis A, polio and measles. Mass vaccination campaigns are commonly run through humanitarian aid organisations and as conflicts escalate, these services are often suspended due to safety concerns [7]. Another study suggested that despite high measles vaccination coverage in the Central African Republic, an outbreak still occurred due to reporting issues and poor cold-chain maintenance [31].

An additional factor seen in conflicts includes healthcare forming the political fabric of the violence, resulting in attacks on health centres and workers. This further reduces uptake of services as people do not perceive seeking care as safe [32] and mistrust can escalate towards both the government and healthcare providers [33]. For example, conflict in the Democratic Republic of Congo has reportedly hampered Ebola response teams in the outbreak with begun in 2018, causing delays in vaccination and reducing vaccine effectiveness by as much as 37.7% (based on the ratio of administered doses to the number of primary and secondary contacts of each case) [34].

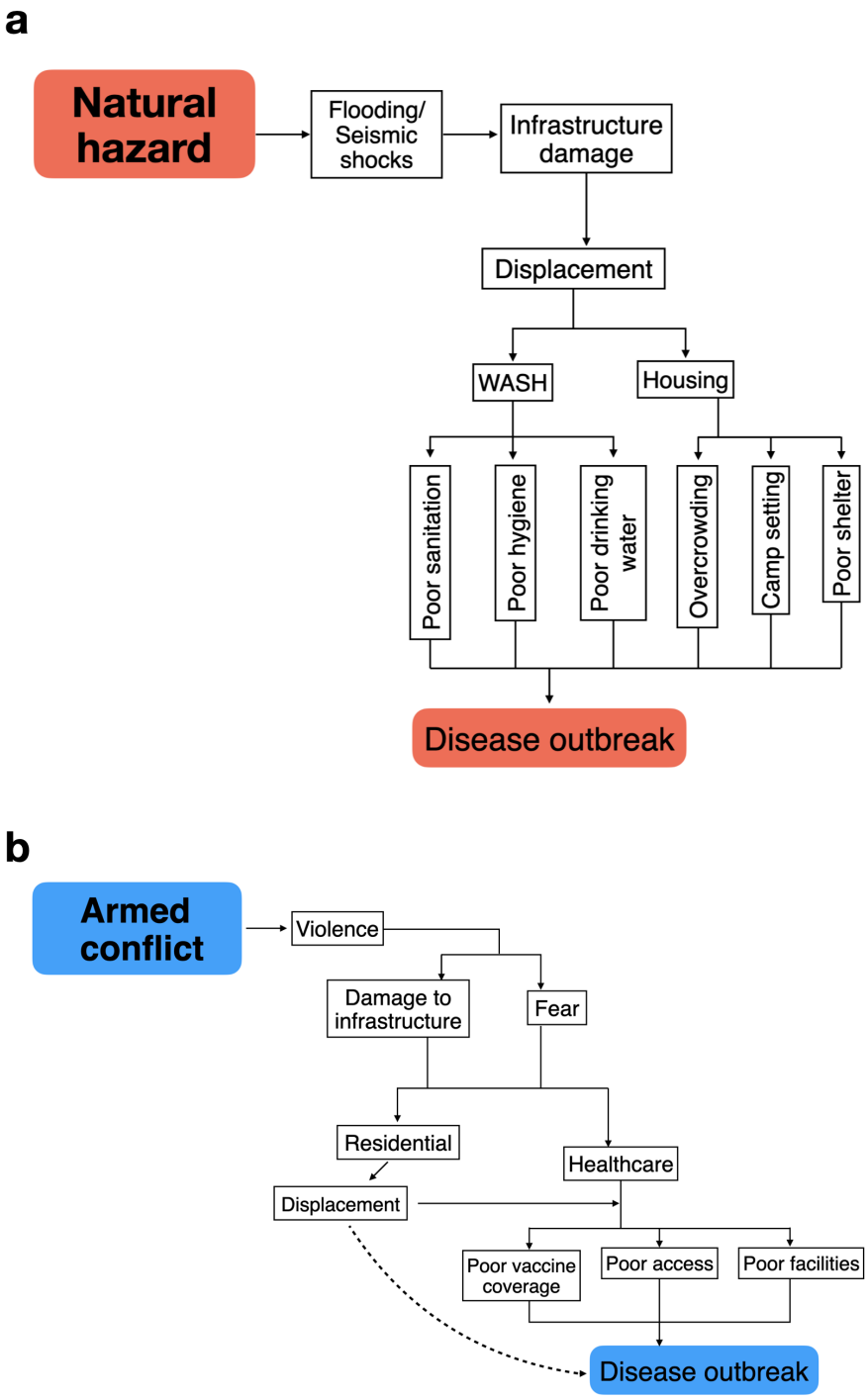


Figure 2.8: Examples of cascading risk factors for **a**, natural hazards and **b**, armed conflicts. The dashed line between displacement and disease outbreaks in **8b** represents the acknowledgement that displacement does not directly lead to disease outbreaks, but instead the conditions it creates when poorly managed.

2.4.2 Displacement

Despite a common negative narrative used for displacement, one study discussed displacement as a protective factor. Reporting on West Nile virus after Hurricane Katrina, the study states that displacement allowed people to move away from floodwater and therefore vector breeding grounds [28]. A lack of displacement may be a sign of inequity and poor socio-economic conditions, as people do not have the financial means to move and therefore become trapped within the affected area [35]. Displacement can be an opportunity to move people out of immediate danger caused by the disaster and provide services quickly and easily to large groups.

Unfortunately, the opportunity to remove this danger is often not capitalised on. For example, of the 25 times being male was reported as a risk factor, 17 of these were during outbreaks where displacement was not reported to have occurred. This is not simply a representation of the commonality of non-displacement, as displacement was reported on more occasions (n = 73) than not (n = 54). Interestingly, exposure to floodwater was also reported 16 times in outbreaks without displacement, compared to just once in studies where displacement did occur. This suggests that without displacement (especially after flooding), risk factor cascades resulted from men being more likely to assist in post-natural hazard clean up (potentially due to gender norms, expectations and stereotypes [36]), exposing them to disease.

2.4.3 Water, Sanitation & Hygiene

These results and clustering help highlight the importance of basic sanitation and hygiene, regardless of disasters, as poor WASH is linked to several infectious diseases and often associated with poverty [37, 38]. For example, in Kenya, only 24.3% of the population have access to adequate sanitation, a figure which is much worse for rural communities [39]. Unfortunately, studies rarely mention these non-disaster-related conditions which impact population vulnerability. Instead, risk factors are solely reported in causing the outbreak but not why they occurred or previous conditions, providing a current state and not a comprehensive view of vulnerability. This is a potentially important area of future research, especially for effective

disaster planning.

Poor WASH conditions may explain why children were often implicated in these results. In India, where 28% of the reported outbreaks occurred, around 1.7 million children died before the age of 5 in 2010 alone, with diarrhoea causing 13% of this mortality [40]. Similar statistics are present throughout South Asia and Africa [41]. Another possible reason is that commonly reported diseases, including polio, measles and cholera, heavily impact young children, due to physiological (rapid onset dehydration and wasting) and social differences (poor hygiene standards) [42, 43]. The gendered and age-specific risk factors found in this review, stress the need for sex and age-disaggregated post-disaster data in order to try and fully understand the impacts on disease outbreaks.

2.4.4 Implications of Global Change

The risks of climate change for health are far-reaching [44, 45] and natural hazards provide an opportunity to attribute a climate-related event to a health outcome. The review findings have several implications for region-specific global change. For example, under climate change scenario RCP4.5 (an intermediate scenario representing moderate emissions reductions), projections are geographically heterogenous suggesting a drier Africa and the Middle East and wetter southern Asia [46]. These changes may therefore alter the frequency/intensity of droughts and floods in the future and therefore, more related disease outbreaks in areas already heavily impacted by post-disaster disease outbreaks.

Two studies [47, 48] also reported that contact with floodwater in conjunction with higher than normal temperatures was a risk factor for developing a water-borne disease. Alterations in temperatures can impact the ways pathogens and vectors behave in the environment, yielding implications from rising global temperatures [30, 49, 50].

Alterations in temperature and precipitation may also occur in conjunction with population growth and urban expansion, with East and South East Asia seeing the highest rates of urbanisation [51]. Many of these areas are low-lying coastal cities, and liable to flooding, sea level

rise [52] and potential post-disaster disease outbreaks. This combination of both climate and population changes may therefore put more people (at higher densities) at risk for post-disaster disease outbreaks.

However, urbanisation provides opportunities to meet the needs of concentrated groups of people and can be an effective low-carbon way of managing and providing services and employment. Successful urban planning through building design, education and provision of healthcare, contributes to effective disaster mitigation strategies and possibly reduces the risk of post-disaster disease outbreaks [53, 54]. While urban residents often have lower emissions per capita, compared to their rural counterparts, reducing their impacts on climate change [55].

2.4.5 Limitations

Difficulties arise when comparing one disaster to another, as disaster severity, population risks and socio-economic conditions of affected populations are substantially different. Thus, this is not a complete list of global post-disaster disease outbreaks and outbreaks are likely to have been missed through excluding grey literature and internationally displaced populations. If populations were displaced internationally by a disaster and an outbreak occurred, it could be argued that this was caused by the disaster. Despite this, several of the reported camps housed refugees from multiple countries, linking them to multiple disasters; therefore, this would have created issues linking the outbreak to a specific disaster.

There is a temporal bias in reporting here, as disaster and disease outbreak reports and the curation of public datasets have increased exponentially in recent years, due to a greater global effort [56, 57]. The increase in reporting has also coincided with an increase in published literature. It is therefore difficult to understand whether disease outbreaks or disasters are increasing or if this is a product of reporting. Additionally, studies which were published before 1950 often did not meet the inclusion criteria here, scoring 'poor' on the NIH tool, mainly due to a lack of methodological detail. This removed a number of conflict-related outbreaks reported during World War II.

Reporting bias may be a cause of the gender-related risks found here. The high number of males reporting disease may not be due to more men contributing to disease cases but because they were more likely to seek formal medical assistance and therefore be reported. More research is needed to understand gender biases and barriers for women accessing care in post-disaster settings. For example, women may have less access to health insurance and financing or may not be allowed to attend hospitals alone due to cultural values or fears around safety [58, 59].

Outbreaks were particularly common in disasters that were highly publicised, contributing to another potential reporting bias. For example, of the 26 earthquake and tsunami-related outbreaks, ten were due to just two natural hazards; the 2011 Japan Earthquake and Tsunami and the 2004 Indian Ocean Tsunami. This may have introduced an over-reporting bias for certain disaster types and regions, raising questions about whether these disasters saw more disease outbreaks, or whether they were more often reported. However, they may have been highly publicised because they were particularly severe in terms of damage and mortality and therefore resultant disease outbreaks. Comparing the results against a "baseline" may be a method to address reporting biases, however disaster databases also suffer from similar reporting issues, resulting in no ideal data or figures for comparison [56].

The over-represented regions found in this review are generally stated as having high numbers of disaster, compared to other regions. Additionally, certain regions may be over-represented due to the removal of non-English studies, e.g., few outbreaks were found in South America, potentially because the search strategy excluded studies written in Spanish and Portuguese. Therefore, the large number of outbreaks may have been a product of the overall higher disaster frequencies in certain areas. For example, in a 2020 review Africa and the Middle East were reported as being the most conflict-prone regions [60], while in a 2018 review of natural hazards, 141/315 hazards were reported in Asia [61]. However, the results still increase the understanding of the risk factors caused by certain disasters. It could be argued that the higher number of disasters in these regions should lead to less outbreaks, as the region is more aware of the risks and how to prevent them. Additionally, figures for the number of disasters have several limitations and are dependent on what was included in the data.

Defining a disaster and attributing an infectious disease outbreak to the event has its difficulties, as there is no consensus on when a disaster ends and recovery begins [62]. This creates limitations in assigning and comparing risk factors and a major limitation of risk factor analysis is its subjectivity. If the authors of the reviewed studies did not clearly state their risk factors and mechanisms, this resulted in an element of subjectivity in trying to interpret the results. Reported risk factors also depend on the data collecting process used during the outbreak and what was asked of participants. Several of the less frequently reported risk factors may link to more common factors but were just listed differently by the authors and resulted in high uncertainty.

The magnitude of a risk factor in one event may differ from others, particularly in terms of how it was perceived by the population and researchers. Confirmed and probable cases ranged from two to 379,000 across the 132 studies; therefore, how risk factors were measured and analysed in such a wide range of case numbers is likely to differ, especially statistically. Despite the studies' limitations, with 132 separate outbreaks and 418 reported risk factors, this review is significantly larger and broader in scope than other studies exploring similar subjects [1, 2, 3].

2.4.6 Conclusion

Our understanding of how global change will alter risks to populations is still relatively incomplete and has become a growing area of study, including population vulnerability to disasters. Chapter 2 is the first comprehensive global overview of disaster-related disease outbreaks and highlights commonly reported risk factors related to both conflicts and natural hazards. Despite displacement being suggested as an important risk factor, displacement may help mitigate several other risks and remove people from hazardous situations, ultimately protecting their health relative to those not displaced. This is an important finding for disaster and public health literature, as this challenges the narrative of many previous studies. It supports the theory and practice of disaster risk reduction and response in terms of recognising that displacement is not inherently detrimental, but the impacts depend somewhat on how the displaced people are supported.

India and several African countries had particularly high outbreak reporting rates compared to other countries. Further evidence is needed to understand why this is the case, or if it is simply a by-product of their very large geographic areas and population sizes. Certain disease aetiologies were common in specific disasters, which were often reported in specific regions. This specificity is essential for international disaster risk reduction, as humanitarian and government sector can effectively prepare for and help communities withstand the impacts of post-disaster disease outbreaks through effective region-specific mitigation. By further understanding the risk factors involved, outbreaks can be reduced, and this chapter identifies better sanitation and housing as areas for prioritisation.

The results of the review highlight several links between certain disasters, regions and diseases and the work presented in subsequent chapters aim to investigate some of these links. As previously discussed the number of reports for certain disasters and diseases may be due to reporting bias, therefore areas of future research were based around identifying possible links but where a research gap still remained. One example of a link identified here, in the absence of high numbers of papers was drought-related outbreaks in which five out of six occurred in Africa. Droughts are a complex disaster, often being slow-onset and issues arise in assigning a drought start and end point. The conditions which create a drought are also not consistent and vary significantly among geographic locations [63, 64, 65]. These complexities may have prevented previous drought-related work, instead focusing on hazards considered less contentious such as floods.

Water-borne disease outbreaks were common in a drought setting (3/6), of which cholera was reported several times. Cholera was also the second most commonly reported disaster-related disease outbreak (third was diarrhoeal outbreaks with no specific pathogen identified) and was the most common multi-risk factor reporting disease. Africa is both a drought and cholera prone region, while being chronically understudied in terms cholera [66]. Limited research has investigated the links between cholera and drought and many outbreaks may have been missed or not attributed to dry conditions. The next chapter aims to investigate these links in more details, being the first quantitative study to evaluate drought and cholera in isolation and the potential risk factors for these outbreaks.

References

[1] I.K. Kouadio et al. “Infectious diseases following natural disasters: prevention and control measures”. en. In: *Expert Rev. Anti-Infect.* 10 (2012), pp. 95–104.

[2] J.E. Suk et al. “Natural disasters and infectious disease in Europe: a literature review to identify cascading risk pathways”. en. In: *Eur. J. Public Health* (2019).

[3] A. Wilder-Smith. “Tsunami in South Asia: what is the risk of post-disaster infectious disease outbreaks?” en. In: *Ann. Acad. Med. Singap.* 34 (2005), p. 625.

[4] D.A. Walton and L.C. Ivers. “Responding to cholera in post-earthquake Haiti”. en. In: *New Engl. J. Med.* 364 (2011), pp. 3–5.

[5] M. Baqir. “Infectious diseases in the aftermath of monsoon flooding in Pakistan”. en. In: *Asian Pac. J. Trop. Biomed.* 2 (2012), pp. 76–79.

[6] F. Fiasca. “Bacterial Meningitis Hospitalizations after the 2009 L’Aquila Earthquake: A Retrospective Observational Study”. en. In: *Asian J. Epidemiol.* 11 (2018), pp. 46–51.

[7] F. Valente. “Massive outbreak of poliomyelitis caused by type-3 wild poliovirus in Angola in 1999”. en. In: *Bull. World Health Organ.* 78 (2000), pp. 339–346.

[8] J.T. Watson, M. Gayer, and M.A. Connolly. “Epidemics after natural disasters”. en. In: *Emerg. Infect. Dis.* 13, 1 (2007).

[9] M. Gayer et al. “Conflict and emerging infectious diseases”. en. In: *Emerg. Infect. Dis.* 13 (2007), p. 1625.

[10] C.C. Hammer, J. Brainard, and P.R. Hunter. “Risk factors and risk factor cascades for communicable disease outbreaks in complex humanitarian emergencies: A qualitative systematic review”. en. In: *BMJ Glob. Health* 3 (2018), p. 000647.

[11] D. Evans. “Hierarchy of evidence: a framework for ranking evidence evaluating healthcare interventions”. en. In: *J. Clin. Nurs.* 12.1 (2003), pp. 77–84.

[12] Z. Munn et al. “Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach”. en. In: *BMC Med. Res. Methodol.* 18.1 (2018), pp. 1–7.

[13] Z. Munn et al. “What kind of systematic review should I conduct? A proposed typology and guidance for systematic reviewers in the medical and health sciences”. en. In: *BMC Med. Res. Methodol.* 18.1 (2018), pp. 1–9.

[14] D. Moher. “Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement”. io. In: *Syst. Rev.* 4, 1 (2015).

[15] K.S. Khan et al. “Five steps to conducting a systematic review”. en. In: *J. Roy. Soc. Med.* 96 (2003), pp. 118–121.

[16] G.E.C. Charnley et al. “Understanding the risks for post-disaster infectious disease outbreaks: a systematic review protocol”. en. In: *BMJ Open* 10 (2020), p. 039608.

[17] World Health Organization. *Communicable diseases following natural disasters. Risk assessment and priority interventions (World Health Organization)*. en. Geneva, 2006.

[18] S. Khan et al. “Prevalence of HCV and HIV infections in 2005-Earthquake-affected areas of Pakistan”. en. In: *BMC Infect. Dis.* 8 (2008), p. 147.

[19] W. Kimbrough et al. “The burden of tuberculosis in crisis-affected populations: a systematic review”. en. In: *Lancet Infect. Dis.* 12 (2012), pp. 950–965.

[20] J.P. Higgins. *Cochrane handbook for systematic reviews of interventions*. en. New Jersey: John Wiley & Sons, 2019.

[21] National Institute Health. National Health, Lung, and Blood Institute. *Study Quality Assessment Tools*. it. 2020. URL: https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools.

[22] EM-DAT. *Classification*. en. 2009. URL: https://www.emdat.be/classification.

[23] Taiyun W. and Viliam S. *R package "corrplot": Visualization of a Correlation Matrix*. (Version 0.84). 2017. URL: https://github.com/taiyun/corrplot.

[24] P.J. Villacorta Iglesias and A. Signorell. *DescTools: Tools for Descriptive Statistics*. R package version 0.99.47. 2022. URL: https://cran.r-project.org/package=DescTools.

[25] J.G. Shaffer. “Lassa fever in post-conflict Sierra Leone”. it. In: *PLoS Neglect. Trop. Dis.* 8, e274 (2014).

[26] H. Takahashi. “Pneumonia after earthquake”. en. In: *Infect. Dis.* 18 (2011).

[27] F.A. Dureab et al. “Yemen: cholera outbreak and the ongoing armed conflict”. en. In: *J. Infect. Dev. Countr.* 12 (2018), pp. 397–403.

[28] K.A. Caillouët et al. “Increase in West Nile neuroinvasive disease after Hurricane Katrina”. en. In: *Emerg. Infect. Dis.* 14 (2008), p. 804.

[29] O.A. Hassan et al. “The 2007 Rift Valley fever outbreak in Sudan”. en. In: *PLoS Neglect. Trop. Dis.* (2011), p. 5.

[30] P.R. Hunter. “Climate change and waterborne and vector-borne disease”. en. In: *J. Appl. Microbiol.* 94 (2003), pp. 37–46.

[31] V. Tricou et al. “Measles outbreak in Northern Central African Republic 3 years after the last national immunization campaign”. en. In: *BMC Infect. Dis.* 13 (2013), p. 103.

[32] R.M. Garfield et al. “Malaria in Nicaragua: community-based control efforts and the impact of war”. en. In: *Int. J. Epidemiol.* 18 (1989), pp. 434–439.

[33] B. McPake et al. “Ebola in the context of conflict affected states and health systems: case studies of Northern Uganda and Sierra Leone”. en. In: *Confl. Health* 9 (2015), p. 23.

[34] C.R. Wells. “The exacerbation of Ebola outbreaks by conflict in the Democratic Republic of the Congo”. en. In: *PNAS* 116 (2019), pp. 24366–24372.

[35] R. Black et al. “Migration, immobility and displacement outcomes following extreme events”. en. In: *Environ. Sci. Policy* 27 (2013), pp. 32–43.

[36] E. Enarson and J. Scanlon. “Gender patterns in flood evacuation: A case study in Canada’s Red River Valley”. In: *Appl. Behav. Sci. Rev.* 7.2 (1999), pp. 103–103.

[37] Z. He et al. “Burden of common childhood diseases in relation to improved water, sanitation, and hygiene (WASH) among Nigerian children”. en. In: *Int. J. Environ. Res. Public Health* 15 (2018), p. 1241.

[38] A. Ellis. “WASH challenges to girls’ menstrual hygiene management in Metro Manila, Masbate, and South Central Mindanao”. fr. In: *Philippines. Waterlines* 306-323 (2016).

[39] R.G. Garriga and A.P. Foguet. “Unravelling the linkages between water, sanitation, hygiene and rural poverty: the WASH poverty index”. en. In: *Water Resour. Manag.* 27 (2013), pp. 1501–1515.

[40] World Health Organization. *Disease outbreaks*. en. 2020. URL: https://www.who.int/.

[41] V. Fauveau et al. “Diarrhoea mortality in rural Bangladeshi children”. en. In: *J. Trop. Pediatrics* 37 (1991), pp. 31–36.

[42] M.A. Muoki, D.S. Tumuti, and D. Rombo. “Nutrition and public hygiene among children under five years of age in Mukuru slums of Makadara Division”. de. In: *Nairobi. E. Afr. Med. J.* 85 (2008), pp. 386–397.

[43] A. Vivas et al. “Knowledge, attitudes, and practices (KAP) of hygiene among school children in Angolela, Ethiopia”. en. In: *J. Prev. Med. Hyg.* 51, 73 (2010).

[44] M. Romanello et al. “The 2021 report of the Lancet Countdown on health and climate change: code red for a healthy future”. In: *Lancet* 398.10311 (2021), pp. 1619–1662.

[45] N. Watts et al. “The Lancet Countdown on health and climate change: from 25 years of inaction to a global transformation for public health”. In: *Lancet* 391.10120 (2018), pp. 581–630.

[46] Climate Explorer. *Climate Change Atlas*. en. 2020. URL: https://climexp.knmi.nl/plot_atlas_form.py.

[47] M.F.M. Radi. “Leptospirosis outbreak after the 2014 major flooding event in Kelantan, Malaysia: a spatial-temporal analysis”. en. In: *Am. J. Trop. Med. Hyg.* 98 (2018), pp. 1281–1295.

[48] X. Xu et al. “Quantifying the impact of floods on bacillary dysentery in Dalian city”. pt. In: *Public* 11 (2004), pp. 190–195.

[49] J.M. Medlock and S.A. Leach. “Effect of climate change on vector-borne disease risk in the UK”. en. In: *Lancet Infect. Dis.* 15 (2015), pp. 721–730.

[50] L. Vezzulli. “Climate influence on Vibrio and associated human diseases during the past half-century in the coastal North Atlantic”. en. In: *PNAS* 113 (2016), pp. 5062–5071.

[51] U.N. Habitat. “Urbanization and development: emerging futures”. en. In: *World Cities Rep.* 3 (2016), pp. 4–51.

[52] S.A. Kulp and B.H. Strauss. “New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding”. en. In: *Nature Commun.* 10 (2019), pp. 1–12.

[53] D. Dodman. “Blaming cities for climate change? An analysis of urban greenhouse gas emissions inventories”. en. In: *Environ. Urban.* 21 (2009), pp. 185–201.

[54] H.B. Dulal. “Making cities resilient to climate change: identifying “win–win” interventions”. en. In: *Local Environ.* 22 (2017), pp. 106–125.

[55] Y. Bai et al. “How does urbanization affect residential CO_2 emissions? An analysis on urban agglomerations of China”. In: *J. Clean. Prod.* 209 (2019), pp. 876–885.

[56] J. Cuthbertson et al. “Improving Disaster Data Systems to Inform Disaster Risk Reduction and Resilience Building in Australia: A Comparison of Databases”. In: *Prehosp. Disaster Med.* 36.5 (2021), pp. 511–518.

[57] E.H. Chan et al. “Global capacity for emerging infectious disease detection”. In: *PNAS* 107.50 (2010), pp. 21701–21706.

[58] P.D. Drummond et al. “Barriers to accessing health care services for West African refugee women living in Western Australia”. en. In: *Health Care Women In.* 32 (2011), pp. 206–224.

[59] Z. Mumtaz et al. “Gender-based barriers to primary health care provision in Pakistan: the experience of female providers”. en. In: *Health Policy Plann.* 18 (2003), pp. 261–269.

[60] C.R.E.D. *Natural Disasters 2018*. en. Brussels, 2019. URL: https://emdat.be/.

[61] P.B. Stares. *Conflicts to Watch in 2020 (2019*. en. URL: https://www.cfr.org/report/conflicts-watch-2020.

[62] L. Brown and V. Murray. “Examining the relationship between infectious diseases and flooding in Europe: A systematic literature review and summary of possible public health interventions”. en. In: *Disaster Health* 1 (2013), pp. 117–127.

[63] R. Ward, K. Lackstrom, and C. Davis. “Demystifying Drought: Strategies to Enhance the Communication of a Complex Hazard”. en. In: *Bull. Am. Meteorol. Soc.* 103.1 (2022), pp. 181–197.

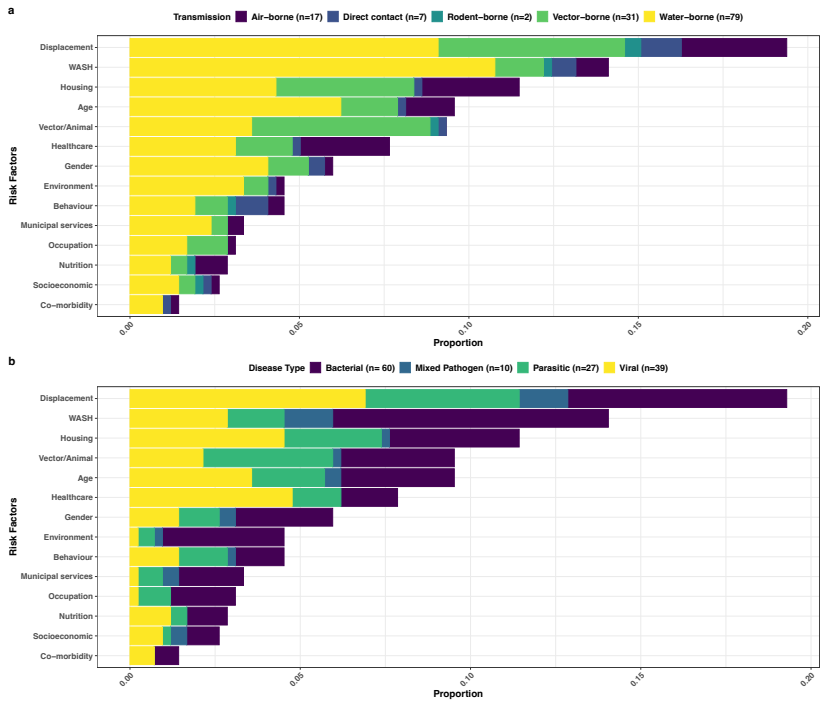
[64] D.A. Wilhite and M. Buchanan-Smith. “Drought and water crises: Science, technology, and management issues”. fr. In: vol. 3. Taylor & Francis, 2005. Chap. Drought as hazard: understanding the natural and social context, p. 29.

[65] D.A. Wilhite, M.D. Svoboda, and M.J. Hayes. “Understanding the complex impacts of drought: A key to enhancing drought mitigation and preparedness”. en. In: *Water Res. Manag.* 21.5 (2007), pp. 763–774.

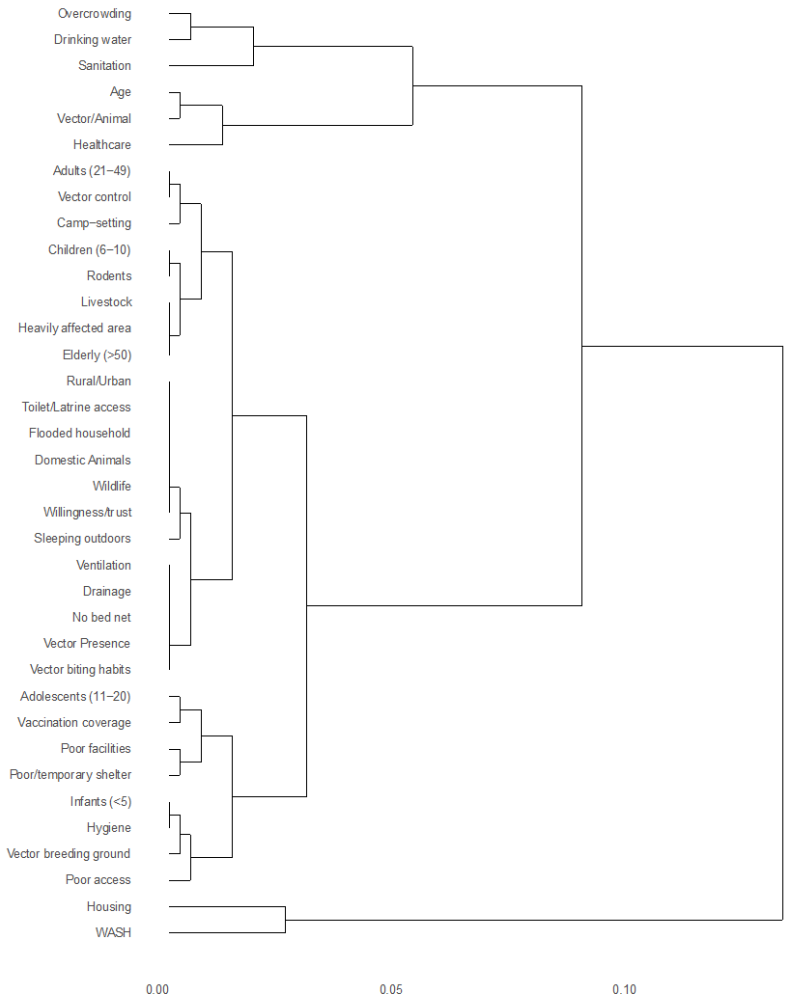
[66] M. Ali et al. “Updated global burden of cholera in endemic countries”. en. In: *PLoS Neglect. Trop. Dis.* 9 (2015), p. 0003832.

Supplementary Material

Supplementary Figures



Supplementary Figure 2.1: Proportions of reported risk factor clusters against disease type and transmission.



Supplementary Figure 2.2: Hierarchical cluster analysis of the top five risk factor clusters, broken down into individual risks reported within the cluster.

Supplementary Tables

Supplementary Table 2.1: The Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) 2015 checklist [14].

Section and topic	Item No	Checklist item
ADMINISTRATIVE INFORMATION		
Title:		
Identification	1a	Identify the report as a protocol of a systematic review
Update	1b	If the protocol is for an update of a previous systematic review, identify as such
Registration	2	If registered, provide the name of the registry (such as PROSPERO) and registration number
Authors:		
Contact	3a	Provide name, institutional affiliation, email address of all protocol authors; provide physical mailing address of corresponding author
Contributions	3b	Describe contributions of protocol authors and identify the guarantor of the review
Amendments	4	If the protocol represents an amendment of a previously completed or published protocol, identify as such and list changes; otherwise, state plan for documenting important protocol amendments

Support:		
Sources	5a	Indicate sources of financial or other support for the review
Sponser	5b	Provide name for the review funder and/or sponsor
Role of sponsor or funder	5c	Describe roles of funder(s), sponsor(s), and/or institution(s), if any, in developing the protocol
INTRODUCTION		
Rationale	6	Describe the rationale for the review in the context of what is already known
Objectives	7	Provide an explicit statement of the question(s) the review will address with reference to participants, interventions, comparators and outcomes (PICO)
METHODS		
Eligibility criteria	8	Specify the study characteristics (such as PICO, study design, setting, time frame) and report characteristics (such as years considered, language, publication status) to be used as criteria for eligibility for the review

Information sources	9	Describe all intended information sources (such as electronic databases, contact with study author, trial registers or other grey literature sources) with planned dates of coverage
Search strategy	10	Present draft of search strategy to be used for at least one electronic database, including planned limits, such that it could be repeated
Study records:		
Data management	11a	Describe the mechanism(s) that will be used to mange records and data throughout the review
Selection process	11b	State the process that will be used for selecting studies (such as two independent reviews) through each phase of the review (that is, screening, eligibility and inclusion in meta-analysis)
Data collection process	11c	Describe planned method of extracting data from reports (such as piloting forms, done independently, in duplicate), any processes for obtaining and confirming data from investigators

Data items	12	List and define all variables for which data will be sought (such as PICO items, funding sources), any pre-planned data assumptions and simplifications
Outcomes and prioritisation	13	List and define all outcomes for which data will be sought, including prioritisation of main and additional outcomes, with rationale
Risk of bias in individual studies	14	Describe anticipated methods for assessing risk of bias of individual studies, including whether this will be done at the outcome or study level, or both; state how this information will be used in data synthesis
Data synthesis	15a	Describe criteria under which study data will be quantitatively synthesised
	15b	If data are appropriate for quantitative synthesis, describe planned summary measures, methods of handling data and methods of combining data from studies, including any planned exploration of consistency
	15c	Describe any proposed additional analyses (such as sensitivity or subgroup analyses, meta-regression)
	15d	If quantitative synthesis is not appropriate, despite the type of summary planned

Meta-bias(es)	16	Specify any planned assessment of meta-bias(es) (such as publication bias across studies, selective reporting within studies)
Confidence in cumulative evidence	17	Describe how the strength of the body of evidence will be assessed (such as GRADE)

Supplementary Table 2.2: Search strategies for MEDLINE, Embase and GlobalHealth.

Category	Keywords	MeSH MEDLINE	MeSH Embase	MeSH GlobalHealth
Natural Hazards	natural hazard* OR natural disaster* OR extreme adj2 event*	climatic process exp, cyclonic storms exp, droughts exp, floods exp, tidal waves exp, geological phenomena exp, avalanches exp, earthquakes exp, landslides exp, tsunamis exp, volcanic eruptions exp, wildfires exp, natural disasters exp	natural disaster exp, disaster victim exp, earthquake exp, drought exp, flooding exp, hurricane exp, tsunami exp, landslide exp, avalanche exp, wildfire exp, volcano exp	natural disaster exp, hurricanes exp, tornados exp, typhoons exp, droughts exp, floods exp, earthquakes exp, landslides exp, avalanches exp, tsunami exp, volcanos exp, wildfire exp
Conflicts	armed conflict* or civil war*	ethnic violence exp, exposure to violence exp, armed conflicts exp, war exposure exp	war exposure exp, ethnic conflict exp	conflict exp, war exp, aggression exp, fighting exp
Disease	infectious disease outbreak* OR communicable disease outbreak*	disease outbreaks exp, epidemics exp, communicable diseases exp, diarrhoea, vibrio infections exp, cholera exp, salmonella infections exp, typhoid fever exp, paratyphoid fever exp, leptospirosis exp, Weil disease, measles exp, measles virus exp, meningitis, bacterial exp, meningitis, escherichia coli exp, meaningococcal exp, pneumococcal exp, respiratory tract infections exp, malaria exp, dengue exp, tetanus exp, clostridium infections exp, haemorrhagic fevers, viral exp, poliomyelitis exp, poliovirus exp, coccidioidomycosis exp, dysentery exp, leishmaniasis, cutaneous exp, leishmaniasis, visceral exp, hepatitis exp, hepatitis a exp, hepatitis e exp	typhoid fever exp, salmonellosis exp, acute hepatitis exp, hepatitis a virus exp, hepatitis e virus exp, hepatitis a exp, hepatitis e exp, leptospirosis exp, measles exp, respiratory tract infection exp, malaria exp, dengue exp, tetanus exp, coccidioidomycosis exp, haemorrhagic fever exp, poliomyelitis exp, dysentery exp, acute diarrhoea exp, meningitis exp, skin leishmaniasis exp, visceral leishmaniasis exp	outbreaks exp, epidemics exp, infectious diseases exp, diarrhoea exp, cholera exp, vibrio cholerae exp, salmonella typhi exp, typhoid exp, salmonella paratyphi exp, paratyphoid exp, hepatitis a exp, hepatovirus a exp, hepatitis e exp, hepatovirus e exp, leptospirosis exp, measles exp, meningitis exp, bacterial meningitis exp, viral meningitis exp, respiratory diseases exp, malaria exp, dengue exp, tetanus exp, coccidioidomycosis exp, leishmaniasis exp, cutaneous leishmaniasis exp, visceral leishmaniasis exp, haemorrhagic fever exp, poliomyelitis exp, dysentery exp

Supplementary Table 2.3: The National Institute of Health's Quality Assessment Tool for the three main study types reviewed in Chapter 2. Studies are rated either 'good', 'fair' or 'poor' depending on how many questions are answered 'yes' vs 'no'.

Observational Cohort and Cross-Sectional Studies

Criteria	**Yes**	**No**	**Other (CD, NR, NA)***
1. Was the research question or objective in this paper clearly stated?			
2. Was the study population clearly specified and defined?			
3. Was the participation rate of eligible persons at least 50%?			
4. Were all the subjects selected or recruited from the same or similar populations (including the same time period)? Were inclusion and exclusion criteria for being in the study prespecified and applied uniformly to all participants?			
5. Was a sample size justification, power description, or variance and effect estimates provided?			
6. For the analyses in this paper, were the exposure(s) of interest measured prior to the outcome(s) being measured?			
7. Was the timeframe sufficient so that one could reasonably expect to see an association between exposure and outcome if it existed?			
8. For exposures that can vary in amount or level, did the study examine different levels of the exposure as related to the outcome (e.g., categories of exposure, or exposure measured as continuous variable)?			
9. Were the exposure measures (independent variables) clearly defined, valid, reliable, and implemented consistently across all study participants?			
10. Was the exposure(s) assessed more than once over time?			

11. Were the outcome measures (dependent variables)
clearly defined, valid, reliable, and implemented consistently
across all study participants?
12. Were the outcome assessors blinded to the exposure
status of participants?
13. Was loss to follow-up after baseline 20% or less?
14. Were key potential confounding variables measured
and adjusted statistically for their impact on the relationship
between exposure(s) and outcome(s)?

Case-Control Studies

1. Was the research question or objective in this paper
clearly stated and appropriate?
2. Was the study population clearly specified and defined?
3. Did the authors include a sample size justification?
4. Were controls selected or recruited from the same or
similar population that gave rise to the cases
(including the same timeframe)?
5. Were the definitions, inclusion and exclusion criteria,
algorithms or processes used to identify or select cases and
controls valid, reliable, and implemented
consistently across all study participants?
6. Were the cases clearly defined and differentiated from controls?
7. If less than 100 percent of eligible cases and/or
controls were selected for the study, were the cases and/or
controls randomly selected from those eligible?
8. Was there use of concurrent controls?
9. Were the investigators able to confirm that the exposure/risk
occurred prior to the development of the condition or event that
defined a participant as a case?

10. Were the measures of exposure/risk clearly defined, valid, reliable, and implemented consistently (including the same time period) across all study participants?

11. Were the assessors of exposure/risk blinded to the case or control status of participants?

12. Were key potential confounding variables measured and adjusted statistically in the analyses? If matching was used, did the investigators account for matching during study analysis?

*CD, cannot determine; NA, not applicable; NR, not reported

Supplementary Table 2.4: Full list of reported risk factors, their cluster and how the cluster was defined (in *italics*).

Risk Cluster	**Risk Factor**	**No. of outbreaks**
Displacement		81
A report of national population movement due to the disaster	General	79
	Rural to urban	1
	Coastal to jungle	1
Age		40
Reported age-related risk factors, either a demographic group e.g., children, or a specific age category e.g., <5 years.	<2	2
	<4	1
	<5	10
	>5	2
	>5	2
	3-8	1
	<7	1
	<10	1
	Child	2
	<14	1
	15-20	1
	15-19	1
	<15	2
	15-34	1
	<18	2
	<20	2
	40-49	1
	Adults	3

	>50	2
	>65	1
	Older	1
	Elderly	2
Gender		25
Males or females being more at risk	Male	20
	Female	5
WASH		59
Any issues with access or quality of water, sanitation and hygiene provisions, separate from disruptions to specific municipal services.	Drainage	1
	Sanitation	25
	Hygiene	13
	Drinking water	17
	Toilet/latrine access	3
Housing		48
Reports of inadequate living conditions or the location of either habitual residence or temporary housing provided after displacement.	Overcrowding	19
	Urban	1
	Rural	1
	Mountainous	1
	Flooded household	3
	Living a heavily impacted area	6
	Camp-setting	5

	Poor shelter	6
	Sleeping outdoors	2
	Collapsed infrastructure	1
	No bed net	1
	Temporary shelter	2
	Ventilation	1
Healthcare		35
Any issue that prevented people seeking formal health care	Poor access	12
	Vaccination coverage	11
	Poor facilities	9
	Willingness/trust to seek care	3
Municipal services		14
Disruption to municipal services	Garbage	4
	Waste	6
	Water	4
Environment		19
Alterations in the natural environment that exacerbated risk of contracting the disease.	Contact with floodwater	13
	Higher temperatures	2
	Alterations in land moisture	1
	Water salinity	1
	Geological changes	1
	Lower temperatures	1
Vector/Animal		38

Changes in animals and vectors that accelerated contact with the population and subsequent disease spread.	Livestock (cattle, pigs, camels)	6
	Rodents	7
	Domestic animals (dogs)	3
	Wildlife (monkeys, bandicoots, beavers)	3
	Alterations in vector breeding ground	14
	Vector control	4
	Vector biting habits	1
	Vector presence	1
	Exposure to animals	1
Behaviour		19
Any report of human behaviour which heightened the risk of contact with the pathogen, except human displacement.	Recreation	1
	Swimming	1
	Hiking	1
	Camping	1
	Assisting in clean up	1
	Method of acquiring water and storing	3
	Mixing with people of different immunities	2
	Sexual contact	1
	Not being covered	1
	Sharing combs	1
	Burial practices	2

	Public distrust	1
	Nomadic	1
Occupation		13
An occupation which was associated with increased pathogen exposure and therefore disease.	Rice paddy farmer	2
	Street vendor	1
	Unemployed	1
	Homemaker	1
	Farmer	3
	Military	2
	Working outdoors	1
	Hunting	1
	International workers	1
Nutrition		12
Issues with insufficient diet or eating specific foods.	No fruits and vegetables	1
	Eating rodents	1
	Eating monkeys	1
	Malnourishment	7
	Drinking sugar cane juice	1
	Eating millet gruel	1
Co-morbidity		6
Significant numbers of infected individuals also presented with another morbidity.	Pregnancy	2
	Co-morbidities	2
	Respiratory tract infections	1

	Psychological conditions	1
Socio-economic		11
Further socioeconomic conditions that increased cases but did not fit into any other cluster. These mainly included education and inequities.	Poverty	4
	Literacy rate	1
	Poor socio-economics	2
	Education level	2
	Attending church school	1
	Fathers education	1

Supplementary Table 2.5: Full list of reported countries.

Region	Country	Frequency
South & South East Asia	India	12
	Nepal	1
	Bangladesh	4
	Pakistan	1
	Sri Lanka	3
	Sumatra	2
	Philippines	3
	Malaysia	1
	Thailand	2
	Indonesia	2
Middle East	Syria	3
	Iraq	2
	Iran	2
	West Bank	1
	Afghanistan	3
	Yemen	4
	Oman	1
	Lebanon	1
Oceania	Australia	1
	New Zealand	1
	Fiji	1
	Papua New Guinea	2
	Solomon Islands	1
Africa	Libya	1
	Kenya	3
	Ethiopia	2
	Somalia	2

	Sierra Leone	2
	Uganda	2
	Liberia	1
	Mauritania	1
	Mali	1
	Sudan	5
	South Sudan	4
	Central African Republic	3
	Democratic Republic of Congo	2
	Angola	1
	Mozambique	2
	Burundi	1
East Asia	China	9
	Japan	4
	Taiwan	2
Europe	Bosnia & Herzegovina	2
	Yugoslavia	2
	Czechoslovakia	1
	Croatia	1
	Turkey	2
	Italy	2
	Netherlands	1
	UK	1
	Germany	1
North America	USA	10
Latin America	Chile	1
	Ecuador	1
	Columbia	1
	Guyana	1

Mexico	1
Nicaragua	2
Costa Rica	1
Honduras	1
Haiti	2

Supplementary Table 2.6: Full list of reported disease outbreaks.

Disease Type	Transmission	Disease	No. of outbreaks
Bacterial	Water borne	Leptospirosis	21
	Water borne	Cholera	15
	Water borne	Diarrhoeal disease	7
	Water borne	Dysentery	6
	Water borne	Typhoid	3
	Vector borne	Typhus	2
	Air borne	Pneumonia	2
	Water borne	Bacteremia	1
	Air borne	Diphtheria	1
	Air borne	Meningitis	1
	Air borne	Respiratory disease	1
	Rodent borne	Tularemia	1
Viral	Air borne	Measles	5
	Water borne	Hepatitis E	5
	Water borne	Polio	4
	Vector borne	Dengue	3
	Water borne	Diarrhoeal disease	3
	Direct contact	Ebola	2
	Vector borne	Rift Valley fever	2
	Water borne	Hepatitis A	2
	Air borne	Influenza	2
	Vector borne	West Nile	2
	Water borne	Norovirus	2
	Vector borne	Chikungunya	1
	Rodent borne	Hantavirus	1
	Direct contact	Lassa fever	1
	Air borne	Monkeypox	1
	Air borne	Respiratory disease	1
	Air borne	Rubella	1
	Vector borne	Yellow Fever	1
	Vector borne	Zika	1
Parasitic	Vector borne	Malaria	10
	Vector borne	Cutaneous leishmaniasis	5
	Water borne	Giardiasis	3
	Vector borne	Visceral leishmaniasis	3
	Water borne	Diarrhoeal disease	2
	Vector borne	Sleeping Sickness	2
	Water borne	Cryptosporidium	1
	Water borne	Fascioliasis	1
	Direct contact	Scabies	1
	Water borne	Schistosomiasis	1
Mixed pathogen	Water borne	Diarrhoeal disease	6
	Direct contact	Dermatological disease	3
	Air borne	Respiratory disease	1
		Total	140

Supplementary Table 2.7: List of full p values from pair-wise comparisons. * shows significance at <0.05. Blank cells in the table indicate no outbreaks that fit into both the categories.

	p value
Risk Factor Cluster	$7.2x10^{-36}$
Conflict/Displacement	0.04
Geophysical/Displacement	0.22
Hydrological/WASH	0.58
Conflict/WASH	0.14
Water-borne/WASH	0.0022
Geophysical/Housing	0.001
Hydrological/Vector	0.023
Conflict/Vector	0.126
Parasitic/Vector	0.004
Water-borne/Age	0.14
Conflict/Age	0.68
Conflict/Healthcare	0.0000003
Viral/Healthcare	0.0014
Hydrological/Gender	0.004
Water-borne/Gender	0.15

Actual p value	**Africa**	**E Asia**	**Europe**	**LAC**	**Middle East**	**N America**	**Oceania**	**S & SE Asia**
Conflict	0.0000197*		0.79565613	0.04237575*	0.00000001*		0.38814874	0.00055075*
Hydrological	0.03742919*	0.08779559	0.71072159	0.49736414	0.00819448*	0.29446618	0.31503477	0.04128529*
Climatological	0.00027878*					0.42591199		
Geophysical		0.2537846	0.55776124	0.01565516*	0.07715278	0.33047572	0.42309951	0.04748551*
Meteorological	0.11744737	0.2876211		0.14102256		0.00488033*		0.63602109
	Water-borne	**Vector-borne**	**Air-borne**	**Rodent-borne**	**Direct contact**			
Conflict	0.00002757*	0.01827592*	0.32309873	0.04164768*	0.16000299			
Hydrological	0.00000015*	0.00217725*	0.0211398*		0.31688118			
Climatological	0.16244772	0.11471884			0.18858594			
Geophysical	0.00972217*	0.21803643	0.01387154*		0.67859545			
Meteorological	0.00595018*	0.10543734	0.52184744					
	Bacterial	**Viral**	**Parasitic**	**Mixed pathogen**				
Conflict	0.00029837*	0.00604762*	0.0463105*	0.11010373				
Hydrological	0.00362129*	0.00795126*	0.16208116	0.04285716*				
Climatological	0.5726781	0.81973714	0.45566813					
Geophysical	0.53167598	0.70071662	0.57786094	0.39523447				
Meteorological	0.01736218*	0.40642793		0.3410774				

Supplementary Information

Supplementary Information 2.1: Full list of included studies in the review (Chapter 2).

1. Bhardwaj P. et al. A case control study to explore the risk factors for acquisition of leptospirosis in Surat city, after flood. Indian J Med Sci 2008;62(11).

2. Hussein N.R. et al. A clinical study of cutaneous leishmaniasis in a new focus in the Kurdistan region, Iraq. PloS One 2018;14(5).

3. Ventura R.J. et al. A community-based gastroenteritis outbreak after Typhoon Haiyan, Leyte, Philippines, 2013. Western Pac Surveill Response J 2015;6(1):1.

4. Togami E. et al. A large leptospirosis outbreak following successive severe floods in Fiji, 2012. Am J Trop Med 2018;99(4):849-851.

5. Roy S. et al. Acute diarrhea in children after 2004 tsunami, Andaman Islands. Emerg Infect Dis 2009;15(5):849.

6. McCarthy M.C. et al. Acute hepatitis E infection during the 1988 floods in Khartoum, Sudan (No. NMRI-94-15). NMRID 1994.

7. Yani F.F. et al. Acute respiratory infection in children after earthquake on September 30th, 2009 and its related factors in Malalak and Maninjau villages at Agam District West Sumatera. Paediatr Respir Rev 2010;(11):S111.

8. Balasubramaniam S.M. and Roy G. An outbreak investigation of measles after the tsunami in Cuddalore district, Tamil Nadu. Ann Trop Med Public Health 2012;5(4):291.

9. Wei L. et al. An outbreak of bacillary dysentery caused by flash floods polluted water. Prev Med 2012;39(13):3203-3204. [Abstract only]

10. Pradutkanchana S. et al. An outbreak of leptospirosis after severe flood in Hat Yai in 2000. J Infect Dis and Antimicrob Agents 2002;19(1):pp.9-13.

11. Hukic M. et al. An outbreak of rubella in the Federation of Bosnia and Herzegovina between December 2009 and May 2010 indicates failure to vaccinate during wartime (1992–1995). Epidemiol & Infect 2012;140(3):447-453.

12. Weniger B.G. et al An outbreak of waterborne giardiasis associated with heavy water runoff due to warm weather and volcanic ashfall. Am J Public Health 1983;73(8):868-872.

13. Zhang F. et al. Association between flood and the morbidity of bacillary dysentery in Zibo City, China: a symmetric bidirectional case-crossover study. Int J Biometeorol. 2016;60(12):pp.1919-1924.

14. Liu Z. et al. Association between floods and typhoid fever in Yongzhou, China: effects and vulnerable groups. Environ Res 2018;167:pp.718-724.

15. Fiasca F. et al. Bacterial Meningitis Hospitalizations after the 2009 L'Aquila Earthquake: A Retrospective Observational Study. Asian J Epidemiol 2018;11(1):pp.46-51.

16. Tanabe K. et al. Bacteriological survey of diarrheal epidemics in the 1998 Bangladesh floods. Kansenshogaku zasshi. Jpn J Infect Dis 1999;73(9):918-922. [Abstract only]

17. Montero D. et al. Characterization of enterotoxigenic Escherichia coli strains isolated from the massive multi-pathogen gastroenteritis outbreak in the Antofagasta region following the Chilean earthquake, 2010. Infect Genet Evol 2017;52:26-29.

18. CDC. Cholera epidemic after increased civil conflict–Monrovia, Liberia, June-September 2003. MMWR 2003;52(45):1093.

19. Fredrick T. et al. Cholera outbreak linked with lack of safe water supply following a tropical cyclone in Pondicherry, India, 2012. J Health Popul Nutr 2015;33(1):31.

20. Yuan X. Concern over reported number of measles cases in Yemen. The Lancet, 2018;391(10133):1886.

21. Rosewell A. et al. Concurrent outbreaks of cholera and peripheral neuropathy associated with high mortality among persons internally displaced by a volcanic eruption. PloS One 2013;8(9).

22. Norris A., et al. Crippling violence: Conflict and incident polio in Afghanistan. PLoS One 2016;11(3).

23. Al-Salem W.S. et al. Cutaneous leishmaniasis and conflict in Syria. Emerg Infect Dis 2016;22(5):931.

24. Fanany D. Dengue hemorrhagic fever and natural disaster: the case of Padang, West Sumatra. Int J Collab Res Intern Med Public Health 2012;4(5).

25. Alghazali K.A. et al. 2019. Dengue outbreak during ongoing civil war, Taiz, Yemen. Emerg Infect Dis 2019;25(7):1397.

26. Schwartz B.S. et al. Diarrheal epidemics in Dhaka, Bangladesh, during three consecutive floods: 1988, 1998, and 2004. Am J Trop Med Hyg 2006;74(6):1067-1073.

27. Dureab, F. et al. Diphtheria outbreak in Yemen: the impact of conflict on a fragile health system. Confl Health 2019;13(1):19.

28. Chretien J.P. et al. Drought-associated chikungunya emergence along coastal East Africa. Am J Trop Med Hyg 2007;76(3):405-407.

29. McPake B. et al. Ebola in the context of conflict affected states and health systems: case studies of Northern Uganda and Sierra Leone. Confl Health 2015;9(1):23.

30. Claude K.M. et al. Ebola virus epidemic in war-torn eastern DR Congo. The Lancet 2018;392(10156):1399-1401.

31. Tauxe R.V. et al. Epidemic cholera in Mali: high mortality and multiple routes of transmission in a famine area. Epidemiol & Infect 1988;100(2):279-289.

32. Karim A.M. et al. Epidemiology and clinical burden of malaria in the war-torn area, Orakzai Agency in Pakistan. PLoS Neglect Trop Dis 2016;10(1).

33. Hashizume M. et al. Factors determining vulnerability to diarrhoea during and after severe floods in Bangladesh. J Water Health 2008;6(3):323-332.

34. Sidley P. Floods in Mozambique result in cholera and displacement. BMJ 2008;336(7642):471-471.

35. Lora-Suarez F. et al. Giardiasis in children living in post-earthquake camps from Armenia (Colombia). BMC Public Health 2002;2(1):5.

36. Bugert J.J. et al. Hantavirus infection—haemorrhagic fever in the Balkans—potential nephrological hazards in the Kosovo war. Nephrol Dial Transplant 1999;14(8):1843-1844.

37. Kaddoura M. et al. Hepatitis A Virus Genotype IB Outbreak among Internally Displaced Persons, Syria. Emerg Infect Dis 2020;26(2):369.

38. Wei W. et al. Impacts of typhoon 'Koppu'on infectious diarrhea in Guangdong Province, China. Biomed Environ Sci 2015;28(12):920-923.

39. Caillouët K.A. et al. Increase in West Nile neuroinvasive disease after hurricane Katrina. Emerg Infect Dis 2008;14(5):804.

40. Schousboe M. et al. Increased incidence of Escherichia coli bacteremia post-Christchurch earthquake 2011: possible associations. Prehosp Disaster Med 2013;28(3):202-209.

41. Jones F.K. et al. Increased rotavirus prevalence in diarrheal outbreak precipitated by localized flooding, Solomon Islands, 2014. Emerg Infect Dis 2016;22(5):875.

42. Chen W. et al. Influence of flood in 1998 on schistosomiasis epidemic. Chin. J. Schistosomiasis Control 2000;12(4):202-205. [Abstract only]

43. Kalthan E. et al. Investigation of an outbreak of monkeypox in an area occupied by armed groups, Central African Republic. Med Mal Infect 2018;48(4):263-268.

44. Enbiale W. and Ayalew A. Investigation of a scabies outbreak in drought-affected areas in Ethiopia. Trop Med Infect Dis 2018;3(4):114.

45. Shaffer J.G. et al. Lassa fever in post-conflict Sierra Leone. PLoS Negl Trop Dis 2014;8(3):e2748.

46. Zúñiga Carrasco I.R. et al. Later leptospirosis after flood in Tabasco, Mexico, 2007. Enferm. Infec 2011;31(1):33-37. [Abstract only]

47. Chiu C.H. et al. Leptospirosis after typhoon in Taiwan. J Med Sci 2009;29(3):131-134.

48. Pellizzer P. et al. Leptospirosis following a flood in the Veneto area, North-east Italy. Ann Ig: Med Prev Comunita 2006;18(5):453-456. [Abstract only]

49. Smith J.K. et al. Leptospirosis following a major flood in Central Queensland, Australia. Epidemiol & Infect 2013;141(3):585-590.

50. Supe A. et al. 2018. Leptospirosis following heavy rains in 2017 in Mumbai: Report of large-scale community chemoprophylaxis. Nat Med J India 2018;31(1):19.

51. Easton A. Leptospirosis in Philippine floods. BMJ 1999;319(7204):212.

52. Gaynor K. et al. Leptospirosis on Oahu: an outbreak associated with flooding of a university campus. American J Trop Med Hyg 2007;76(5):882-886.

53. Radi M. et al. 2018. Leptospirosis outbreak after the 2014 major flooding event in Kelantan, Malaysia: a spatial-temporal analysis. American J Trop Med Hyg 2018;98(5):1281-1295.

54. World Health Organization. Leptospirosis, India: report of the investigation of a post-cyclone outbreak in Orissa, November 1999. WER 2000;75(27):217-223.

55. Salahi-Moghaddam A. Low-altitude outbreaks of human fascioliasis related with summer rainfall in Gilan province, Iran. Geospat Health 2011:133-136.

56. Wiwanitkit V. Correspondence-Malaria and dengue infection after Tsunami in Southern Thailand. Trop Doct 2007;37(3):194-194.

57. Manimunda S.P. et al. Malaria in Car Nicobar Island in the aftermath of the tsunami: Some observations. Natl Med J India 2009;22(4):217-218.

58. Snuparek J. Malaria in Czechoslovakia after World War II. J Parasitol 1947;33(6):506-508.

59. Garfield R.M. et al. Malaria in Nicaragua: community-based control efforts and the impact of war. Int J Epidemiol 1989;18(2):434-439.

60. Baomar A.T. and Mohamed A.G. Malaria outbreak in a malaria-free region in Oman 1998: unknown impact of civil war in Africa. Public Health 2000;114(6):480-483.

61. Muriuki D., et al. Cross-sectional survey of malaria prevalence in tsunami-affected districts of Aceh Province, Indonesia. Int J Emerg Med 2012;5(1):11.

62. Valente F. et al. 2000. Massive outbreak of poliomyelitis caused by type-3 wild poliovirus in Angola in 1999. Bull World Health Organ 2000;78:339-346.

63. Tekpa G. et al. 2019. Measles Epidemic in Post-Conflict Period in Rural Central African Republic. Med Afr Noire 2019;66(7):387–92. [Abstract only]

64. Ahmad K. Measles epidemic sweeps through Afghanistan. The Lancet 2000;355(9213):1439.

65. Cambaza E. et al. Outbreak of cholera due to Cyclone Kenneth in northern Mozambique, 2019. Int J Environ Res Public Health 2019;16(16):2925.

66. CDC. Outbreak of diarrheal illness associated with a natural disaster–Utah. MMWR 1983;32(50):662.

67. Gertler M. et al. Outbreak of Cryptosporidium hominis following river flooding in the city of Halle (Saale), Germany, August 2013. BMC Infect Dis 2015;15(1):88.

68. Vanasco N.B. et al. Outbreak of human leptospirosis after a flood in Reconquista, Santa Fe, 1998. Rev. Argent. Microbiol 2001;34(3):124-131. [Abstract only]

69. Al-Shere T.A. et al. Outbreak of leptospirosis after flood, the Philippines, 2009. Emerging Infect Dis 2012;18(1):91.

70. Nomura K. et al. 2008. Outbreak of norovirus gastroenteritis in elderly evacuees after the 2007 Noto Peninsula earthquake in Japan. J Am Geriat Soc 2008;56(2):361-363.

71. Takahashi H. et al. Pneumonia after earthquake, Japan, 2011. Emerg Infect Dis 2012;18(11):1909.

72. Kamadjeu R. et al. Polio outbreak investigation and response in Somalia, 2013. J Infect Dis 2014;210:S181-S186.

73. Sáenz R. et al. Post-disaster malaria in Costa Rica. Prehosp Disaster Med 1995;10(3):154-160.

74. Fakoorziba M.R. Post-earthquake outbreak of cutaneous leishmaniasis in a rural region of southern Iran. Ann Trop Med Parasitol 2011;105(3):217-224.

75. Karmakar S. et al. 2008. Post-earthquake outbreak of rotavirus gastroenteritis in Kashmir (India): An epidemiological analysis. Public Health 2008;122(10):981-989.

76. Ortiz M.R. et al. Post-earthquake Zika virus surge: Disaster and public health threat amid climatic conduciveness. Sci Rep 2017;7(1):1-10.

77. Hatta M. et al. Post-tsunami outbreaks of influenza in evacuation centers in Miyagi Prefecture, Japan. Clin Infect Dis 2012;54(1):e5-e7.

78. Frawley A.A. et al. Postflooding leptospirosis - Louisiana, 2016. MMWR 2017;66(42):1158-1159.

79. Xu X. et al. uantifying the impact of floods on bacillary dysentery in Dalian city, China, from 2004 to 2010. Disaster Med Public Health Prep 2017;11(2):190-195.

80. Liu X. et al. Quantitative analysis of burden of bacillary dysentery associated with floods in Hunan, China. Sci Total Environ 2016;547:190-196.

81. Agampodi S.B. et al. Regional differences of leptospirosis in Sri Lanka: observations from a flood-associated outbreak in 2011. PLoS Neglected Trop Dis 2014;8(1).

82. Mbaeyi C. et al. Response to a large polio outbreak in a setting of conflict—Middle East, 2013–2015. MMWR 2017;66(8):227.

83. Moore A. Resurgence of sleeping sickness in Tambura County, Sudan. Am J Trop Med Hyg 1999;61(2):315-318.

84. Sur D. Severe cholera outbreak following floods in a northern disrict of West Bengal. Indian J Med Res 2000;112:178. [Abstract only]

85. Mamova A. et al. Severe malaria including cerebral malaria among 3707 admissions in South Sudanese hospital for internally displaced population after tribual conflicts in 2012-2013. Clin Soc Work 2015:65.

86. Suzuki M. et al. 2011. Shelter-acquired pneumonia after a catastrophic earthquake in Japan. J Am Geriat Soc 2011;59(10):1968-1970.

87. Naranjo M. et al. Study of a leptospirosis outbreak in Honduras following Hurricane Mitch and prophylactic protection of the vax-SPIRAL® vaccine. MEDICC Rev 2008;10(3):38-42.

88. Hassan O.A. The 2007 rift valley fever outbreak in Sudan. PLoS Neglect Trop Dis 2011;5(9).

89. Liu X. et al. The effects of floods on the incidence of bacillary dysentery in Baise (Guangxi Province, China) from 2004 to 2012. Int J Environ Res Public Health 2017;14(2):179.

90. Vahaboglu H. Transient increase in diarrheal diseases after the devastating earthquake in Kocaeli, Turkey: results of an infectious disease surveillance study. Clin Infect Dis 2000;31(6):1386-1389.

91. Reintjes R. Tularemia outbreak investigation in Kosovo: case control and environmental studies. Emerg Infect Dis 2002;8(1):69.

92. Bradarić N. Two outbreaks of typhoid fever related to the war in Bosnia and Herzegovina. Eur J Epidemiol 1996;12(4):409-412.

93. Su H.P. Typhoon-related leptospirosis and melioidosis, Taiwan, 2009. Emerg Infect Dis 2011;17(7):1322.

94. O'Connor K.A. et al. Risk factors early in the 2010 cholera epidemic, Haiti. Emerg Infect Dis 2011;17(11):2136.

95. CDC. Vibrio illnesses after Hurricane Katrina–multiple states, August-September 2005. MMWR 2005;54(37):928.

96. Gorski S. et al. Visceral leishmaniasis relapse in Southern Sudan (1999–2007): a retrospective study of risk factors and trends. PLoS Neglect Trop Dis 2010;4(6).

97. Punda-Polić V. War-associated cases of typhoid fever imported to Split-Dalmatia County (Croatia). Military Med 2007;172(10):1096-1098.

98. Bhunia R. and Ghosh S. Waterborne cholera outbreak following cyclone Aila in Sundarban area of West Bengal, India, 2009. Trans R Soc Trop Med Hyg 2011;105(4):214-219.

99. Dureab F.A. et al. Yemen: cholera outbreak and the ongoing armed conflict. J Infect Dev Countr 2018;12(05):397-403.

100. Qadri F. et al. Enterotoxigenic Escherichia coli and Vibrio cholerae diarrhea, Bangladesh, 2004. Emerg Infect Dis 2005;11(7), p.1104.

101. DeMan H. et al. Gastrointestinal, influenza-like illness and dermatological complaints following exposure to floodwater: a cross-sectional survey in The Netherlands. Epidemiol & Infect 2016;144(7):1445-1454.

102. Ford L.B. Civil conflict and sleeping sickness in Africa in general and Uganda in particular. Confl Health 2007;1(1):6.

103. Shikanga O.T. et al. 2009. High mortality in a cholera outbreak in western Kenya after post-election violence in 2008. Am J Trop Med Hyg 2009;81(6):1085-1090

104. Marlet M.V.L. et al. Emergence or re-emergence of visceral leishmaniasis in areas of Somalia, northeastern Kenya, and south-eastern Ethiopia in 2000–2001. Trans R Soc Trop Med Hyg 2003;97(5):515-518.

105. Bastola A. et al Aftermath earthquake in Nepal: burden of scrub typhus cases and their presentations. J Trop Dis 2017;5(1):236.

106. Trevejo R.T. et al. Epidemic leptospirosis associated with pulmonary hemorrhage—Nicaragua, 1995. J Infect Dis 1998;178(5):1457-1463.

107. Ding G. et al. Quantitative analysis of burden of infectious diarrhea associated with floods in northwest of Anhui Province, China: a mixed method evaluation. PLOS One, 2013;8(6).

108. Reacher M. et al. Health impacts of flooding in Lewes: a comparison of reported gastrointestinal and other illness and mental health in flooded and non-flooded households. Commun Dis Public Health 2004;7:39-46.

109. Tricou V., et al. Measles outbreak in Northern Central African Republic 3 years after the last national immunization campaign. BMC Infect Dis 2013;13(1):103.

110. Dechet A.M. et al. Leptospirosis outbreak following severe flooding: a rapid assessment and mass prophylaxis PLoS One 2012;7(7):e39672.

111. Maskey M. et al. Leptospirosis in Mumbai: post-deluge outbreak 2005. Indian J Med Microbiol 2006;24(4):337.

112. Raoult D. et al. Outbreak of epidemic typhus associated with trench fever in Burundi. The Lancet 1998;352(9125):353-358.

113. Öztaş M.O. Early skin problems after Düzce earthquake. Int J Dermatol 2000;39(12):952-953.

114. Alawieh A. Revisiting leishmaniasis in the time of war: the Syrian conflict and the Lebanese outbreak. Int J Infect Dis 2014;29:115-119.

115. El Mamy A. et al. Unexpected rift valley fever outbreak, Northern Mauritania. Emerg Infect Dis 2011;17(10):1894.

116. Reithinger R. et al. Anthroponotic cutaneous leishmaniasis, Kabul, Afghanistan. Emerg Infect Dis 2003;9(6):727

117. Dahanayaka N.J. et al. Massive hepatitis A outbreak in Sri Lanka in 2009: an indication of increasing susceptibility and epidemiological shift?. Sri Lanka J Infect Dis 2013;3(2).

118. Pal S. et al. An outbreak of hepatitis A virus among children in a flood rescue camp: a post-disaster catastrophe. Indian J Med Microbiol 2016;34(2):233.

119. Boccia D. et al. High mortality associated with an outbreak of hepatitis E among displaced persons in Darfur, Sudan. Clinical infectious diseases, 2006;42(12):1679-1684.

120. Brooks H.M. Malaria in an Internally Displaced Persons Camp in the Democratic Republic of the Congo. Clin Infect Dis 2017;65(3):529-530.

121. Sugunan A.P. et al. Outbreak of rotaviral diarrhoea in a relief camp for tsunami victims at Car Nicobar Island, India. J Public Health 2007;29(4):449-450.

122. Yee E.L. et al. Widespread outbreak of norovirus gastroenteritis among evacuees of Hurricane Katrina residing in a large "megashelter" in Houston, Texas: lessons learned for prevention. Clin Infect Dis 2007;44(8):1032-1039.

123. Elfaituri S.S., Skin diseases among internally displaced Tawerghans living in camps in Benghazi, Libya. Int J Dermatol 2016;55(9):1000-1004.

124. Watanabe H. et al. Possible prevalence and transmission of acute respiratory tract infections caused by Streptococcus pneumoniae and Haemophilus influenzae among the internally displaced persons in tsunami disaster evacuation camps of Sri Lanka. Intern Med, 2007;46(17):1395-1402.

125. Benny E. et al. A large outbreak of shigellosis commencing in an internally displaced population, Papua New Guinea, 2013. Western Pacific surveillance and response journal: WPSAR, 2014;5(3):18.

126. Malholland K. Cholera in Sudan: an account of an epidemic in a refugee camp in eastern Sudan, May–June 1985. Disasters, 1985;9(4):247-258.

127. Liu L. et al. Influenza A (H3) outbreak at a Hurricane Harvey Megashelter in Harris County, Texas: Successes and challenges in disease identification and control measure implementation. Disaster Med Public Health Prep 2019;13(1):97-101.

128. Thomson K. et al. Investigation of hepatitis E outbreak among refugees—Upper Nile, South Sudan, 2012–2013. MMWR 2013;62(29):581.

129. Nogodalla M.A. Yellow fever outbreak investigation and response, Darfur State, Sudan, September-November 2012. Int J Infect Dis 2014;21:259.

130. Abu-Alrub S.M. et al. Prevalence of Cryptosporidium spp. in children with diarrhoea in the West Bank, Palestine. Journal Infect Dev Countr 2008;2(01):059-62.

131. Hulland E. et al. Increase in Reported Cholera Cases in Haiti Following Hurricane Matthew: An Interrupted Time Series Model. American J Tropical Med Hyg 2019;100(2):368-373.

132. Shaman J. et al. 2005. Drought-induced amplification and epidemic transmission of West Nile virus in southern Florida. J Med Entomol 2005;42(2):134-141.

Chapter 3

Exploring Relationships between Drought and Epidemic Cholera in Africa using Generalised Linear Models

Dissemination

An extended version of the introduction for this chapter is published at:

Charnley GEC, Kelman I, Murray KA. Drought-related cholera outbreaks in Africa and the implications for climate change: a narrative review. *Pathogens and Global Health* 2022;116(1):3-12.

A modified version of the full chapter is published at:

Charnley GEC, Kelman I, Green N, Hinsley W, Gaythorpe KAM, Murray KAM. Exploring relationships between drought and epidemic cholera in Africa using generalised linear models. *BMC Infectious Diseases* 2021;21:1177.

Abstract

Temperature and precipitation are known to affect *Vibrio cholerae* outbreaks, but 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. Chapter 3 analysed a range of environmental and socio-economic publicly available national data and fit generalised linear models to test for associations with drought. Using the best fit model, cholera outbreak projections for Africa were calculated to 2070 under three scenarios of varying trajectories of CO_2 emissions and socio-economic development. The best fit model found 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. Despite an effect of drought in explaining recent cholera outbreaks, future projections highlighted the potential for sustainable development and emissions reductions to offset drought-related impacts on cholera risk in the future.

3.1 Introduction

Cholera was reintroduced into Africa in the 1970s during the seventh and continuing cholera pandemic, with many countries showing signs of endemicity. It has since caused significant mortality and morbidity, especially amongst the most vulnerable [1]. Despite over 94% of WHO reported cholera cases occurring in Africa and some of the highest mortality rates [2], previous research has heavily focused on South America, the Indian subcontinent and more recently Haiti. Additionally, other disease outbreaks have drawn attention away from cholera in recent years, including COVID-19 and Ebola [3, 4].

The relationships between temperature, precipitation and water-borne disease outbreaks, including cholera, have implications for outbreaks following natural hazards, especially droughts. Several links between drought and cholera outbreaks have been described [5, 6, 7], particularly those relating to food (e.g., difficulties accessing certain foodstuff and less reheating due to limited cooking fuel) and water behaviours (e.g., limited water sources leading to multi-use water), which are important for cholera transmission [5, 8]. Despite this, research has heavily focused on the link between flooding and cholera, despite droughts potentially posing a considerably greater risk than floods [7].

The socio-economic links (e.g., poverty, WASH, healthcare) associated with cholera 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 [9]. WASH factors are considered particularly significant, as the importance of cholera environmental reservoirs (e.g., waterbodies) depends on the sanitary conditions of the community [10]. In Africa, 418 million people lack basic drinking water, 779 million are without basic sanitation services (including 208 million who still practice open defecation) and 839 million don't have basic hygiene services. [11]. This puts millions of people on the continent at risk of water-borne disease outbreaks including cholera and more vulnerable to the effects of drought.

Understanding the implications of drought on cholera is especially important considering the suggested changes in drought frequency and intensity which climate change may cause [12], while considering how socio-economic factors may play a role in these changes and the populations ability to adapt. Few studies have investigated the link between drought and cholera outbreaks in isolation in Africa [5, 7], or projected outbreak changes into the future.

Chapter 3 aims to address the research gap of drought-related health outcomes by investigating the implications of drought on cholera outbreak occurrence at a continental scale across Africa, after accounting for important socio-economic factors. These results will provide further understanding of the hypothesis that droughts lead to cholera outbreaks through increased concentrations of infectious bacteria, shed by symptomatic and asymptomatic cholera cases into more limited drinking water sources. Elevated pathogen concentrations (due to less water diluting the bacteria) and risky drinking water behaviours, leads to higher pathogen exposure and an increased probability of ingesting an infective dose. Figure 3.1 shows a schematic to help visualise the hypothesised pathway.

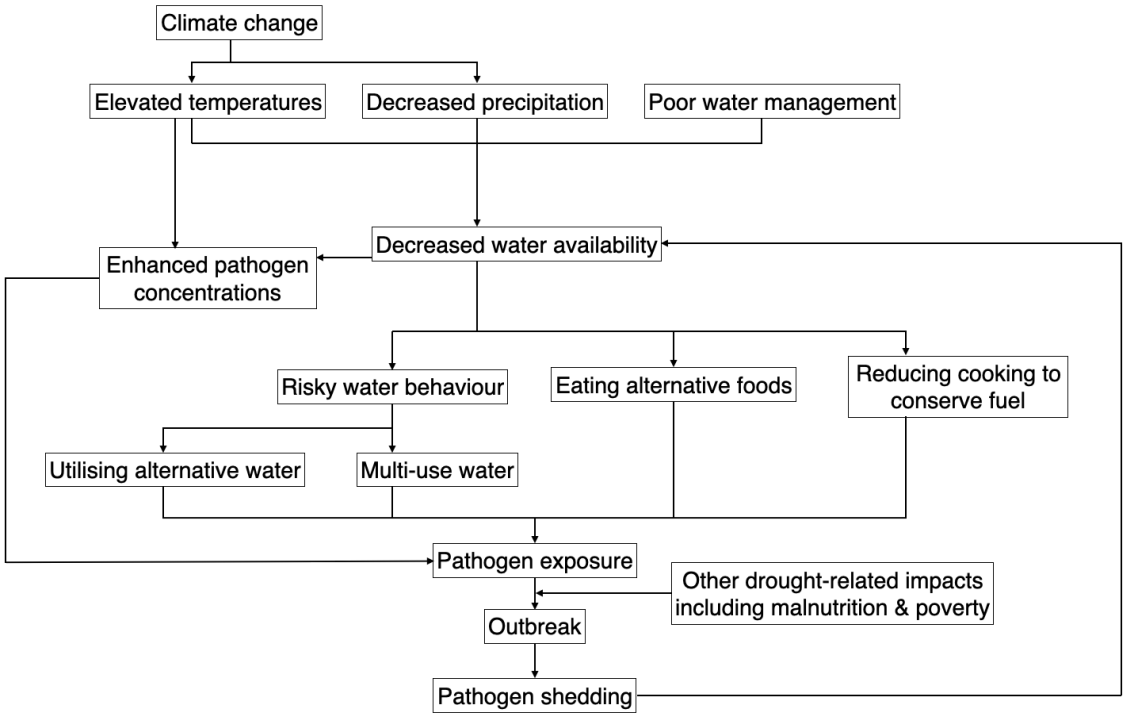


Figure 3.1: Pathways from water shortages to cholera outbreaks: Suggested mechanism through which drought can lead to cholera outbreaks in Africa [5, 8].

In addition, the work aims to evaluate how future changes in drought area and risk due to climate change [13, 14], alongside other development factors, may impact future cholera outbreak occurrence. Several projection scenarios incorporating different greenhouse gas emissions and socio-economic development trajectories will be developed here. 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. The objectives of Chapter 3 are as follows:

1. Identify a potential relationship between cholera and drought at a national level in Africa using statistical modelling.

2. Consider risk factors and variables which may contribute to this relationship.

3. Using the best fit model, project drought into the future with several degrees of global change.

3.2 Methods

3.2.1 Datasets and Study Period

Data were compiled on cholera outbreaks and a range of social and environmental covariates over the period 1970 to 2019. Annual cholera cases were retrieved from the WHO's Global Health Observatory [15], which provides reported annual cholera case for each country, which were confirmed either clinically, epidemiologically, or by laboratory investigation. Cholera case numbers were transformed into a binary outcome to reflect outbreak occurrence for a specific country and year (i.e., set at 0 for no outbreak and 1 for an outbreak in the country and year (>=1 case/death, based on the WHO outbreak definition [16])), which was then used as the outcome variable in the models. Raw case data were not analysed to minimise the effect of unmeasured observations and reporting biases among countries. A sensitivity analysis was used to test alternative methods of dealing with missing values. For years with no outbreak data, either the complete row of data for the year was removed (to avoid the following assumptions), the outcome was set to 0, assuming if cholera cases/deaths occurred within a country then they would have been identified and reported, or 1, to account for known cholera under-reporting:

- Alternative 1 - Removing all the rows with missing cholera data and only assigning 1 or 0 when it was reported
- Alternative 2 - Setting all missing data points to 0
- Alternative 3 - Setting all missing data points to 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 Supplementary Table 3.1). Environmental data were extracted from a variety of sources and included climate (temperature and precipitation) [17], meteorological drought (Palmer Drought Severity Index, PDSI) [18], agricultural drought (soil moisture and potential evapotranspiration (PET) [19, 20] and hydrological drought (runoff and freshwater withdrawal annually and per capita) [21, 22]. The definitions of the types of drought evaluated here are below and are based on those provided by the National Oceanic and Atmospheric Administration's National Weather Service [23]:

- Meteorological drought - the degree of dryness or rainfall deficit and the length of the dry period

- Hydrological drought - the impact of rainfall deficits on the water supply such as stream flow, reservoir and lake levels, and ground water table decline
- Agricultural drought - the impacts on agriculture by factors such as rainfall deficits, soil water deficits, reduced ground water, or reservoir levels needed for irrigation
- Socio-economic drought - considers the impact of drought on socio-economics conditions, which should be captured through the other covariates included in the analysis

Climate data were from the European Centre for Medium-Range Weather Forecast's ERA5 which combines historical observations into global estimates using advanced modelling and data assimilation systems. This provides hourly data for a single grid at a 0.1x0.1 resolution, which is averaged to three grids per month. Shapefiles are then used to lookup what administrative level the grid represents, transforming the data to monthly administrative level 1 data [17, 24].

PDSI was chosen over other metrics of meteorological drought, due to its superior ability to capture long-term change and global warming. PDSI was calculated from gridded (2.5x2.5) monthly surface air temperature, precipitation and meteorological forcing data using the Penman-Monteith PET equation [18]. The equation is below where R_n is the net radiation, G is the soil heat flux, $(e_s - e_a)$ represents the vapour pressure deficit of the air, r_a is the mean air density at constant pressure, c_p is the specific heat of the air, D represents the slope of the saturation vapour pressure temperature relationship, g is the psychrometric constant and r_s and r_a are the surface and aerodynamic resistances.

$$PET_{penman} = \frac{\Delta(R_n - G) + p_a c_p \frac{(e_s - e_a)}{r_a}}{\Delta + \gamma(1 + \frac{r_s}{r_a})}.$$

Where monthly or sub-national data were available, a national yearly mean was calculated. A yearly mean was taken for the monthly climate data to lessen the impact of seasonality and provided a better metric of an environmental extreme, if one occurred. Additionally, using the mean of the raw data, rather than assigning a threshold extreme value, reduced the assumptions needed to do so, which would be different for each country. 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, The Comoros, Mauritius and Seychelles. Environmental data for these countries were derived by taking the mean of their neighbouring countries (assuming their climate would be

similar), 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 [25] and the United Nations Development Programme [26] 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, the designated instrumental period for analysis was 2000 to 2016, to limit omitting missing data and interpolation. If rows with missing values remained, they were removed from the data before model fitting, as missing values created issues with covariate selection. The data fit to the model were to an annual and national granularity, as this best captured the scale of the various data sources (national/annual socioeconomic data), removed seasonality from the environmental data and matched the outcome variable data (cholera outbreak occurrence for a specific country and year). Summary figures of the climate and cholera data over the instrumental period are shown in Supplementary Figure 3.1. Further information about the drought indices (PDSI, soil moisture, PET, runoff and freshwater withdrawal) are in Supplementary Information 3.1.

3.2.2 Model Fitting and Covariate Selection

Generalised linear models (GLM), using maximum likelihood estimation (MLE), were fitted to the dataset describing cholera outbreak occurrence (set to 1 if >= 1 case of cholera was reported for the country and year) for the instrumental period (2000-2016), for all countries in mainland Africa and Madagascar. GLMs are a flexible generalisation of ordinary linear regression, by allowing the linear model component to be related to the response variable via a link function, based on an assumed probability distribution. The linear component (η) is expressed as linear combinations of unknown parameters β. The coefficients of the linear combination are represented as a matrix of independent covariates X, where *water* represents perCapita freshwater withdrawal and *poverty* in poverty headcount at <$1.90/day. $ISO3$, refers to the national ISO3 country code.

$$\eta = X\beta.$$

$$\eta(\text{cholera outbreak occurrence}_{year,ISO3}) = \beta^1 PDSI_{year,ISO3} + \beta^2 Population_{year,ISO3} +$$

$$\beta^3 Temperature_{year,ISO3} + \beta^4 Poverty_{year,ISO3} + \beta^5 Water_{year,ISO3}.$$

The link function provides the relationship between the linear predictor and the mean of the distribution function. The binary outcome variable used for cholera outbreak occurrence, meant that a binomial likelihood with a complementary log-log link function was chosen for all models. Y is the outcome of the dependent variables based on the assumed binomial probability distribution. The probability that $Y = 1$ is the mean of the Y values, expressed as μ and the probability that $Y = 0$ is 1 - μ. Therefore, the link function (g) is expressed in terms of the means.

$$g(\mu) = log(-log(1 - \mu)).$$

Parameter estimation allows an approximation of the unknown parameters (β) of the distribution without seeing all the data. Here, MLE was used and works by finding parameters that yield the maximum of the likelihood function shown above and was chosen as a means of statistical inference due to the assumed probability distribution (binomial), the availability of a relatively complete dataset and prior knowledge of the relationship between drought and cholera being unclear.

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.* [27] and Gaythorpe *et al.* [28]. In summary, univariate linear regression 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. As this was a relatively crude method of excluding covariates, a slightly higher p-value threshold was chosen here, compared to other chapters ($p < 0.05$).

Of the remaining covariates, absolute pairwise correlations were calculated, and highly correlated variables ($r >= 0.75$) were then clustered into groups, to prevent multicolinearity. Co-linearity is where several independant variables in the model are correlated, resulting in the model not fitting independant relationships, causing overfitting. A correlation matrix of the Pearson correlations used in the covariate selection is shown in Supplementary Figure 3.2.

The covariates from each cluster most strongly correlated with the outcome variable was then selected for inclusion in the multivariate generalised linear models. Model fit was evaluated using Bayesian Information Criterion (BIC) and a single best fit model was found using stepAIC. BIC and Akaike Information Criterion (AIC) were used for model selection as they introduce a penalty term for the addition of more parameters from the covariate selection, which can increases the likelihood due to overfitting, not independant relationships. In addition, area under the receiver operator characteristic curve (AUC) was used to quantify model performance. All statistical analyses were carried out in R version 3.6.2 (packages: tidyr, MASS, ggplot2, dplyr, magrittr, corrplot, caret, nlme, MuMIn, car, boot [29, 30, 31, 32, 33]).

3.2.3 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, the degree of similarity between a time series and a lagged version of itself). To investigate the potential influence on the covariate selection and the subsequent best fit model, separate analyses were run including year and ISO3 country code as predictor variables to the outcome variable of cholera outbreak occurrence (set to 1 if >= 1 case of cholera was reported for the country and year). The additional analysis followed the same step-wise covariate selection process and multivariate model fitting approach as described above.

Autocorrelation diagnostics were run on selected spatial and temporal covariates by testing the significance of the linear relationship and the model fit according to AIC, with and without consideration of AR1 (autoregressive model of order 1) autocorrelation. Evidence of autocorrelation in the residuals was also assessed, using autocorrelation function (ACF). ACF measures the correlation coefficient of the residuals, between the time series and it lagged values. The lag is set to $10log_{10}(N/m)$, where N is the number of observations and m the number of series. Additionally, out-of-sample testing was used including leave-one-out (LOO) cross-validation, to assess model performance in terms of predictive accuracy, of both the original (without year/ISO3) and the updated (with year/ISO3) multivariate model. K-fold cross-validation prediction error, in terms of Mean Squared Error (MSE) was used to compare model accuracy.

3.2.4 Projection Scenarios

Three scenarios (S1, S2 and S3) were developed for 2020-2070 (at decadal increments). 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 (RCP4.5) and strong progress towards the SDGs (meeting the 2030 targets), S2 represents an intermediate scenario with median values between S1 and S3, and S3 a "worst-case" scenario with slower progress towards emissions reductions (RCP8.5) and the SDGs (targets not achieved until much later in the century).

Detailed descriptions and justifications of the projected changes for each variable are provided in full in Supplementary Information 3.2. Briefly, projected temperature data were taken from WorldClim [34] which are Coupled Model Intercomparison Project 6 (CMIP6) downscaled future climate projections, processed for nine global climate models using three Representative Concentration Pathways (RCP). RCP4.5, 6.0 and 8.5 were used for scenarios S1, S2 and S3, respectively. The numbers in the RCPs (4.5, 6.0 and 8.5), refers to the radiative forcing (change of energy flux in the atmosphere) that would occur in the scenario. RCP4.5 requires emissions to peak in 2040, RCP6.0 peaks around 2080 and RCP8.5 assumes emission will continue to rise throughout the century. These three scenarios resulted in mean global warming by 2081-2100 of 1.8°C, 2.2°C and 3.7°C, respectively [35]. The projections provided values for 2050 and 2070 and the instrumental period average (2000-2016) was used for the 2020-2040 values, to account for interannual climate variability. Supplementary Figure 3.3 summaries the projected climate data for each pathway and year.

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 and disagreement [13, 18], 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 duration remained constant despite periods of extreme dryness, over a range of time scales. The PDSI dataset used here for both the full data range (1850-2016) and the instrumental period (2000-2016) also remained fairly constant and the data accurately captured past drought as its changes tracked with soil moisture, a good index of drought (see Supplementary Figure 3.4 and 3.5) [40].

Given these disagreements and the likelihood that calculating drought at this scale (national) would result in loss of important spatial heterogeneity. Future drought conditions were estimated for each scenario at the assumption that drought would remain at a similar level (in the "best-case" scenario) or continue in the direction of its historical trend (as a "worst-case" scenario). Therefore for S1, no change relative to a current "baseline" was used 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"), univariate linear regression models (drought $\sim$ year) were fit to the full historical data for each country (1850-2016). The models were then used to extrapolated the trend, by using the coefficients as a yearly multiplier (up until the extreme values of +4 for extreme wetness and -4 for extreme dryness). S2 was an intermediate value between S1 and S3. The method used here followed other drought projection studies [41], and the results of the univariate drought models are available in Supplementary Table 3.2.

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. The selected covariates in the best fit model were maintained at the 2016 levels and only drought was altered. The selected six alternative drought scenarios were created to account for a wide range of drought changes that could occur in the future. Changes were only made up to the PDSI limits (+4 to -4) and if the scale limit was reached then this was taken as the value. The six alternative drought scenarios for the sensitivity analysis are shown below:

1. Drought-average – Average national drought for 2000-2016
2. Drought-S1 – 2016 value + 0.5
3. Drought-S2 – 2016 value - 0.5
4. Drought-S3 – 2016 value + 1
5. Drought-S4 – 2016 value - 1
6. Drought-S5 – 2016 value + 2
7. Drought-S6 – 2016 value - 2

Poverty changes were based on SDG1.1 and 1.2 [42], despite the limitations of the SDGs (e.g., ambiguous terms), they are a globally recognised standard for sustainable development and also feature

heavily within the GTFCC cholera targets. 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 threshold 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 (substantially increase water-use efficiency across all sectors and ensure sustainable withdrawals) and either increased or decreased based on each country's historical freshwater withdrawal relative to available water resources, taken from the same data source used in the model [21]. The data visualisation of freshwater security for each country is plotted in Figure 3.2 and countries categorised as high freshwater withdraw countries (HWC) were those who withdraw >357m^3/year, which were countries in the top 10% of freshwater withdrawal in Africa (water withdrawal includes all domestic, industrial and agricultural uses). The threshold was set on withdrawal only (Figure 3.2 x-axis), as resource is difficult to alter beyond the geography of the country, whereas withdrawal is the more changeable component of water security. 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 sustainable freshwater availability was increased by 20% (10% for S2 and 5% for S3) by the middle of the projection period (2050), for countries with sufficient resources.

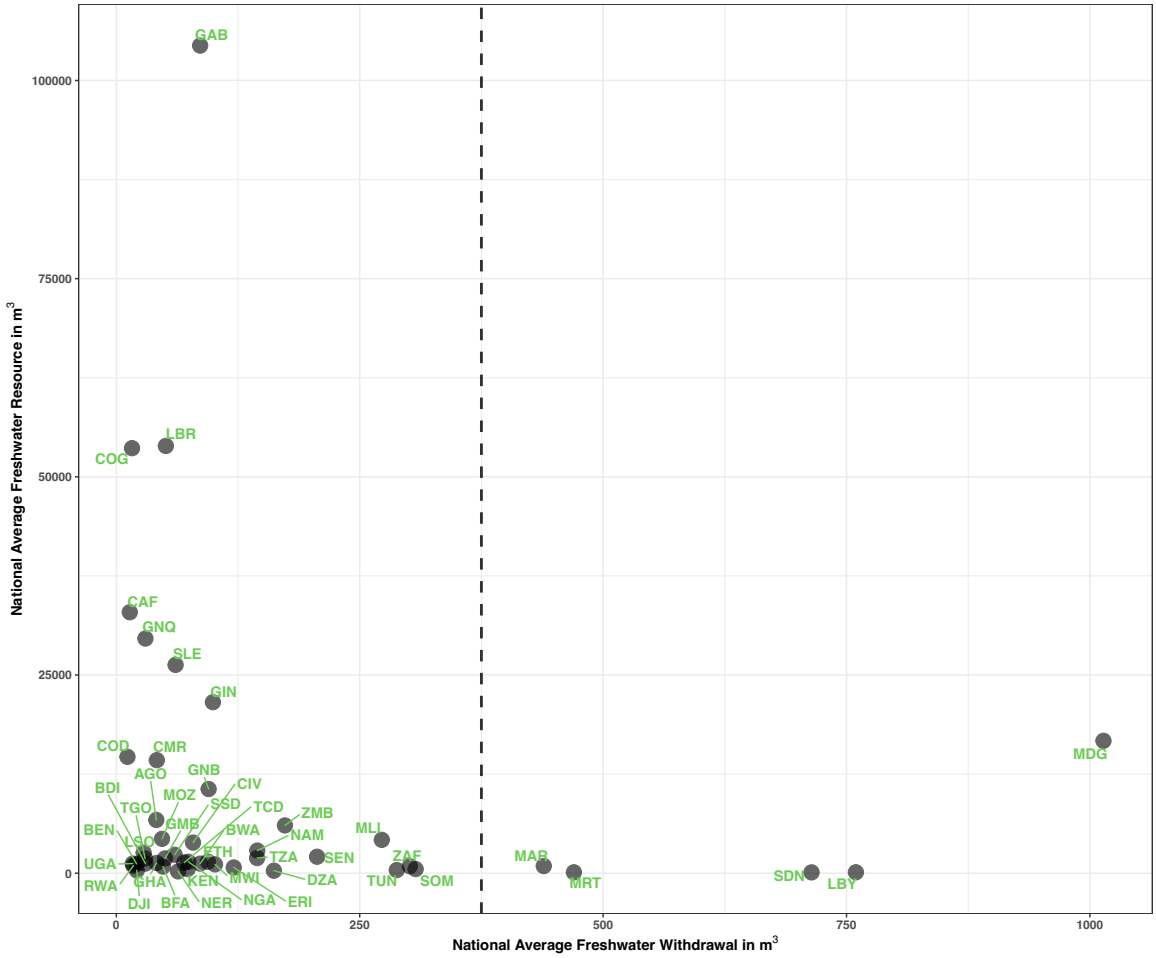


Figure 3.2: Historical water security in Africa as freshwater withdrawal against freshwater resource. Both are taken as national per capita averages for the full dataset (1985-2014) in m^3. The dashed line represents the cutoff for countries to be categorised as high withdraw ($375m^3$), and countries are represented by their ISO3 code [21].

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, a single medium population size was used to isolate the effects of the other environmental and socio-economic covariates. Additionally, the larger population size would likely result in greater cholera transmission (and higher projected outbreak occurrence here) due to the larger number of susceptible people, not necessarily because of changes in important risk factors. The three scenarios are summarised in Table 3.1.

Table 3.1: Cholera projection scenarios for 2020-2070 at decadal intervals for Scenario 1 (S1), Scenario 2 (S2) and Scenario 3 (S3). HWC = high water withdraw countries including Madagascar (MDG), Libya (LBY), Sudan (SDN), Mauritania (MRT) and Morocco (MAR). RCP = Representative Concentration Pathway.

Scenario	Year	Drought	Temperature	Poverty	Water withdrawal
S1	2020	2000-2016 average	2000-2016 average	2016	2016
S1	2030	2000-2016 average	2000-2016 average	Reduce 2016 by 50%	2016
S1	2040	2000-2016 average	2000-2016 average	Reduce 2016 by 50%	2016
S1	2050	2000-2016 average	RCP4.5 2050	Medium value between 2030 & 2070	20% increase and 20% decrease for HWC
S1	2060	2000-2016 average	RCP4.5 2050	Medium value between 2030 and 2070	20% increase and 20% decrease for HWC
S1	2070	2000-2016 average	RCP4.5 2070	Poverty elimination (0%)	20% increase and 20% decrease for HWC
S2	2020	Median value between S1 and S2	2000-2016 average	2016	2016
S2	2030	Median value between S1 and S2	2000-2016 average	2016	2016
S2	2040	Median value between S1 and S2	2000-2016 average	2016	2016
S2	2050	Median value between S1 and S2	RCP6.0 2050	Reduce 2016 by 50%	10% increase and 10% decrease for HWC
S2	2060	Median value between S1 and S2	RCP6.0 2050	Medium value between 2050 and 2070	10% increase and 10% decrease for HWC
S2	2070	Median value between S1 and S2	RCP6.0 2070	Poverty elimination (0%)	10% increase and 10% decrease for HWC
S3	2020	((Coefficient*4) + 2016 value)	2000-2016 average	2016	2016
S3	2030	((Coefficient*10) + 2020 value)	2000-2016 average	2016	2016
S3	2040	((Coefficient*10) + 2030 value)	2000-2016 average	2016	2016
S3	2050	((Coefficient*10) + 2040 value)	RCP8.5 2050	2016	5% increase and 5% decrease for HWC
S3	2060	((Coefficient*10) + 2050 value)	RCP8.5 2050	Medium value between 2050 and 2070	5% increase and 5% decrease for HWC
S3	2070	((Coefficient*10) + 2060 value)	RCP8.5 2070	Reduce 2016 value by 50%	5% increase and 5% decrease for HWC

3.3 Results

3.3.1 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 3.2. 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.

Table 3.2: Univariate model outputs and goodness-of-fit measures for the tested covariates against cholera outbreak occurrence, including p-values, coefficients, BIC and AUC. * represents p <0.1.

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	784.57	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 hand washing facilities (% population)	0.018*	0.189	766.43	0.5882
Percentage living in informal settlement (% urban population)	0.013*	-0.467	778.73	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

For the sensitivity analysis investigating the effects of assigning the outcome variable to 0 if cholera was not reported, the original assumption was taken as the best option here. The results found that the alternative assumptions removed too much data for the covariate selection, which resulted in large numbers of variables being removed in the univariate analysis as they did not meet the threshold of p <0.1.

3.3.2 Output from the Best Fit Model

After covariate selection, five covariates were retained in the best fit model. These include total population on a logarithmic scale, mean meteorological drought (in PDSI), average temperature (in °C), poverty headcount at <$1.90/day and per capita freshwater withdrawal (in m^3/person/year). Goodness of fit measures and outputs for the best fit model are shown below in Table 3.3. 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 Figure 3.3. Marginal effects indicate how a dependent variable changes when a specific independent variable

changes, if other covariates are held constant (at the mean).

Table 3.3: Output and goodness of fit measures for the best fit model.

	Coefficient		**Exp(Coefficient)**		**p-value**
Mean national drought (PDSI)	-0.0927813		0.911392813		0.051172
Population, total (log)	1.3125412		3.71560365		$2.85\text{x}10^{-13}$
Average temperature (°C)	0.0927423		1.097178975		0.000113
Poverty headcount (at <$1.90/day)	0.0327487		1.03329089		$4.23\text{x}10^{-16}$
Per capita freshwater withdrawal (m^3/person/year)	-0.0024225		0.997580455		$5.43\text{x}10^{-7}$
Residuals	**Min**	**1Q**	**Median**	**3Q**	**Max**
	-2.0286	-0.7974	0.4069	0.8601	2.2564
R^2: 0.276254					
AUC: 0.7784					

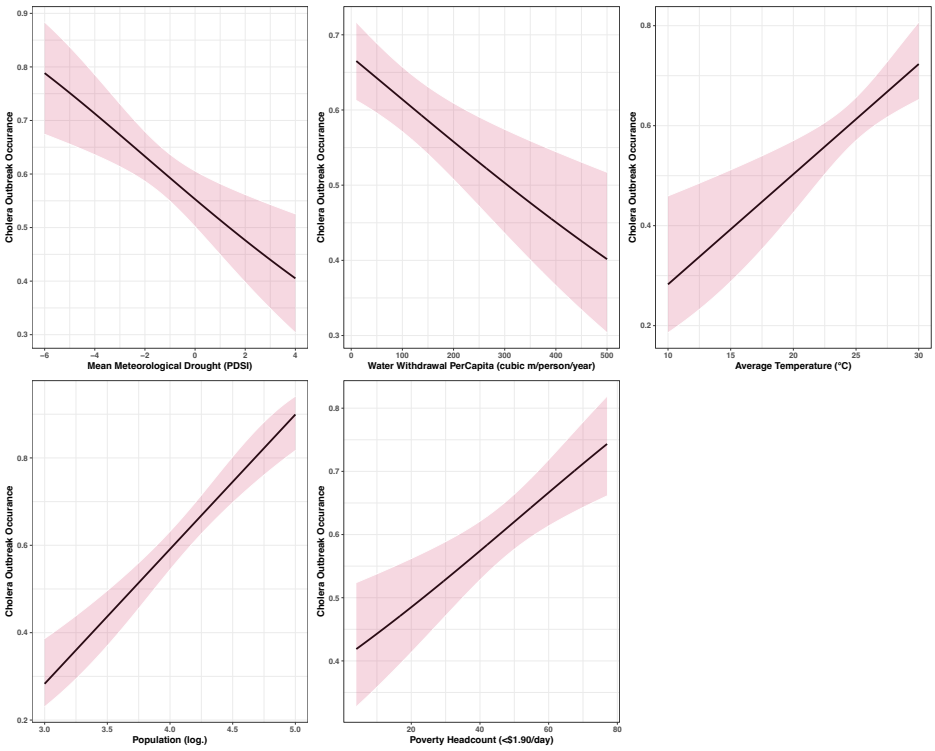


Figure 3.3: Marginal effect plots for the five selected covariates for the best fit model, showing cholera outbreak occurrence. The other covariates are held at the mean, with a 95% prediction confidence interval.

3.3.3 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 as the original model and additionally the percentage of the population with basic hand

washing facilities and Human Development Index (HDI). The linear relationship between year and cholera was visualised using loess curves for each country and are shown in Supplementary Figure 3.6.

The diagnostic results for the out-of-sample validation using AIC are as follows. The time series for lm(Outbreak $\sim$ Year) showed year to be significant at p <0.05 and the model performance for lm(Outbreak $\sim$ Year) showed some temporal autocorrelation (Figure 3.4). The re-estimated linear trend and model performance to account for autocorrelation gls(Outbreak $\sim$ Year, correlation = corAR1(form = $\sim$Year)) showed Year to no longer be significant (p <0.05), but the intercept remained the same. When the two linear models were compared, with or without correction for autocorrelation, there was no appreciable difference (difference of 1 in the AIC values) (Figure 3.5).

In a model which shows no signs of autocorrelation, the points in a Turkey-Anscombe plot and Scale location plot are not correlated and equally spread around a mean of 0. The Normal Q-Q plot has a linear trend line with no outliers and the points are normally distributed. A Leverage plot shows any points that are particularly influential (leverage) and therefore may be biasing the results, ideally points fall inside the Cook's line and confidence interval, showing they all have low leverage. ACF is measured as the correlation coefficient of the residuals between the observed and lagged values, stationary time series (not temporally influenced) have an exponential decrease in ACF and all points are within the confidence interval.

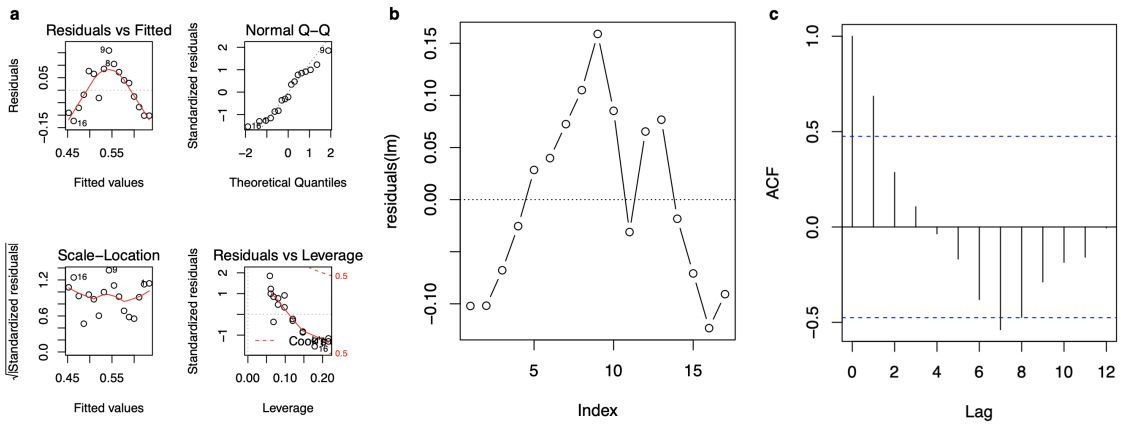


Figure 3.4: Autocorrelation diagnostic results for the model without autocorrelation accounted for. **a**, model performance diagnostics including a Turkey-Anscombe plot (top left), Normal Q-Q plot (top right), Scale location (bottom left) and a Leverage plot (bottom right), **b**, time series of the residuals and **c** evidence of autocorrelation in the residuals using ACF (autocorrelation function) with 95% confidence intervals (blue-dashed line).

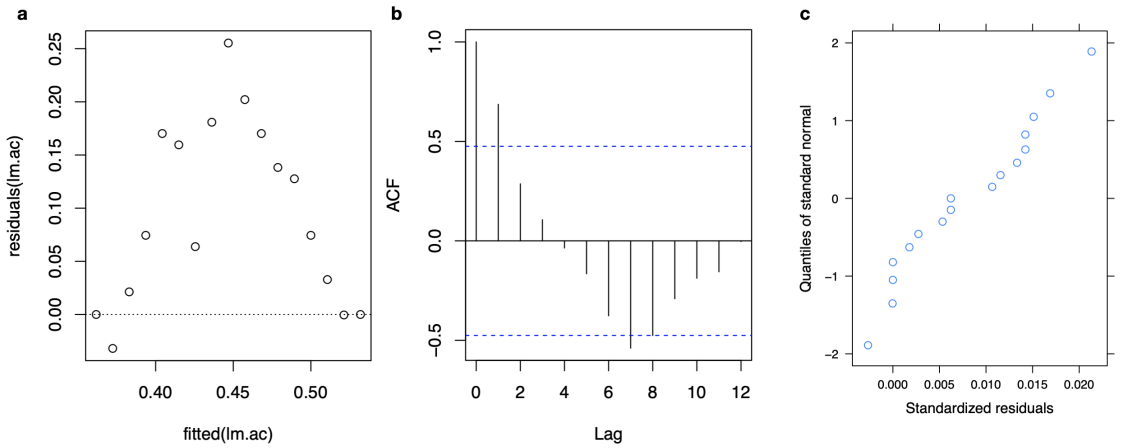


Figure 3.5: Autocorrelation diagnostic results for the model with autocorrelation accounted for. **a**, time series of the residuals, **b** evidence of autocorrelation in the residuals using ACF (autocorrelation function) with 95% confidence intervals (blue-dashed line) and **c**, a Normal Q-Q plot.

The two multivariate GLM models (Outbreak ∼ drought + temperature + poverty + population + water withdrawal and the original model + year + HDI + hand washing), selected by the covariate selection process were then run through LOO cross-validation to assess for an appreciable model performance difference. Model cross-validation prediction error, in terms of MSE, differed by 0.02. Therefore, the model selected without the inclusion of year and country code in the selection process was used as the best fit model.

Using the model without the inclusion of year was selected as cholera outbreak occurrence appears conditionally independent of year given the other covariates in the model. 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 in detection of all-pathogen outbreaks from the mid 1990s onwards, especially in low- and middle-income countries, improving countries' capacity for detection, response and therefore reporting [44, 45].

3.3.4 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 Figure 3.6. 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.

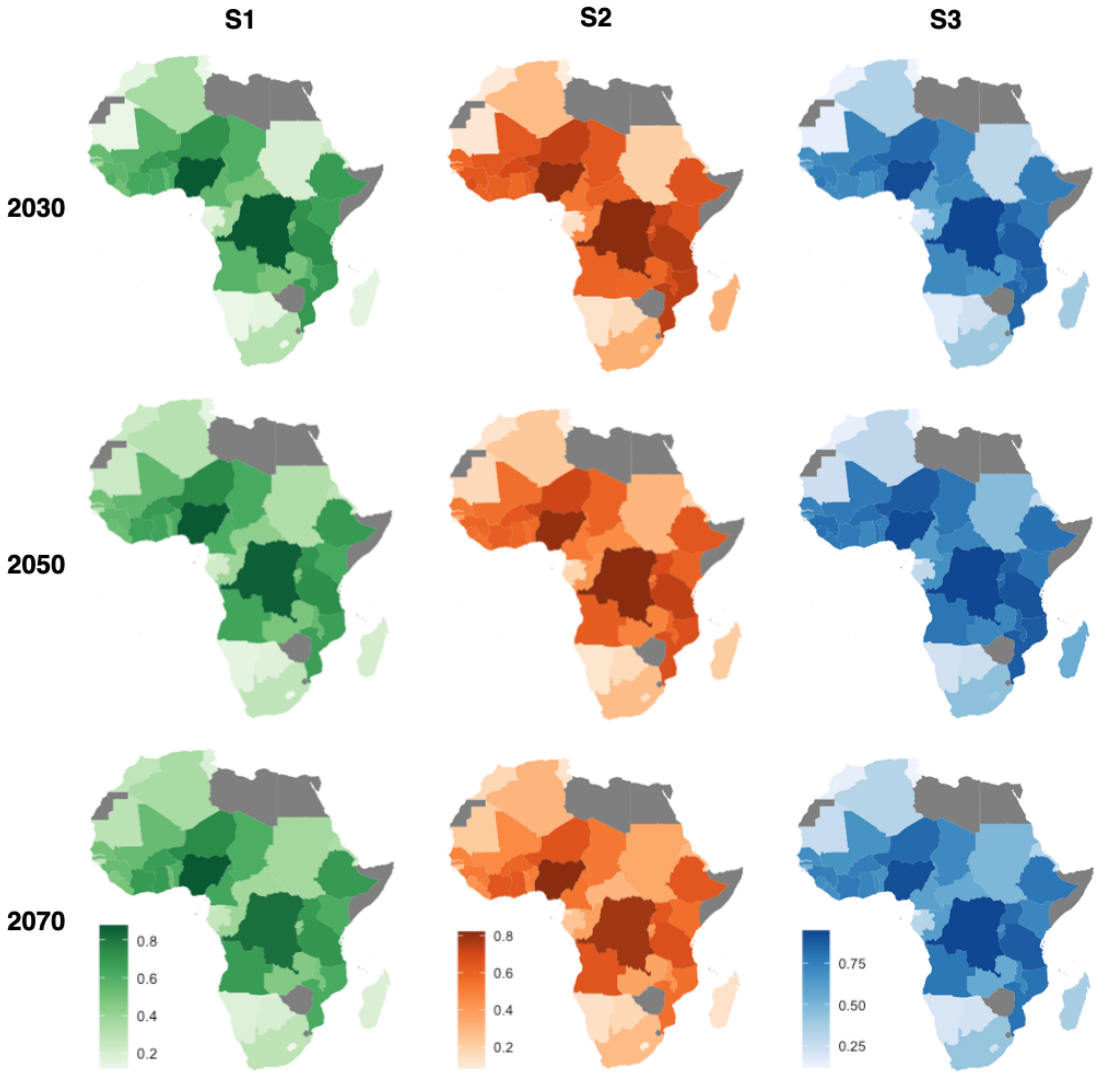


Figure 3.6: Projected cholera outbreak occurrence (0-1) for the three scenarios (S1 - green, S2 - orange and S3 - blue) in 2030, 2050 and 2070. Grey represents countries where covariate data were missing (Botswana, Zimbabwe, Somalia, Egypt, Eswatini, Western Sahara, Algeria, Libya and Eritrea) and therefore could not be included in the model.

Figure 3.7 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.

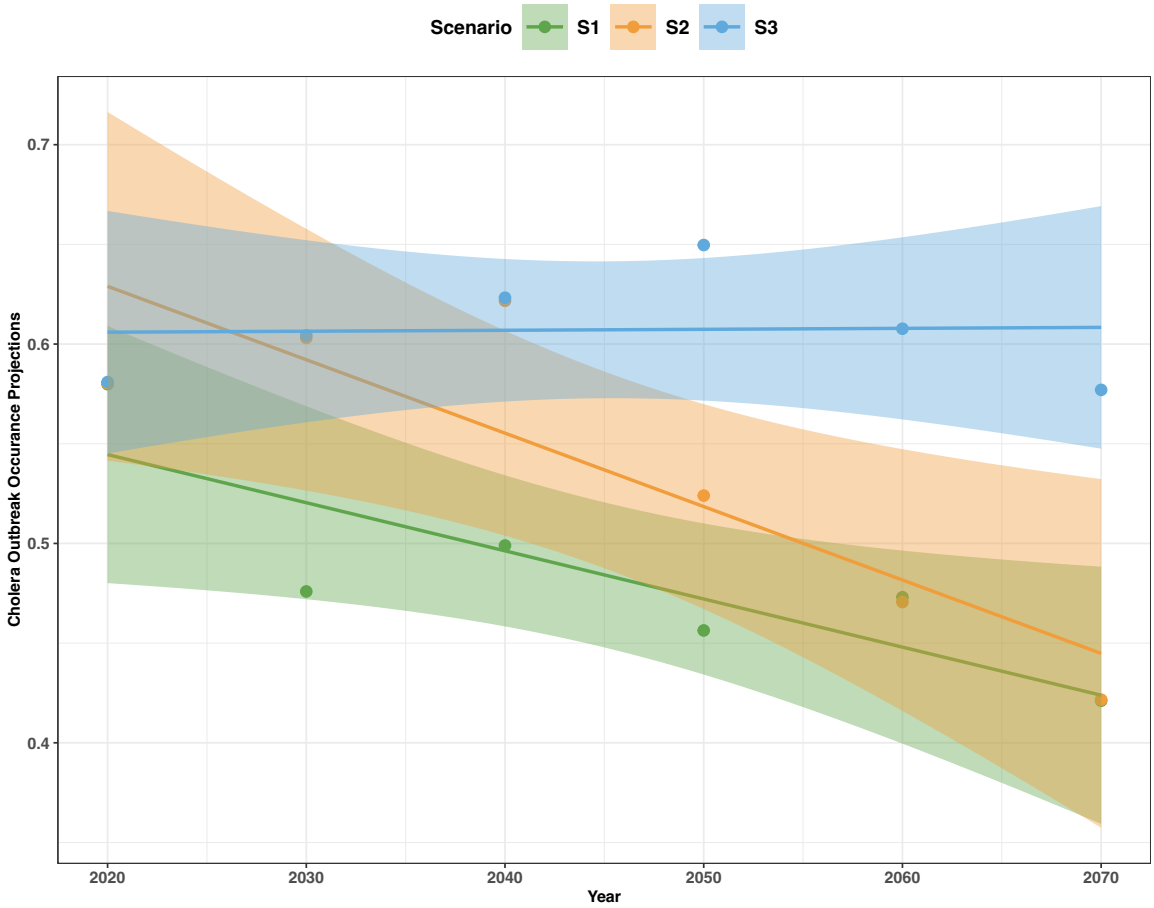


Figure 3.7: Mean continental cholera outbreak occurrence for the projected period (2020-2070) using the three scenario datasets.

The drought sensitivity analysis showed modest changes through the six alternative drought scenarios (Table 3.4), with more negative values of PDSI seeing higher cholera outbreak occurrence (Supplementary Figure 3.7). Despite this, these changes were not excessive with a 0.06 average increase in continental cholera outbreak occurrence from the 2000-2016 average 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.

Table 3.4: Predicted continental cholera outbreak occurrence using the best fit model and different drought sensitivity scenarios. National averages are shown in Supplementary Figure 3.7.

Sensitivity analysis	Continental averages
Drought-average	0.540106952
Drought-S1	0.561436986
Drought-S2	0.576812737
Drought-S3	0.553412242
Drought-S4	0.584072274
Drought-S5	0.53701047
Drought-S6	0.597395878

3.4 Discussion

Cholera has well established environmental [46, 47, 48] and socio-economic links [49, 50, 51], such as poverty, poor WASH conditions and changes in temperature and precipitation (and therefore links to climatic patterns e.g., Intertropical Convergence Zone and El Niño Southern Oscillation). Here, multiple environmental variables were important covariates in the model. Meteorological drought (measured in 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, it is largely understudied [5, 6, 7]. Chapter 3 modelled drought in isolation allowing a more in-depth investigation of its impacts, and tested whether drought is likely to influence cholera outbreaks under scenarios of climate change and socio-economic development (attainment of the SDGs). According to these results, drought will continue to be an important hazard for cholera outbreaks in the future, but gains in sustainable development (reduction of poverty, increased water security) may offset cholera risk.

3.4.1 Cholera-Environment Links

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 growth and biofilm formation at elevated temperatures (20-45°C, depending on salinity [52, 53]) [54]. This may also represent an independent effect of temperature from drought and why both variables were selected in the model. For

example, a 1°C rise in temperature was associated with a 2-fold increase in cholera cases in Zanzibar [48]. 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 (meteorological drought in PDSI and hydrological drought in 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. Furthermore, there are multiple drought metrics, which all have different limitations and will be assessed in a later chapter (Chapter 5).

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 dangerous water practices, in terms of health. 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. For example, water should be stored and not overused for industry and agriculture such as irrigation, dam building should be avoided and surface water run off reduced as much as possible [55].

3.4.2 Cholera-Socio-economic Links

Cholera is a disease of inequity and poverty and is often seen in combination with poor WASH facilities [50, 56]. 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 socio-economic covariates are more appropriate predictors [57]. For example, droughts have been known to impact the United States and Europe [58, 59], 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 socio-economic 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 are 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, nearly 90% of diarrhoeal disease has been attributed to sub-optimal WASH [60]. 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.

3.4.3 Cholera Projections to 2070

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 heterogeneous 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. Despite the challenges, understanding future risks is vitally important for climate change mitigation and adaptation.

3.4.4 Limitations

Environmental and socio-economic data were missing for several countries and years, meaning that data had to be averaged or omitted for model fitting. 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 Chapter 3, such as seasonal changes in cholera and the presence of waterbodies within countries facilitating transmission [2]. 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

[61], possibly due to a spectrum of transmission dynamics, lulls in cases causing decreased awareness [62] and disincentives to report outbreaks due to implications for tourism and trade [63].

Considering the problems in cholera under-reporting, issues may have arisen from assigning the outcome variable to zero for missing years, as this could have led to the under-representation of cholera outbreaks. However, given the results of the sensitivity analysis for this assumption, this is the best interpretation of missing values, as removing values created issues when trying to select covariates from small numbers of data points.

GLMs assume a monotonic relationship and therefore non-linear effects of several covariates might not be captured and these non-linear effects are evaluated in later chapters (Chapter 4 and 5). An example of these non-linearities are seen in Supplementary Figure 3.4 and 3.5 and Supplementary Table 3.2, as some countries fit the linear trend better than others. Additionally, the covariate selection process used here may have resulted in important covariates (in terms of their relationship with cholera) being lost. Removing covariates at a p threshold of 0.1 (in a univariate linear regression) is a relatively crude method of reducing a pool of potential variables and in future analysis (Chapter 5), the performance of the covariates will be first be assessed, before all of them will be considered for model selection using the correlations and clustering.

As with any projections and the creation of scenarios, uncertainty can be high, as assumption must often be made e.g., increasing freshwater availability would result in greater access. Therefore, theoretical, methodological and computational challenges in projecting future climate change and its consequences can occur. There are also the realities of meeting or exceeding the SDGs, which were used to guide the scenarios here, when no clear pathway to success has been defined. Additionally, a "worst-case" scenario where poverty and water withdrawal regresses was not considered and will therefore be revisited in a later chapter (Chapter 6). Finally, the possibility of new *V. cholerae* strains being introduced and changing both natural- and vaccine-derived immunity could also complicate cholera eradication efforts. Nevertheless, with decreasing poverty and the expansion of freshwater availability, even the effect of new cholera cases and strain introductions could be reduced.

3.4.5 Conclusion

In conclusion, the relationships between temperature, drought and water withdrawal add further evidence to the original hypothesis that hotter and drier conditions and a lack of available freshwater, increases cholera outbreak occurrence, potentially through risky water behaviours. Future qualitative work on drought and cholera would be needed to understand the complexities of these potentially risky 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 and future groundwater studies would be needed to confirm this hypothesis. Socio-economic variables were highly significant in the best fit model, showing the impact of vulnerability in times of water shortage and the need to lift people out of poverty to improve health and reduce mortality.

Chapter 3 offers additional insight into how climate change may yield health impacts in the future and the importance of socio-economic development and emissions reductions to offset this risk. Chapters 4-6 aim to build on these findings, to understand some of these relationships on a finer spatial scale. High burden countries such as the DRC and Nigeria saw only marginal changes in the cholera projections, even with high levels of sustainable development and stringent emissions reductions. Both Nigeria and the DRC have similar climates and socio-economic vulnerabilities [64, 65] and therefore it would be expected that the more optimistic scenarios would reduce cholera risk. A potential explanation for why cholera risk was not reduced in these countries, may be other drivers of cholera that were not accounted for in the model (residual confounding), such as conflicts. Both Nigeria and the DRC have active conflicts, ranking 12th and 5th, respectively, on the Fragile States Index [66] and many conflict-related outbreaks were found in the review. Chapter 4 aims to understand the impact of conflict on cholera in Nigeria and the DRC and to use a novel modelling approach to understand how cholera risk could be reduced.

References

[1] M. Ali et al. “Updated global burden of cholera in endemic countries”. en. In: *PLoS Neglect. Trop. Dis.* 9 (2015), p. 0003832.

[2] D.B. Nkoko et al. “Dynamics of cholera outbreaks in Great Lakes region of Africa, 1978–2008”. en. In: *Emerg. Infect. Dis.* 17.11 (2011).

[3] S.E. Carter. “What questions we should be asking about COVID-19 in humanitarian settings: perspectives from the social sciences analysis cell in the Democratic Republic of the Congo”. en. In: *BMJ Glob. Health* 5 (2020), p. 003607.

[4] S.S. Musa. “Dual tension as Nigeria battles cholera during the COVID-19 pandemic”. en. In: *Clin. Epidemiology Glob. Health* 12 (2021).

[5] R.V. Tauxe et al. “Epidemic cholera in Mali: high mortality and multiple routes of transmission in a famine area”. en. In: *Epidemiol. Infect.* 100.2 (1988), pp. 279–289.

[6] A.F. Abdussalam. “Modelling the Climatic Drivers of Cholera Dynamics in Northern Nigeria Using Generalised Additive Models”. en. In: *Int. J. Geog. Environ. Manag.* 2.1 (2016), pp. 84–97.

[7] A. Rieckmann et al. “Exploring droughts and floods and their association with cholera outbreaks in sub-Saharan Africa: a register-based ecological study from 1990 to 2010”. en. In: *Am. J. Trop. Med. Hyg.* 98.5 (2018), pp. 1269–1274.

[8] O. Mark et al. “A new methodology for modelling of health risk from urban flooding exemplified by cholera–case Dhaka, Bangladesh”. en. In: *J. Flood Risk. Manag.* 11:S28-S42 (2018).

[9] A. Jutla, R. Khan, and R. Colwell. “Natural disasters and cholera outbreaks: current understanding and future outlook”. nl. In: *Curr. Environ. Health. Rep.* 4.1 (2017), pp. 99–107.

[10] C.T. Codeço. “Endemic and epidemic dynamics of cholera: the role of the aquatic reservoir”. fr. In: *BMC Infect. Dis.* 1.1 (2001).

[11] UNICEF West and Central Africa. *Africa to drastically accelerate progress on water, sanitation and hygiene – report.* 2022. URL: https://www.unicef.org/senegal/en/press-releases/africa-drastically-accelerate-progress-water-sanitation-and-hygiene-report.

[12] O. Hoegh-Guldberg et al. "Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty". en. In: World Meteorological Organization, 2018. Chap. Impacts of 1.5°C Global Warming on Natural and Human Systems. URL: https://www.ipcc.ch/sr15.

[13] G.G. Haile et al. "Projected impacts of climate change on drought patterns over East Africa". en. In: *Earths Future* 8.7 (2020).

[14] A. Ahmadalipour et al. "Future drought risk in Africa: Integrating vulnerability, climate change, and population growth". en. In: *Sci. Total Environ.* 662 (2019), pp. 672–686.

[15] World Health Organization. *The Global Health Observatory*. en. 2020. URL: https://www.who.int/data/gho.

[16] World Health Organization. *Cholera*. en. 2020. URL: https://www.who.int/news-room/fact-sheets/detail/cholera.

[17] E.C.M.W.F. *ERA5*. it. 2020. URL: https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5.

[18] N.C.A.R. *Dai Global Palmer Drought Severity Index (PDSI)*. en. 2020. URL: https://rda.ucar.edu/datasets/ds299.0/index.html#!sfol-wl-/data/ds299.0.

[19] Copernicus. *Soil moisture gridded data from 1978 to present*. en. 2018. URL: https://cds.climate.copernicus.eu/cdsapp#!/dataset/satellite-soil-moisture?tab=form.

[20] N.C.A.R. *CRU TS GRIDDED PRECIPITATION AND OTHER METEOROLOGICAL VARIABLES SINCE 1901*. en. 2015. URL: https://climatedataguide.ucar.edu/climate-data/cru-ts-gridded-precipitation-and-other-meteorological-variables-1901.

[21] H. Ritchie. *Water Use and Stress*. en. 2017. URL: https://ourworldindata.org/water-use-stress.

[22] U.N.H./G.R.D.C. *UNH/GRCD Composite Runoff Fields V1.0*. en. 2000. URL: https://www.compositerunoff.sr.unh.edu.

[23] National Weather Service. *Drought Types*. 2014. URL: https://www.weather.gov/safety/drought-types.

[24] GADM. *GADM maps and data*. 2022. URL: https://www.gadm.org.

[25] WorldBank. *World Bank Open Data*. en. 2017. URL: https://data.worldbank.org.

[26] United Nations Development Programme. *Human Development Data (1990-2018)*. en. 2018. URL: http://hdr.undp.org/en/data#.

[27] T. Garske et al. "Yellow fever in Africa: estimating the burden of disease and impact of mass vaccination from outbreak and serological data". en. In: *PLoS Med.* 11.5 (2014).

[28] K.A. Gaythorpe et al. "The global burden of yellow fever". de. In: *eLife* 10:e64670 (2021).

[29] M. Kuhn. *caret: Classification and Regression Training*. en. 2021. URL: https://CRAN.R-project.org/package=caret.

[30] J. Pinheiro et al. *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-148. 2020. URL: https://CRAN.R-project.org/package=nlme.

[31] J. Fox and S. Weisberg. *car: An R Companion to Applied Regression*. Third. Thousand Oaks CA: Sage, 2019. URL: https://socialsciences.mcmaster.ca/jfox/Books/Companion/.

[32] A. Canty and B. Ripley. *boot: Bootstrap R (S-Plus) Functions*. R package version 1.3-25. 2020. URL: https://cran.r-project.org/web/packages/boot/boot.pdf.

[33] K. Bartoń. *MuMIn: Multi-Model Inference*. R package version 1.43.17. 2022. URL: https://cran.r-project.org/web/packages/MuMIn/MuMIn.pdf.

[34] WorldClim. *Future climate data*. it. 2020. URL: https://worldclim.org/data/cmip6/cmip6climate.html.

[35] Intergovernmental Panel on Climate Change. *Climate Change 2014 Synthesis Report Fifth Assessment Report Future Climate Changes, Risks and Impacts*. 2014. URL: https://www.ipcc.ch/site/assets/uploads/2018/05/SYR_AR5_FINAL_full_wcover.pdf.

[36] Y. Yang et al. "Little change in Palmer Drought Severity Index across global land under warming in climate projections". en. In: *Hydrol. Earth Syst. Sci. Discuss.* (2020).

[37] J. Sheffield, E.F. Wood, and M.L. Roderick. “Little change in global drought over the past 60 years”. en. In: *Nature* 491.7424 (2012), pp. 435–438.

[38] R. Touchan et al. “Long term context for recent drought in northwestern Africa”. en. In: *Geophys. Res. Lett* 35.13 (2008).

[39] D. Verschuren, K.R. Laird, and B.F. Cumming. “Rainfall and drought in equatorial east Africa during the past 1,100 years”. en. In: *Nature* 403.6768 (2000), pp. 410–414.

[40] Z. Tian-Jun and H. Tao. “Projected changes of palmer drought severity index under an RCP8.5 scenario”. en. In: *AOSL* 6.5 (2013), pp. 273–278.

[41] A. Ahmadalipour and H. Moradkhani. “Multi-dimensional assessment of drought vulnerability in Africa: 1960–2100”. en. In: *Sci. Total Environ.* 644 (2018), pp. 520–535.

[42] United Nations. *The 17 Goals*. en. https://sdgs.un.org/goals. 2015.

[43] United Nations Department for Economic and Social Affairs. *Population Dynamics. World Population Prospectus*. en. 2019. URL: https://population.un.org/wpp/.

[44] S.A. Kluberg et al. “Global capacity for emerging infectious disease detection, 1996–2014”. en. In: *Emerg. Infect. Dis.* 22.10 (2016).

[45] R. Ratnayake et al. “Early detection of cholera epidemics to support control in fragile states: estimation of delays and potential epidemic sizes”. en. In: *BMC Med.* 18.1 (2020), pp. 1–16.

[46] G.C. De Magny et al. “Cholera outbreak in Senegal in 2005: was climate a factor?” en. In: *PLoS ONE* 7.8 (2012).

[47] S. Rebaudet et al. “Environmental determinants of cholera outbreaks in inland Africa: a systematic review of main transmission foci and propagation routes”. en. In: *J. Infect. Dis.* 208.Suppl 1 (2013).

[48] R. Reyburn et al. “Climate variability and the outbreaks of cholera in Zanzibar, East Africa: a time series analysis”. en. In: *Am. J. Trop. Med. Hyg.* 84.6 (2011), pp. 862–869.

[49] D. Olago et al. “Climatic, socio-economic, and health factors affecting human vulnerability to cholera in the Lake Victoria basin, East Africa”. en. In: *Ambio* 36.4 (2007), pp. 350–358.

[50] A. Talavera and E.M. Perez. “Is cholera disease associated with poverty?” en. In: *J. Infect. Dev. Ctries.* 3.06 (2009), pp. 408–411.

[51] L. Mari et al. “Modelling cholera epidemics: the role of waterways, human mobility and sanitation”. en. In: *J. R. Soc. Interface.* 9.67 (2012), pp. 376–388.

[52] R.M. Martinez, C.J. Megli, and R.K. Taylor. “Growth and laboratory maintenance of Vibrio cholerae”. In: *Curr Protoc Microbiol.* 17.1 (2010), 6A–1.

[53] F.L. Singleton et al. “Effects of temperature and salinity on Vibrio cholerae growth”. In: *Appl. Environ. Microbiol.* 44.5 (1982), pp. 1047–1058.

[54] R.J. Borroto. “Ecology of Vibrio cholerae serogroup 01 in aquatic environments”. fr. In: *Pan. Am. J.* 2 (1997), pp. 328–333.

[55] F. Pearce. *When the Rivers Run Dry: The Global Water Crisis and How to Solve It.* Portobello Books, 2018.

[56] K. Penrose et al. “Informal urban settlements and cholera risk in Dar es Salaam, Tanzania”. et. In: *PLoS. Neglect. Trop. Dis.* 4.3 (2010).

[57] F.X. Weill et al. “Genomic history of the seventh pandemic of cholera in Africa”. en. In: *Science* 358.6364 (2017), pp. 785–789.

[58] D.R. Easterling et al. “Effects of temperature and precipitation trends on US drought”. en. In: *Geophys. Res. Lett.* 34.20 (2007).

[59] V. Moravec et al. “A 250-year European drought inventory derived from ensemble hydrologic modelling”. en. In: *Geophys. Res. Lett.* 46.11 (2019), pp. 5909–5917.

[60] A. Ramesh et al. “Evidence on the effectiveness of water, sanitation, and hygiene (WASH) interventions on health outcomes in humanitarian crises: a systematic review”. en. In: *PloS ONE* 10.9 (2015).

[61] M. Ali et al. “The global burden of cholera”. en. In: *Bull. World. Health Organ.* 90 (2012), pp. 209–218.

[62] A.S. Azman, S.M. Moore, and J. Lessler. “Surveillance and the global fight against cholera: setting priorities and tracking progress”. en. In: *Vaccine* 38.Suppl 1 (2020).

[63] D. Legros. “Global cholera epidemiology: opportunities to reduce the burden of cholera by 2030”. en. In: *J. Infect. Dis.* 218.Suppl 3 (2018).

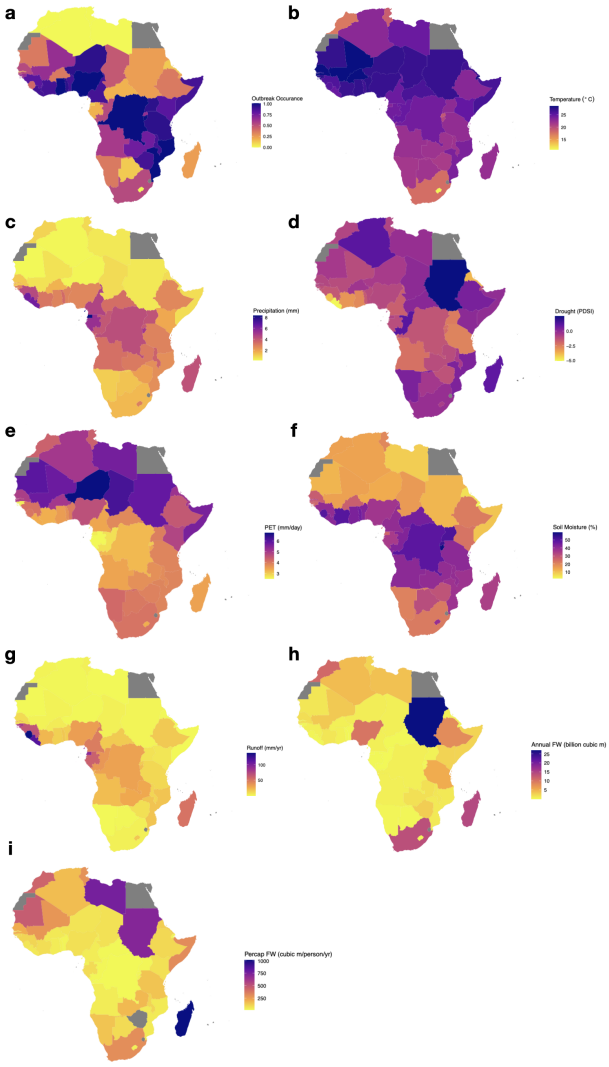
[64] M. Peel, B. Finlayson, and T. Mcmahon. “Updated World Map of the Koppen-Geiger Climate Classification”. In: *Hydrol. Earth Syst. Sci. Discuss.* 4 (Oct. 2007). DOI: 10.5194/hess-11-1633-2007.

[65] United Nations Statistical Division. *Millennium Development Goal Indicators*. en. 2015. URL: https://unstats.un.org/unsd/mdg/SeriesDetail.aspx?srid=580.

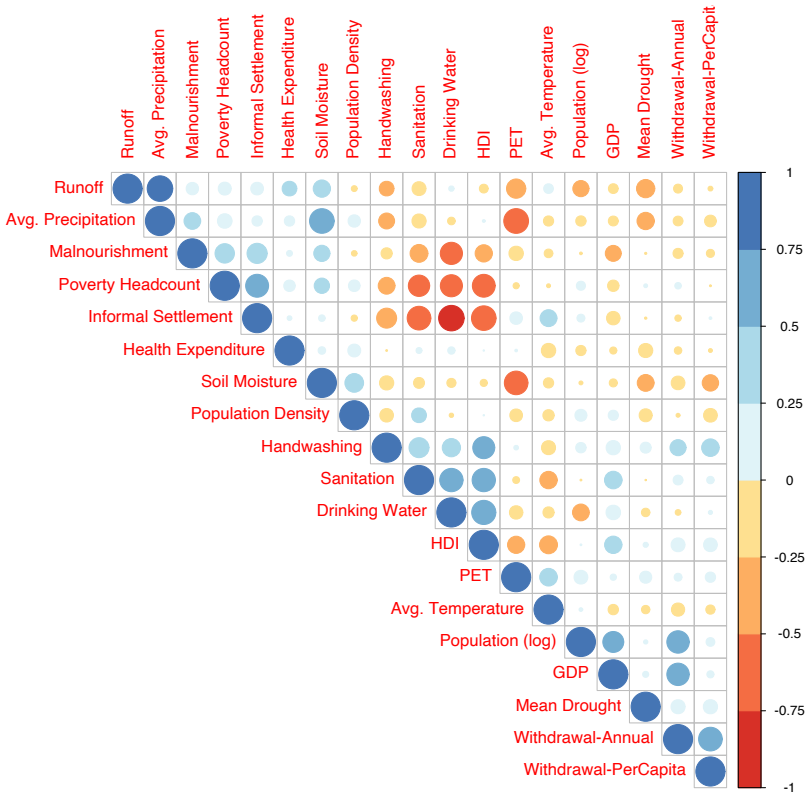
[66] Fragile States Index. *Measuring Fragility Risk and Vulnerability in 179 Countries*. en. 2021. URL: https://fragilestatesindex.org.

Supplementary Material

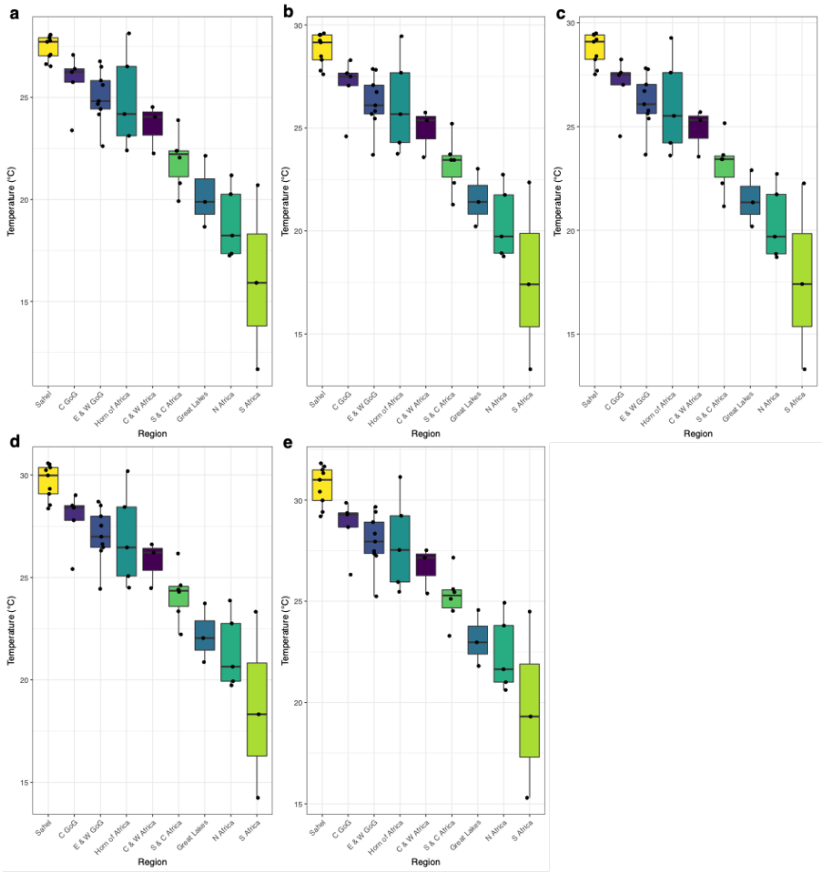
Supplementary Figures



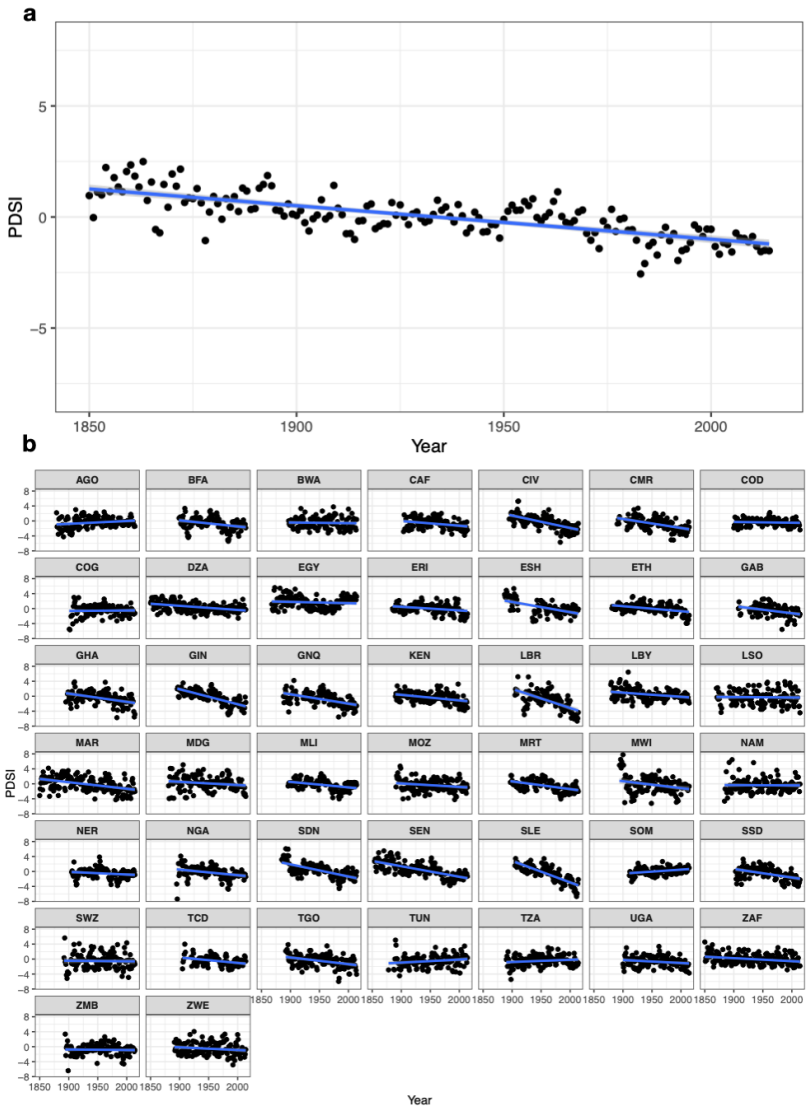
Supplementary Figure 3.1: National average for the instrumental period (2000-2016) for **a**, cholera outbreak occurrence, **b**, temperature (CO_2), **c**, precipitation (mm), **d**, Palmers Drought Severity Index (PDSI), **e**, potential evapotranspiration (PET), **f**, soil moisuture (%), **g**, water runoff (mm/year), **h**, annual freshwater withdrawal (FW, billion m^3) and **i**, Per capita freshwater withdrawal (FW, billion m^3/person/year).



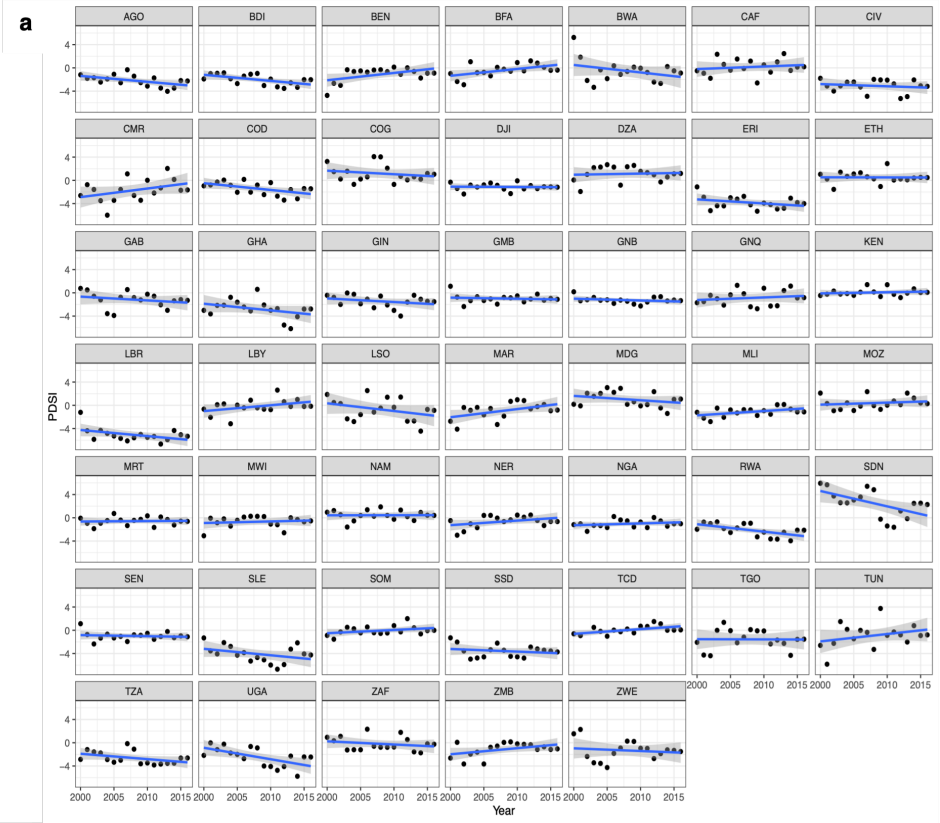
Supplementary Figure 3.2: Correlation plot for the Pearson correlation coefficients of the nineteen covariates included in the covariate selection process in Chapter 3. Positive residuals are blue suggesting a strong positive association between the corresponding row and column and negative residuals are in red, suggesting a negative association.

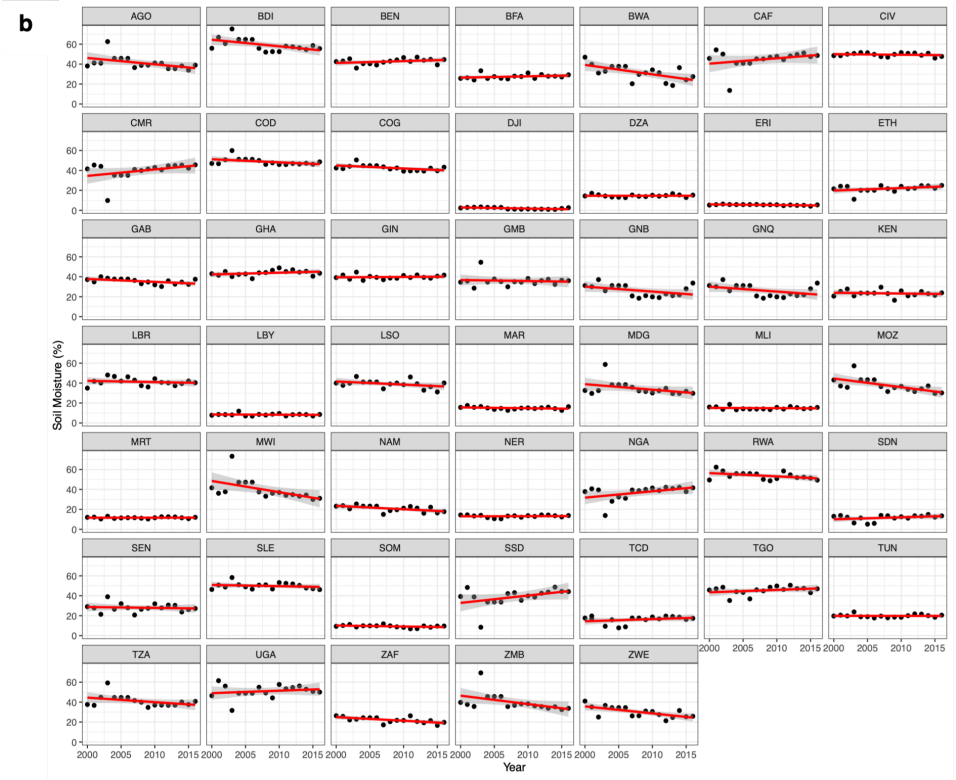


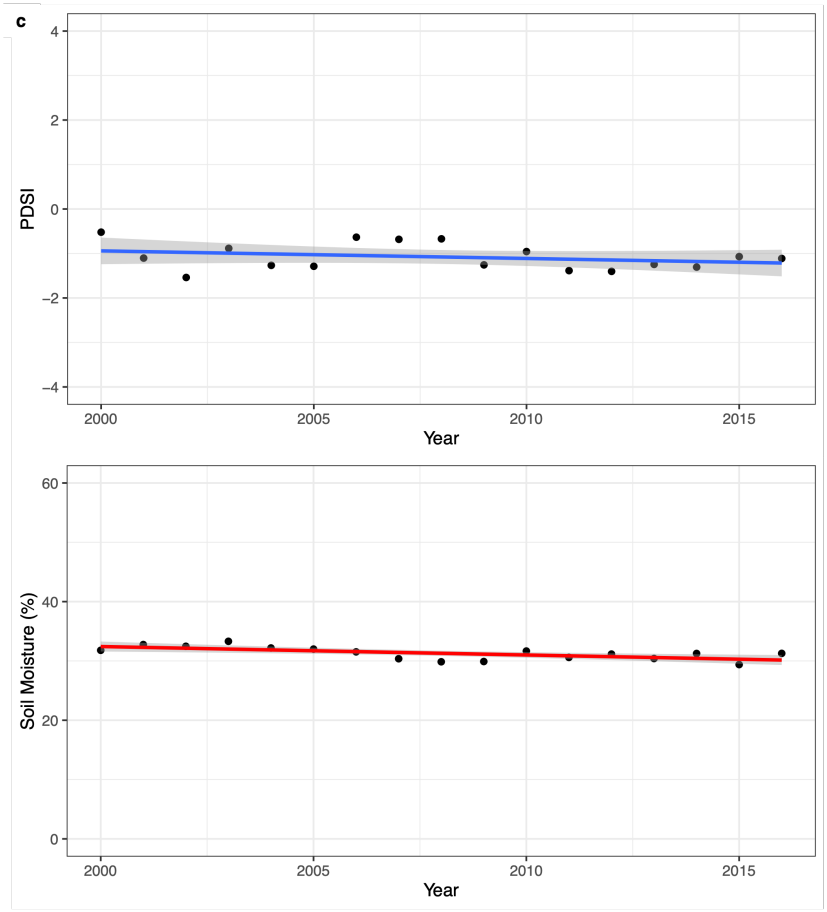
Supplementary Figure 3.3: Summary boxplots with jitters of the projected temperature data from WorldClim for **a**, 2016, **b**, RCP2.6 2050, **c**, RCP2.6 2070, **d**, RCP8.5 2050 and **e**, RCP8.5 2070. The colours represent the different regions and are explained below:
C & W Africa: Central and West Africa (Angola, Congo, Democratic Republic of Congo).
C GoG: Central Gulf of Guinea (Benin, Côte d'Ivoire, Ghana, Togo, Tanzania).
E & W GoG: East and West Gulf of Guinea (Cameroon, Central African Republic, Equatorial Guinea, Gabon. Guinea, Liberia, Nigeria, Sierra Leone, South Sudan).
Great Lakes (Burundi, Rwanda, Uganda).
Horn of Africa (Djibouti, Eritrea, Ethiopia, Kenya, Somalia).
N Africa: Northern Africa (Algeria, Libya, Morocco, Tunisia).
S & C Africa: South and Central Africa (Madagascar, Malawi, Mozambique, Namibia, Zambia, Zimbabwe).
S Africa: Southern Africa (Botswana, Lesotho, South Africa).
Sahel (Burkina Faso, Chad, Gambia, Guinea-Bissau, Mali, Mauritania, Niger, Senegal, Sudan)



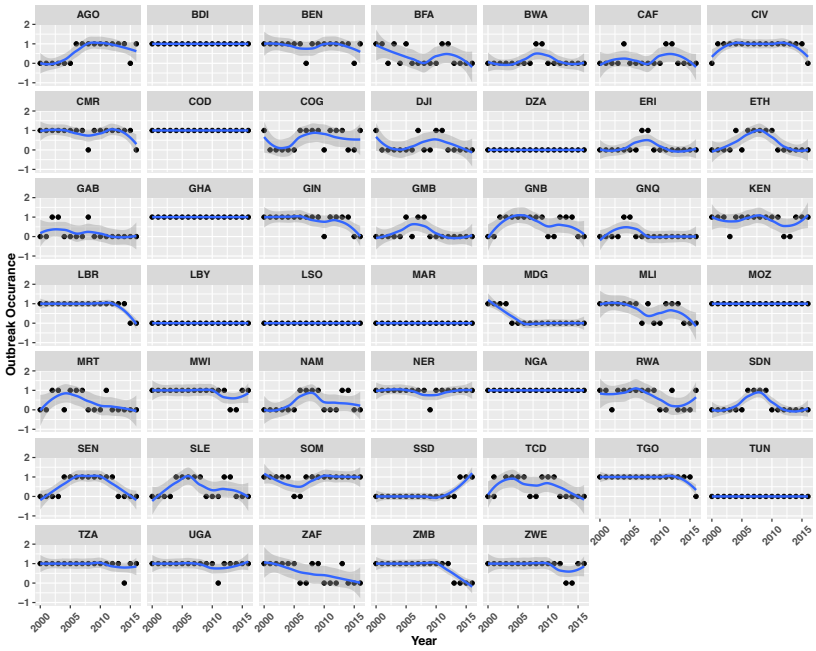
Supplementary Figure 3.4: Historical trends in PDSI on **a**, averaged over a continental scale and **b**, averaged over a national scale for the full dataset (1879-2016). Only 8 (BWA, COD, COG, NAM, NER, TZA, ZMB, ZWE) of the 47 countries showed an insignificant cholera trend at $p < 0.05$.


a
AGO
BDI
BEN
BFA
BWA
CAF
CIV
CMR
COD
COG
DJI
DZA
ERI
ETH
GAB
GHA
GIN
GMB
GNB
GNQ
KEN
LBR
LBY
LSO
MAR
MDG
MLI
MOZ
MRT
MWI
NAM
NER
NGA
RWA
SDN
SEN
SLE
SOM
SSD
TCD
TGO
TUN
TZA
UGA
ZAF
ZMB
ZWE
PDSI
Year

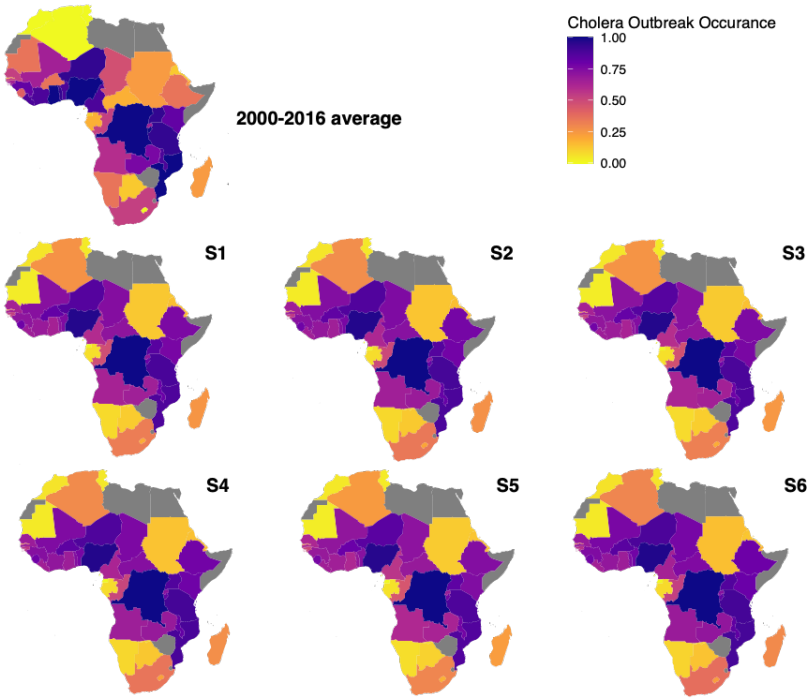

b
AGO
BDI
BEN
BFA
BWA
CAF
CIV
CMR
COD
COG
DJI
DZA
ERI
ETH
GAB
GHA
GIN
GMB
GNB
GNQ
KEN
LBR
LBY
LSO
MAR
MDG
MLI
MOZ
MRT
MWI
NAM
NER
NGA
RWA
SDN
SEN
SLE
SOM
SSD
TCD
TGO
TUN
TZA
UGA
ZAF
ZMB
ZWE
Soil Moisture (%)
Year



Supplementary Figure 3.5: A comparison of historical trends in **a**, national PDSI, **b**, national soil moisture and **c**, continental averages of PDSI and soil moisture, over the instrumental period.



Supplementary Figure 3.6: Loess curves showing year and outbreak for each country (2000-2016) with a smoothed trend line and standard error.



Supplementary Figure 3.7: Projected cholera outbreak occurrence using the best fit model fitted to the 2016 data and drought data altered in isolation from the 2000-2016 national averages as a baseline and the six different drought sensitivity analysis.

Supplementary Tables

Supplementary Table 3.1: Studies linking cholera outbreaks to specific risk factors used to establish the prior hypothesis and nineteen tested covariates in Chapter 3. This is in addition to those referenced in the main text.

Source	Covariate
Nsagha DS, et al. Assessing the risk factors of cholera epidemic in the Buea Health District of Cameroon. BMC Pub Health. 2015;15:1-7.	Hygiene and water
Labite H, et al. Quantitative Microbial Risk Analysis to evaluate health effects of interventions in the urban water system of Accra, Ghana. Journal Water Health. 2010;8:417-430.	Sanitation and water
Alsan MM, et al. Poverty, global health, and infectious disease: lessons from Haiti and Rwanda. Infect Dis Clin. 2011;25:611-22.	Poverty, health expenditure and development
Talavera A, Perez EM. Is cholera disease associated with poverty?. Journal Infect Dev Countr. 2009;3:408-11.	Poverty
Gidado S, et al. Cholera outbreak in a naïve rural community in Northern Nigeria: the importance of hand washing with soap, September 2010. Pan Afr Med J. 2018;30.	Hand washing
Aggrey-Korsah E, Oppong J. Researching urban slum health in Nima, a slum in Accra. Spatial Inequalities 2013 (pp. 109-124). Springer, Dordrecht.	Informal settlement
Penrose K, et al. Informal urban settlements and cholera risk in Dar es Salaam, Tanzania. PLoS Negl Trop Dis. 2010;4:e631.	Informal setttlement, population density
Ververs M, Narra R. Treating cholera in severely malnourished children in the Horn of Africa and Yemen. Lancet. 2017;390:1945-6.	Malnourishment
Osei FB, Duker AA. Spatial and demographic patterns of cholera in Ashanti region-Ghana. Inter J Health Geogr. 2008;7:1-0.	Population, total and density

Supplementary Table 3.2: Linear model results for the S3 drought extrapolation projections in Chapter 3, Res = Residual.

Code	Coeff	P value	SE	R^2	Res. Min	Res. 1Q	Res. Median	Res. 3Q	Res. Max
AGO	0.007584	1.38E-02	0.003	0.0405	-3.4003	-0.9893	-0.1943	1.0217	3.6879
BFA	-0.016124	4.80E-04	0.004	0.1018	-4.3346	-0.7317	0.0433	1.0693	3.4803
BWA	-0.001625	6.84E-01	0.004	0.001437	-2.7141	-1.1711	-0.2378	0.9759	4.3036
CAF	-0.012484	2.20E-03	0.004	0.08274	-2.7737	-0.9622	-0.1477	0.9576	2.8916
CIV	-0.033197	6.41E-15	0.004	0.4091	-4.3856	-0.8907	0.1215	0.7528	4.2415
CMR	-0.02412	1.21E-08	0.004	0.2526	-3.6014	-0.8186	-0.1533	0.992	3.515
COD	-0.002054	4.30E-01	0.003	0.005524	-2.0415	-0.6651	0.01018	0.6305	2.6456
COG	0.001273	7.72E-01	0.004	0.0007691	-5.0383	-0.806	0.1266	0.9635	3.6098
DZA	-0.011053	1.71E-08	0.002	0.1777	-2.7972	-0.82673	0.07252	0.70509	2.95153
ERI	-0.009764	1.18E-03	0.082	0.08161	-2.64799	-0.90982	-0.04446	0.79464	3.08663
ETH	-0.01341	2.14E-07	0.002	0.185	-3.11721	-0.63626	-0.04943	0.64698	2.80778
GAB	-0.019066	3.92E-04	0.005	0.1297	-4.6503	-0.7585	0.0266	1.0521	2.9602
GHA	-0.22088	8.64E-07	0.004	0.1891	-4.6036	-0.7854	0.0653	0.9259	4.4206
GIN	-0.040172	2.00E-16	0.003	0.5475	-3.4633	-0.7489	0.0282	0.8407	3.4562
GNQ	-0.024931	2.12E-09	0.004	0.2669	-4.8	-0.9298	0.0909	1.0086	3.8341
KEN	-0.014909	7.84E-06	0.003	0.1505	-2.5614	-0.9082	-0.0979	0.9024	4.3328
LBR	-0.050961	3.19E-12	0.006	0.3861	-5.8082	-1.2037	0.0361	1.3154	4.8078
LBY	-0.011042	5.01E-04	0.003	0.08674	-3.4071	-0.9486	-0.0678	0.9751	5.5876
LSO	-0.001181	1.18E-03	0.004	0.0006166	-4.1272	-1.4807	-0.0993	1.5032	4.2181
MAR	-0.01744	3.44E-09	0.003	0.1934	-4.4451	-1.2923	0.0284	1.2875	3.8972
MDG	-0.008485	3.89E-02	0.004	0.03191	-4.6279	-0.8901	-0.1081	1.0532	4.5956
MLI	-0.014778	1.21E-05	0.003	0.1515	-3.3191	-0.6955	-0.0572	0.8934	2.9524
MOZ	-0.009139	2.35E-02	0.004	0.04134	-4.0799	-1.0939	-0.0518	1.0867	4.6014
MRT	-0.020671	2.17E-08	0.003	0.245	-2.9844	-0.9335	0.0472	0.7782	2.7171
MWI	-0.018477	1.81E-03	0.006	0.08019	-5.6662	-1.2043	0.1717	1.2017	7.0808
NAM	-0.0006466	8.77E-01	0.004	0.0001893	-4.2969	-1.1374	-0.1989	0.6849	6.7976
NER	-0.007139	6.45E-02	0.004	0.0313	-3.4058	-0.7212	0.0868	0.7722	4.3451
NGA	-0.014961	5.12E-04	0.004	0.09844	-7.9324	-0.5473	0.1837	0.8922	3.6286
SDN	-0.032495	2.00E-16	0.003	0.4256	-2.7332	-0.9581	0.1599	0.9727	3.9191
SEN	-0.02812	2.00E-16	0.002	0.4758	-2.8209	-0.987	0.0366	0.6876	3.6725
SLE	-0.057589	2.00E-16	0.004	0.6542	-3.1287	-1.0012	-0.171	0.9413	3.7745
SOM	0.01051	7.76E-04	0.003	0.1053	-2.33652	-0.61295	0.08567	0.58038	2.50439
SSD	-0.022848	5.11E-07	0.004	0.2057	-3.8764	-0.9331	0.1218	1.0791	2.9932
TCD	-0.014613	5.53E-04	0.004	0.1141	-2.9722	-0.9354	0.0408	0.8491	3.682
TGO	-0.017806	1.62E-05	0.004	0.1429	-4.9207	-0.9181	0.26437	1.0905	4.2196
TUN	0.008081	3.95E-02	0.004	0.03347	-3.6059	-1.0549	0.0231	0.9955	6.0856
TZA	0.005739	9.72E-02	0.003	0.02221	-4.6694	-1.0467	-0.0035	1.072	3.4672
UGA	-0.008039	4.09E-02	0.004	0.03648	-3.0538	-0.9776	-0.0226	0.8873	4.1223
ZAF	-0.007972	4.19E-04	0.002	0.07372	-2.6528	-1.0104	-0.2118	0.9701	3.9207
ZMB	-0.001333	7.63E-01	0.004	0.001035	-5.654	-0.9028	0.0596	0.8424	4.1035
ZWE	-0.007225	8.62E-02	0.004	0.02394	-4.0278	-1.1522	-0.2328	1.0939	4.3964

Supplementary Information

Supplementary Information 3.1: Additional information about the drought indices data used in Chapter 3.

Palmer Drought Severity Index - The calculation of PDSI relies on a water balance-based two-bucket system, where there is a surface layer with a storage capacity of 1 in and an underlying layer with a storage capacity of -1 (Jacobi *et al.*, 2013). PDSI was first developed by Palmer (1965) and uses temperature and PET data, capturing the basic effects of climate change through PET changes (Dai & NCAR, 2019). Here, PDSI was used to quantifying long-term drought and was taken as a national average from the annual administrative level 1 data.

Sources: Jacobi J, Perrone D, Duncan LL, Hornberger G. A tool for calculating the Palmer drought indices. Water Resources Research 2014;49(9):6086-6089. Palmer WC. Meteorological drought (Vol. 30). 1965.US Department of Commerce, Weather Bureau. Dai A. and National Centre for Atmospheric Research. The Climate Data Guide: Palmer Drought Severity Index (PDSI). [On-line].

Soil moisture - The dataset provides estimates of soil moisture from satellite sensors over the globe. It is based on the European Space Agency's Climate Change Initiative soil moisture version 3.3. Data were provided as a netCDF file for each month from 2000-2016 on a regular latitude/longitude grid at 0.25x0.25 resolution on a monthly temporal resolution. The data were processed by transforming the co-ordinates to each country (by creating a function using the R packages "sp" and "rworldmap") and creating a national annual mean. This provided surface soil moisture as a percentage content of liquid water in a surface soil layer of 2-5cm depth, expressed as the percentage of total saturation.

Source: Copernicus, 2018. Soil moisture gridded data from 1978 to present. [On-line]. European Centre for Medium-Range Weather Forecasts.

Potential Evapotranspiration - Data were taken from the National Centre for Atmospheric Research (NCAR) climate data guide which comes from the Climate Research Unit Timeseries (CRU TS) series of datasets, specifically CRU TS4.0. The data contains several climate fields including precipitation, temperature and cloud cover, which are used to compute variables including evapotranspiration in mm per day. The data covers all land surface and spans from 1901-2015, although data here were only extracted for 2001-2015, as this best captures the instrumental period of this research. The data were gridded at 0.5x0.5° resolution based on 4,000 individual weather station data sources; this requires

homogenisation which can introduce limitations as the data will not be strictly homogeneous. The data were provided as two netCDF format cells (2001-2010 and 2011-2015), which were then extracted into a .csv file as a national average for each year.

Source: NCAR, 2015. CRU TS GRIDDED PRECIPITATION AND OTHER METEOROLOGICAL VARIABLES SINCE 1901. [On-line].

Runoff - Observed river discharge information and a climate driven water balance model were combined in order to develop composite runoff fields which are consistent with observed discharges. The aim of the methods used is to provide a best estimate of terrestrial runoff over large domains in mm per year. The composite fields use data from a gridded river network at 30-minute spatial resolution to represent the riverine flow pathways and to link the continental land mass to oceans through river channels. This used selected gauging stations from the Global Runoff Data Centre (GRDC). Inter-station discharge and runoff were calculated to compare observed runoff with outputs from water balance model simulations. Correction coefficients based on the ratio of observed and simulated runoff for inter-station areas were calculated and applied against simulated runoff to create composite runoff fields. The GRDC also provided data for each station as individual .txt files but once processed the files were missing large amounts of data and therefore the composite fields were instead used, given as .grd files for each month.

Source: UNH/GRDC, 2000. UNH/GRCD Composite Runoff Fields V1.0. [On-line].

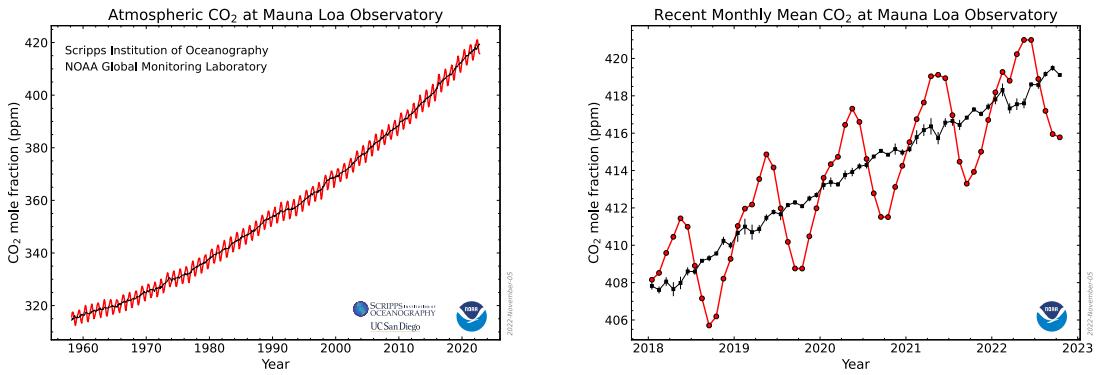
Annual freshwater withdrawal - Is provided as a .csv file containing county, year and annual freshwater withdrawals in billion m^3 from 1962-2016. All data outside the instrumental period (2000-2016) were removed and a national average taken. Annual freshwater withdrawals refer to total water withdrawals, not counting evaporation losses from storage in basins. Withdrawals also include water from desalination plants in countries where they are a significant source. Withdrawal for agriculture and industry are total withdrawals for irrigation and livestock production and for direct industrial use. Domestic uses include drinking water, municipal use or supply and public services. The source of this data is the Food and Agriculture Organization's AQUASTAT data. Data were collected intermittently and subject to variations in collection and estimation methods, which may hide significant variations in water availability within countries. Data for smaller countries or in arid and semi-arid areas are less reliable than those for large countries and countries with greater rainfall.

Per capita freshwater withdrawal - The source of the per capita data were also AQUASTAT and presented in the same way. The length of the dataset was 1960-2015 and was calculated from the annual quantity of water withdrawn for agricultural, industrial and municipal purposes. It can include water from primary and secondary resources as well as from over-abstraction of renewable groundwater, fossil groundwater, agricultural drainage water, treated wastewater and desalinated water. It does not include in-stream use which typically has very low net consumption rates such as recreation, navigation, hydropower and inland capture fisheries.

Source: Ritchie, H., 2017. Water Use and Stress. [On-line]. Our World In Data.

Supplementary Information 3.2: Additional information about the projection data used in Chapter 3.

Projected temperature data were available through WorldClim, which gives projections for 2050 and 2070, under the four Representative Concentration Pathways (RCP) emissions scenarios. RCP2.6 requires emissions to start declining by 2020 and reach net zero by 2100, RCP4.5 requires emissions to peak in 2040, RCP6.0 peaks at around 2080 and RCP8.5 assumes emission will continue to rise throughout the century. These four scenarios are projected to have mean global warming by 2081-2100 of 1°C, 1.8°C, 2.2°C and 3.7°C, respectively (IPCC, 2014). RCP2.6 was not used in the projections, as this requires CO_2 emissions to peak in 2020, a goal which has not been reached (See Figure below). The data were presented as a monthly average at administrative level 1, which were then transformed into a national yearly mean.



Monthly mean carbon dioxide measured at Mauna Loa Observatory, Hawaii. The left panel shows the full record of combined Scripps data and NOAA data. Every monthly mean is the average of daily means, which are in turn based on hourly averages. The right panel shows the last five complete years, plus the current year. The red lines represent the monthly mean values, centered on the middle of each month. The black lines represent the same, after correction for the average seasonal cycle (NOAA, 2022).

Projected PDSI data were difficult to obtain, and previous modelling studies have found spatial heterogeneity, making projecting drought across the continent challenging. There is disagreement over drought changes in Africa under climate change and how populations will adapt to alterations in water security (Ahmadalipour *et al.*, 2019; Calow *et al.*, 2010; Haile *et al.*, 2020; Shanahan *et al.*, 2009; Touchan *et al.*, 2008; Verschuren *et al.*, 2000). One of the main difficulties in calculating projected PDSI and the discrepancies in the results are the algorithms used to calculate PET (Tian-Jun and Tao,

2013). PET algorithms do not account for changes in vegetation cover expected due to elevated CO_2, making the "warming leads to drying" narrative flawed and over-simplistic, causing over-estimations (Yang *et al.*, 2020). Several studies which have taken a more critical approach to the methodology found that drought did not significantly change over long time periods and changes are likely to be seen on finer spatial scales (Sheffield *et al.*, 2012; Yang *et al.*, 2020). These findings are consistent with tree-ring data (Touchan *et al.*, 2008), lake sediment records (Verschuren *et al.*, 2000) and the PDSI dataset used here. This suggests that droughts are neither more severe nor longer now than historically.

Despite the large number of projected drought studies, extracting these results and using them here is challenging, as they are on different spatial scales, using different methodology and data. There are also issues with the projected indices used, as these often vary and are potentially not comparable. To account for this, three scenarios were created from historical PDSI data using univariate linear regression models for drought and year. This method used the coefficients for each country to project the future drought data, with Scenario 3 continually plotting the coefficient values until 2070 (or it reached -4 or 4, the extremes of the PDSI scale), Scenario 1 accepted the above hypotheses that PDSI is an overestimation and drought will not change over the projected period and Scenario 2 was a median value.

Projected population data were based on the United Nation's World Population Prospectus (2019). The projections are based on available data on population size, levels of fertility, mortality and international migration. Data are from censuses, registration of births and deaths, demographic and health surveys, official statistics and population registers. More recently, the data have taken into account refugee statistics, prevalence of HIV and anti-retroviral coverage, infant and under five morality and migration flows. The projections use the cohort component method using a variety of demographic assumptions concerning fertility, mortality and migration. This takes into account the past experiences of the country and reflects uncertainty and other countries in similar conditions. The medium variant projection corresponds to the median of several trajectories of each demographic component derived using a probabilistic model of the historical variability over time. Prediction intervals represent the spread in the distribution of outcomes across the projected trajectories and thus provide an assessment of the uncertainty inherent in the medium variant projection. Therefore, only the medium variant projection was included in the model and not as an offset.

Projected poverty headcount at <$1.90 a day was based on SDG1.1 (UN, 2015), which states that by 2030, extreme poverty (<$1.25/day) will be eliminated and SDG1.2, that the population living in poverty will be reduced by >50%. The goals are very ambitious and will require significant human and economic resources. Several of the terms within the SDGs are also ambiguous, meaning the aims and roadmap to achieving them are not clear. With regards to poverty, the setting used in the SDGs is slightly lower ($1.25 compared to $1.90) and it is difficult to distinguish the level of poverty within the data; therefore, the projected scenario will mainly align with SDG1.2, to halve the population in poverty by 2030. Despite their limitations, the SDGs provide a globally recognised pathway to a sustainable future and what institutions and governments should be aiming to achieve, making their use important in scientific research.

Creating projected freshwater withdrawal per capita data for the scenarios also presented challenges, as this is largely down to human behaviour and therefore hard to predict. While SDG6.4 states that by 2030, water use efficiency will be sustainably increased across all sectors. Projected data are not freely available and may not take into account climate change and alterations in societal behaviour, all of which will alter water stress in the future. To maintain sustainable levels of water resources, rates of withdrawal need to be lower than replenishment. Renewable resources come from internal river flows and groundwater from rainfall. To understand the national historical water security, data were plotted for both freshwater withdrawal and freshwater resources (Ritchie, 2017). Most countries in Africa have both low water use and resources (Figure 3.2), exceptions to this include Gabon, Republic of Congo and Liberia which have comparatively high resources and low withdrawal as well as Madagascar, Libya, Sudan, Mauritania and Morocco which have comparatively low resources and high use. Given the relationship between cholera outbreaks and water withdrawal, most countries in Africa could increase their freshwater withdrawal, except for Madagascar, Libya, Sudan, Mauritania and Morocco which should reduce their use, to improve sustainability.

Sources:

Intergovernmental Panel on Climate Change, 2014. Climate Change 2014 Synthesis Report Fifth Assessment Report Future Climate Changes, Risks and Impacts. [On-line].

NOAA, 2022. Trends in Atmospheric Carbon Dioxide. [On-line].

Ahmadalipour, A., Moradkhani, H., Castelletti, A. and Magliocca, N., 2019. Future drought risk

in Africa: Integrating vulnerability, climate change, and population growth. Science of the Total Environment, 662, pp.672-686.

Calow, R.C., MacDonald, A.M., Nicol, A.L. and Robins, N.S., 2010. Ground water security and drought in Africa: linking availability, access, and demand. Groundwater, 48(2), pp.246-256.

Haile, G.G., Tang, Q., Hosseini-Moghari, S.M., Liu, X., Gebremicael, T.G., Leng, G., Kebede, A., Xu, X. and Yun, X., 2020. Projected impacts of climate change on drought patterns over East Africa. Earth's Future, 8(7), p.e2020EF001502.

Shanahan, T.M., Overpeck, J.T., Anchukaitis, K.J., Beck, J.W., Cole, J.E., Dettman, D.L., Peck, J.A., Scholz, C.A. and King, J.W., 2009. Atlantic forcing of persistent drought in West Africa. Science, 324(5925), pp.377-380.

Touchan, R., Anchukaitis, K.J., Meko, D.M., Attalah, S., Baisan, C. and Aloui, A., 2008. Long term context for recent drought in northwestern Africa. Geophysical Research Letters, 35(13).

Verschuren, D., Laird, K.R. and Cumming, B.F., 2000. Rainfall and drought in equatorial east Africa during the past 1,100 years. Nature, 403(6768), pp.410-414.

Tian-Jun, Z. and Tao, H., 2013. Projected changes of palmer drought severity index under an RCP8. 5 scenario. Atmospheric and Oceanic Science Letters, 6(5), pp.273-278.

Yang Y,, Zhang S., Roderick M. L., McVicar, T. R., Yang D., Liu W. And Li X. 2020. Little change in Palmer Drought Severity Index across global land under warming in climate projections. Hydrology and Earth System Sciences Discussions.

Sheffield, J., Wood, E.F. and Roderick, M.L., 2012. Little change in global drought over the past 60 years. Nature, 491(7424), pp.435-438.

United Nations, 2019. World Population Prospectus 2019. [On-line].

United Nations, 2015. The 17 Goals. [On-line]. United Nations Department for Economic and Social Affairs. Population Dynamics.

Ritchie, H., 2017. Water Use and Stress. [On-line]. Our World In Data.

Chapter 4

Using Self-Controlled Case Series to Understand the Relationship between Conflict and Cholera in Nigeria and the Democratic Republic of Congo

Dissemination

An extended version of the methods regarding data for this chapter is published at:

Charnley GEC, Kelman I, Gaythorpe KAM, Murray KA. Accessing sub-national cholera epidemiological data for Nigeria and the Democratic Republic of Congo during the seventh pandemic. *BMC Infectious Diseases* 2022;22:288.

A modified version of the full chapter is published at:

Charnley GEC, Jean K, Kelman I, Gaythorpe KAM, Murray KA. Using self-controlled case series to understand the relationship between conflict and cholera in Nigeria and the Democratic Republic of Congo. *Emerging Infectious Diseases* 2022;28:2472-2481.2021;21:1177.

Abstract

Cholera outbreaks significantly contribute to disease mortality and morbidity in low-income countries. Cholera outbreaks have several social and environmental risk factors and extreme conditions can act as catalysts. A social extreme with known links to infectious disease outbreaks is conflict, causing disruption to services, loss of income and displacement. Here, the self-controlled case series method was used in a novel application and found that conflict increased the risk of cholera in Nigeria by 3.6 times and 19.7% of cholera outbreaks were attributable to conflict. In the DRC, conflict increased the risk of cholera by 2.6 times and 12.3% of cholera outbreaks were attributable to conflict. Several states/provinces with the strongest relationship were also areas of high reported conflict. Our results help highlight the importance of rapid and sufficient assistance during social extremes and the need for conflict resolution and addressing pre-existing vulnerabilities such as poverty and access to healthcare.

4.1 Introduction

Chapter 3 investigated the implications of environmental extremes on cholera outbreaks but social extremes (e.g., conflict) are also known to cause disease outbreaks, including cholera. Here, the aim is to investigate the impact of conflict on cholera outbreaks, as a potential explanation for why some high burden countries did not project decreased cholera outbreak risk in Chapter 3. The high cholera occurrence, even in the most optimistic future scenarios may be due to additional variables, not considered in the previous chapter (residual confounding). Few studies have investigated the impacts of conflict on cholera outbreaks, especially quantitatively. Studies have commonly focused on cholera and conflict in Yemen [1, 2, 3], its effect on vaccination efforts [4] or the impact of conflict on other diseases such as Ebola [5] and COVID-19 [6].

The impacts of conflict on cholera outbreaks will be compared across two countries in Africa, Nigeria and the DRC in the past 23 years. Nigeria and the DRC share several social and environmental similarities, as well as experiencing cholera outbreaks, making them comparative countries in this analysis. Both have active conflicts including the Boko Haram insurgency in northeastern Nigeria [7] and political unrest in the eastern DRC [8]. In terms of cholera burden, Nigeria and the DRC have the second and third highest numbers of estimated cases per year in Africa, respectively [9], with the Kivu

provinces being the most active cholera foci in the world [10]. In addition, Nigeria and the DRC have a tropical climate, poor access to WASH and a large proportion of the population living in poverty (<$1.25/day) at 87.7% for the DRC and 62% for Nigeria [11], which are all known cholera risk factors.

A novel methodological approach will be used to bridge the research gaps between conflict and cholera. The Self-Controlled Case Series (SCCS) has previously been used as a method to quantify the effectiveness of drug and vaccine intervention on an individual level [12, 13, 14] and more recently, at a population level [15]. Chapter 4 aims to enhance the understanding of the method and promote its use in other contexts. The method will be applied both nationally and sub-nationally and sensitivity analyses used to provide insight into the duration and lag of the effect, along with the implications of different cholera definitions. Furthermore, the recently developed percentage attributable fraction (PAF) equations will be adapted to the work presented here [15], to understand the proportion of cholera outbreaks attributable to conflict.

Based on the results from the analysis, mechanisms through which conflict is driving cholera and potential risk factors will be suggested, building on previous research in this area. The conclusions of Chapter 4 can be used to strengthen disease prevention in conflict settings and reduce additional mortality and morbidity in conflicts. The objectives for Chapter 4 include:

1. Understand the relationship between conflict and cholera in Nigeria and the DRC on a national and a sub-national level.
2. Investigate the duration of this effect and any potential lag effects, while accounting for temporal and spatial autocorrelation.
3. Suggest potential risk factors that may be driving conflict-related cholera outbreaks.

4.2 Methods

4.2.1 Datasets

Cholera data were compiled from a range of publicly available sources and all data on cases and deaths available were included, which spanned from 1971-2021 for Nigeria and 1978-2021 for the DRC on a

daily temporal scale and was provided at the finest spatial scale possible. The sources included WHO disease outbreak news [16], ProMED (including ReliefWeb) [17], WHO regional office for Africa weekly outbreak and emergencies [18], UNICEF cholera platform [19], EM-DAT [20] and the Nigerian Centre for Disease Control [21]. A literature search of MEDLINE, Embase, Global Health and Google Scholar (with snowballing of reference lists) in both English and French was also used to identify additional sources. The eligibility criteria for the literature are shown below in Table 4.1 (the included literature is in Supplementary Information 4.1).

Table 4.1: Eligibility criteria for the literature included in the data fitted to the SCCS models.

Inclusion Criteria	
Population	Any local population/community impacted by a cholera outbreak in Nigeria or the DRC
Intervention	Any investigation carried out to quantify and understand cholera cases/deaths and risk factors
Comparator	Anyone in the affected community which did not become infected with cholera
Outcomes	The outcome is to understand epidemiological features of cholera outbreaks in Nigeria and the DRC
Study type	Retrospective observational reports/studies including cross-sectional, case-control and cohort studies
Exclusion Criteria	
Studies which investigate a diarrhoeal disease outbreak with no specific mention of cholera	
Serological surveys evaluating antibody levels in the population and not a specific outbreak with active cases	
Review papers, as only primary sources were used	
Publications looking at public health and cholera prevention more generally and not in relation to a specific outbreak response	
Non-English and non-French abstracts and full-texts, due to linguistical constraints	
Multiple sources where the information reported was the same	

A data charting form was used to compile and update the data and the full compiled datasets are available in a GitHub repository (https://github.com/GinaCharnley/cholera_data_drc_nga). The

data were at a daily temporal scale and to the finest spatial scale the source allowed, which in some instances was to local government area in Nigeria and health zone in the DRC. Additionally, data on cases, deaths, confirmed cases, age, sex, case fatality ratio and any reported risk factors were noted if available and ordered by the date of the new reported cases or deaths for the outbreak. To investigate differences in cholera reporting in each country, the cholera outbreak definition was consulted and shown below:

- Nigeria Centre for Disease Control - Suspected case: Severe dehydration or death from acute watery diarrhoea in a patient aged 5 years or more. In an epidemic situation: A suspected case in any person aged 5 years or more with acute watery diarrhoea with or without vomiting. Confirmed case: A suspected case in which *Vibrio cholerae* O1 or O139 has been isolated in the stool [22].
- Ministère de la Santé Publique de la République démocratique du Congo - Suspected case: Severe dehydration or death following acute watery diarrhoea in a patient aged 5 years or more. In an epidemic situation: Acute watery diarrhoea with or without vomiting in a patient aged 1 year or more [23].

Both definitions are very similar and are relatively standard case definitions for cholera, therefore it is unlikely that this would alter cholera reporting in Nigeria and the DRC. Although the slightly younger age threshold in the DRC may have resulted in more reported cases. As a method of validating the reports used, multiple reports of the same outbreak were recorded and an average of several sources used. Additionally, the data collected was significantly correlated ($p < 0.05$) to the WHO Global Health Observatory Data and two private data sources provided by the Nigeria Centre for Disease Control and Johns Hopkins Bloomberg School of Public Health. The private data sources were not later used for model fitting, as this was a retrospective analysis for data validation, due to the timing of data availability. The data collation and validation has undergone peer review and additional information is available in a complementary data paper [24].

Conflict data were provided by the United Nations Office for the Coordination of Humanitarian Affairs Humanitarian Data Exchange, which provides data from the Armed Conflict Location & Event Data Project (ACLED) [25]. The data included sub-national conflict events for both countries given to the exact location in longitude/latitude. The data were reported on a daily temporal scale and spanned

from 1997 to 2020 and categorised by event type. The categories included battles, explosions, protests, riots, strategic developments and violence against civilians. This was further sub-categorised within these groups and reported number of fatalities.

The spatial granularity of the analysis was to administrative level 1 (states for Nigeria and provinces for the DRC) and all data points that were reported on a finer spatial scale were aggregated to the upper level. The study period was specified as January 1997 to May 2020, as these were the first and last reports in the conflict data (the cholera data had reports from 1970 to 2021). The temporal scale was set to weekly, with continuous epidemiological weeks from week 1 (January) in 1997 to week 20 (May) in 2020, this resulted in 1,220 continuous weeks (1-1,220). Continuous weeks was chosen for compatibility with the model and to include periods of conflict that endured from one year into the next (as epidemiological weeks re-set to 1 at the end of each year). Weeks was chosen, rather than days, to account for reporting lags. Previous work has reported issues in the granularity of data and timeliness of reporting, especially in humanitarian crises, due to different sources of data and logistical difficulties [26, 27, 28].

4.2.2 Model Structure and Fitting

The SCCS method investigates the association between an exposure and an outcome event. The aim was to estimate the effect, by comparing the relative incidence of the adverse events (outbreaks) within an exposure period of hypothesised excess risk (conflicts), compared to all other times (peace, according to the dataset used). The method is a case only method and has the advantage of not needing separate controls, by automatically controlling for fixed confounders that remain constant over the observational period using stratification [29, 30, 31].

Both the event (outbreaks) and exposure (conflict) were set as binary outcomes. Conflict data were daily reports of a new conflict, transformed to either being reported in the epidemiological week (1), or not (0). Cholera data were daily reports of new cases or deaths, transformed to either being reported in the epidemiological week (1), or not (0). The observation period was the full study period from epidemiological week 1 in 1997 to epidemiological week 20 in 2020 (1-1,220). The exposure period was the first week after conflict report onset and was reported as multiple onsets for each conflict, not one long exposure period incorporating all conflicts in that period. The event was defined by the

week the new cholera cases or deaths were reported. Each event and exposure that occurred in the same state/province and week were designated an identification number (termed: indiv, which here represents the person-time (state/province-week)) and a pre-exposure, exposure, and post exposure period (interval) (see Supplementary Table 4.1 for data setup and additional information).

The data were then fit to either conditional logistic regression models (R package "survival", function clogit() [32]) or generalised non-linear models (R package "gnm", function gnm() [33]). Both model options are comparable and yield similar results but conditional logistic regression models were chosen, due to their superior model fit according to BIC and AIC. However, AIC and BIC were only expected to see marginal differences, as model complexity was not altered. Additionally, conditional logistic regression models are superior if the events are recurrent and independent [31], which was assumed here (although the potential for spatial and temporal autocorrelation was tested and discussed).

Conditional logistic regression is an extension of logistic regression that allows for matching and stratification. The ability of conditional logistic regression to fit matched data means it is often used in case-control studies, as it allows for comparison between exposed and unexposed groups. Here, MLE was used for parameter estimation, based on an assumed Poisson probability distribution, with the number of outcome events (outbreaks) in the interval as the dependant variable, stratified by the individual (indiv). The outcome event, $Y = 1$, is the probability of cholera being reported. The likelihood, π, is a function of β which are the associated coefficients values, x_{ij} which is the exposure to a conflict event for i in interval j and t_{ij} which is the person-time (state/province-week) for i in interval j. The total log likelihood is the sum over all intervals j and person-times i, offset by the log of the time spent in the interval [34, 35].

$$log(\pi) = \frac{\pi}{1-\pi}.$$

$$log(\pi) = \sum_{i}\sum_{j}(\beta_1 x_{ij} + \beta_2 t_{ij} + 1log(j_{length})).$$

The length of the interval between exposed to unexposed was logarithmically transformed and used as an offset term in the function. An offset term is a predictor with fixed beta coefficients (at 1). The interval length was offset because it may have some predictive power e.g., a longer interval would

increase the chances of the event occurring within it. However, the predictive power would not be because the exposure increased the probability of the event but because there was a greater period of time for it to occur by chance. By holding the beta coefficient at 1, this evaluated the effect of j in time, in relation to cholera and conflict, not its length. To control for the fixed confounders, each exposure and event were allocated a unique number (indiv) and used as a stratified term in the model.

The model coefficient was given as a log rate ratio, which was then transformed exponential to incidence rate ratio (IRR) (IRR = exp(coefficient)). IRR is the ratio of the incidence rates in exposed and unexposed individuals and indicates the magnitude in which conflict increased the risk of cholera outbreaks. Expressed algebraically, ρ represents the rates in the exposed (ρ_1) and unexposed (ρ_2), e is the number of events and n is the size of the populations:

$$IRR = \frac{\rho_1}{\rho_2} = \frac{\frac{e_1}{n_1}}{\frac{e_2}{n_2}}.$$

The datasets for each country were then split by state/province and the analysis repeated for each. The aim of the sub-national analysis was to understand if the significance of conflict on cholera outbreaks varied by sub-national location and if conflict was more important in some states/provinces compared to others. All statistical analyses were carried out in R version 3.6.2 and the threshold for significance was $p < 0.05$.

4.2.3 Sensitivity Analysis

A sensitivity analysis was used to test different methods of defining the exposure end point (set to one week in the main analysis) and to understand the impact this would have on the results, at both a national and sub-national level. It allowed for further understanding of how long after a conflict event the risk of cholera is heightened. Five alternative exposure periods were tested from the original exposure period (1 week after the onset of exposure, lag 1) and were named lag periods due to the potential lag effect from conflict onset to cholera outbreaks, these included:

1. Lag 2 - Week of conflict onset + 2 weeks
2. Lag 4 - Week of conflict onset + 4 weeks

3. Lag 6 - Week of conflict onset + 6 weeks

4. Lag 8 - Week of conflict onset + 8 weeks

5. Lag 10 - Week of conflict onset + 10 weeks

Supplementary Figures 4.1 & 4.2 show additional swimmer plots of lag 10 and line plots of the temporal trends.

A second sensitivity analysis was used to understand the effect of altering the cholera outbreak definition (>1 cholera case reported in a specific week and state/province) and to test for the presence of temporal autocorrelation, by removing preceding events and re-testing. The analysis involved two scenarios, Scenario 1 removed all outbreaks within 2 weeks of each other (based on cholera biology, <10 days shedding the bacteria + <5 days incubation period) [36, 37]. Scenario 2 was an extreme scenario to fully test model robustness and removed all outbreaks within 6 months of each other.

4.2.4 Percentage Attributable Fraction

The recently developed percentage attributable fraction (PAF) equations [15] were adapted to the model results and the data used. PAF is an epidemiological measure of the impact of an exposure and is expressed as the fraction of all cases attributed to the exposure. Here, the PAF values were the percentage of outbreaks that could be attributed to conflict at a national level, using the datasets and the IRR values from the model output. Sub-national PAF values could not be obtained, as the data from the unexposed vs exposed states/provinces are needed to calculate the percentage. To calculate PAF, the number of outbreaks attributable to conflicts are first estimated, A_i, for each province i using the formula:

$$A_i = \lambda_i d_i^{E+}(IRR - 1).$$

Where d_i^{E+} is the total duration of conflict exposure for the province i (if no conflict in province i, thus $d_i^{E+} = 0$), λ_i is the rate of outbreak occurrence in a Poisson process in the absence of conflict, and IRR is the incidence rate ratio associated with exposure to conflict. The outbreak data were fit to a Poisson probability distribution to test the assumed probability distribution and simulated counts were

obtained from 10,000 random realizations of a Poisson process of rate λ = number of total national outbreaks/number of states or provinces. This leads to $\hat{A}$, the total number of outbreaks attributable to conflicts, where N_i^{E-} is the number of outbreaks observed in the province i during the un-exposed period and T being the total period of observation, an estimator of λ_i is $\hat{\lambda}_i = N_i^{E-}/(T - d_i^{E+})$:

$$\hat{A} = \sum_i \frac{N_i^{E-} d_i^{E+}}{T - d_i^{E+}} (I\hat{R}R - 1).$$

Bootstrap resampling (1,000 samples) was used to obtain 95% confidence intervals for $\hat{A}$. For each sample, a value of IRR was randomly re-sampled around the mean, based on the parameters estimated in the SCCS analysis (R packages "dplyr" [38] and "resample" [39], functions confint(), do() and resample()). Using $\hat{A}$ and N, the total number of outbreaks observed, the equivalent of the population attributable fraction can be easily obtained, PAF, this is equivalent to the PAF obtained in classical epidemiological studies, but here population refers to the "population of states/provinces":

$$PAF = \frac{\hat{A}}{N}.$$

4.3 Results

4.3.1 Conflict and Cholera Occurrence

The distribution of conflict and cholera in Nigeria and the DRC in the datasets used here is shown temporally and spatially in Figure 4.1. The data shows an increase in reported conflict and cholera, especially after 2010 (Figure 4.1a) and a large proportion of the cholera cases have been reported in conflict-affected areas, namely eastern DRC and northeastern Nigeria (Figure 4.1b).

The total number of conflicts and outbreaks for each state/province during the study period is shown below in Figure 4.2 and totaled 8,190 conflicts and 782 cholera outbreaks for Nigeria and in the DRC, 4,639 conflicts and 396 cholera outbreaks. The outbreak distribution applied satisfactorily to the Poisson probability distribution (Supplementary Figures 4.3).

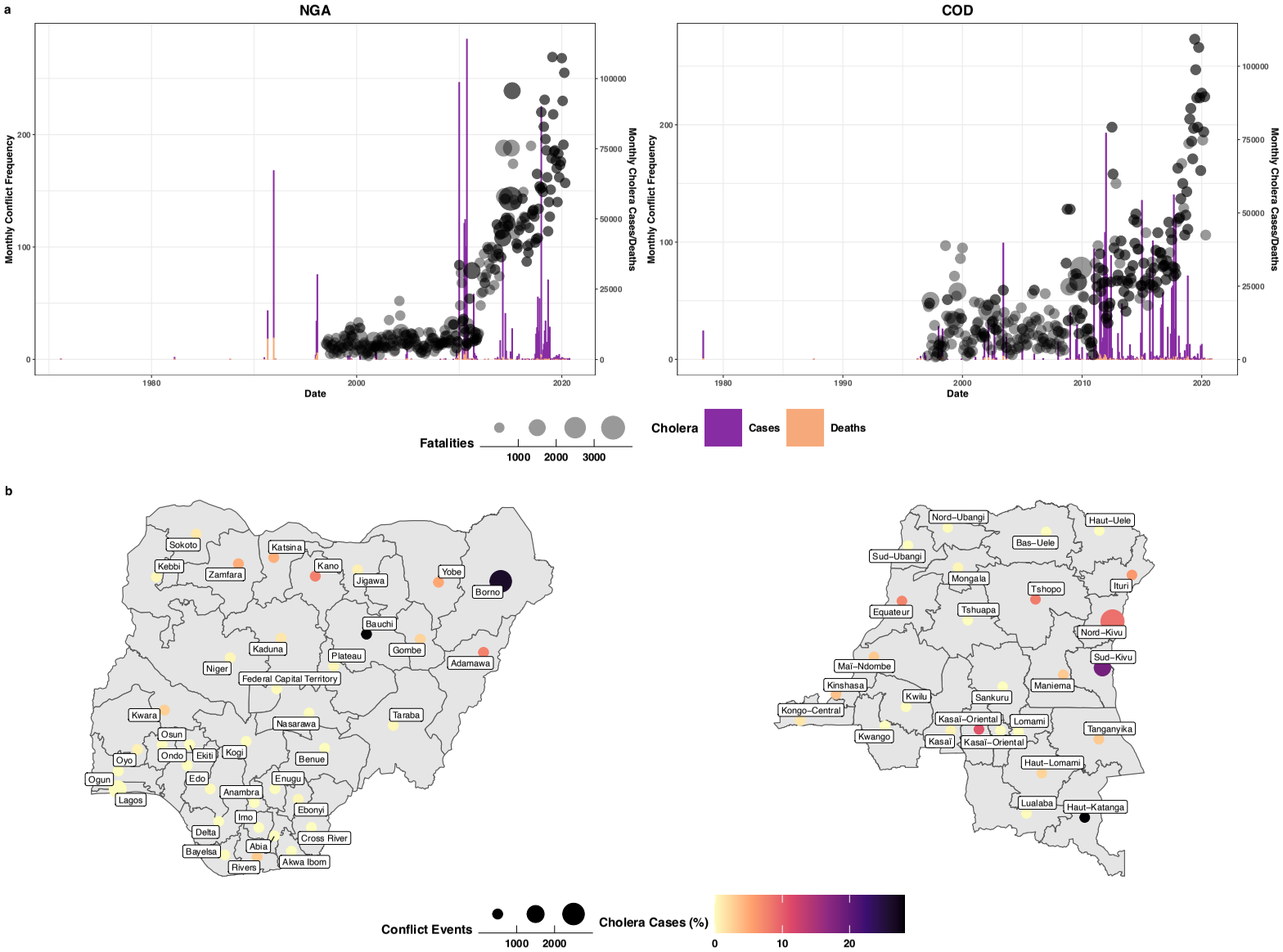


Figure 4.1: Changes in cholera and conflict for the full datasets for Nigeria (left panel) and the Democratic Republic of Congo (right panel). **a**, monthly cholera cases and deaths and monthly frequency of conflict events and fatalities and **b** the number of conflict events and cholera cases as a percentage of the total number of national cases by administrative level 1.

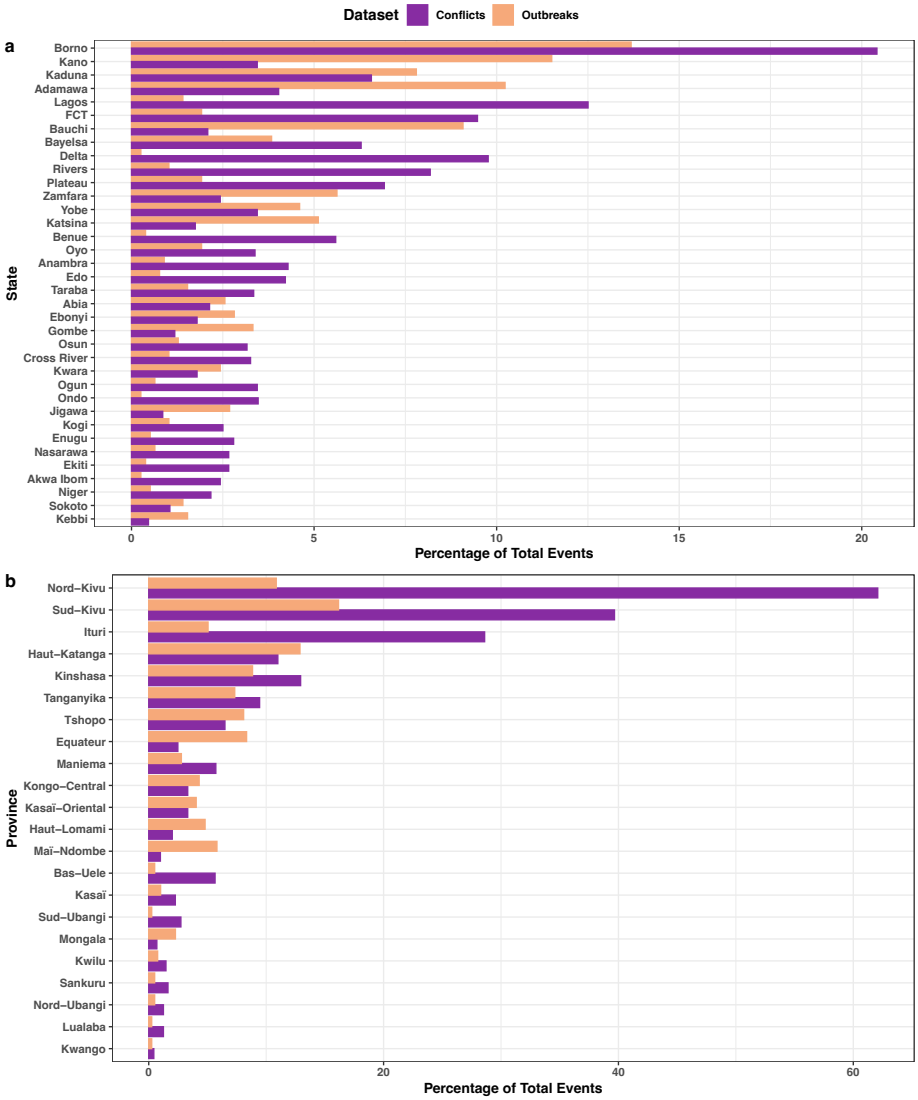


Figure 4.2: Percentage of events (conflicts and cholera outbreaks) in each dataset, over the instrumental period (1997-2020), for **a**, Nigeria and **b**, the Democratic Republic of Congo by administrative level 1. FCT - Federal Capital Territory.

To be included in the analysis, the state/province had to report both outbreaks and conflicts in the datasets during the study period. No states/provinces reported only outbreaks in the observation period. As such, 22 provinces were included for the DRC and 36 states for Nigeria. States and provinces excluded were (and the number of conflicts removed) Imo (239 conflicts) for Nigeria and for the DRC, Haut-Uélé (629 conflicts), Kasaï-Central (234 conflicts), Lomani (101 conflicts) and Tshuapa (70 conflicts). The temporal distribution of the exposure periods and outbreaks fit to the model for each state/province are shown in Figure 4.3.

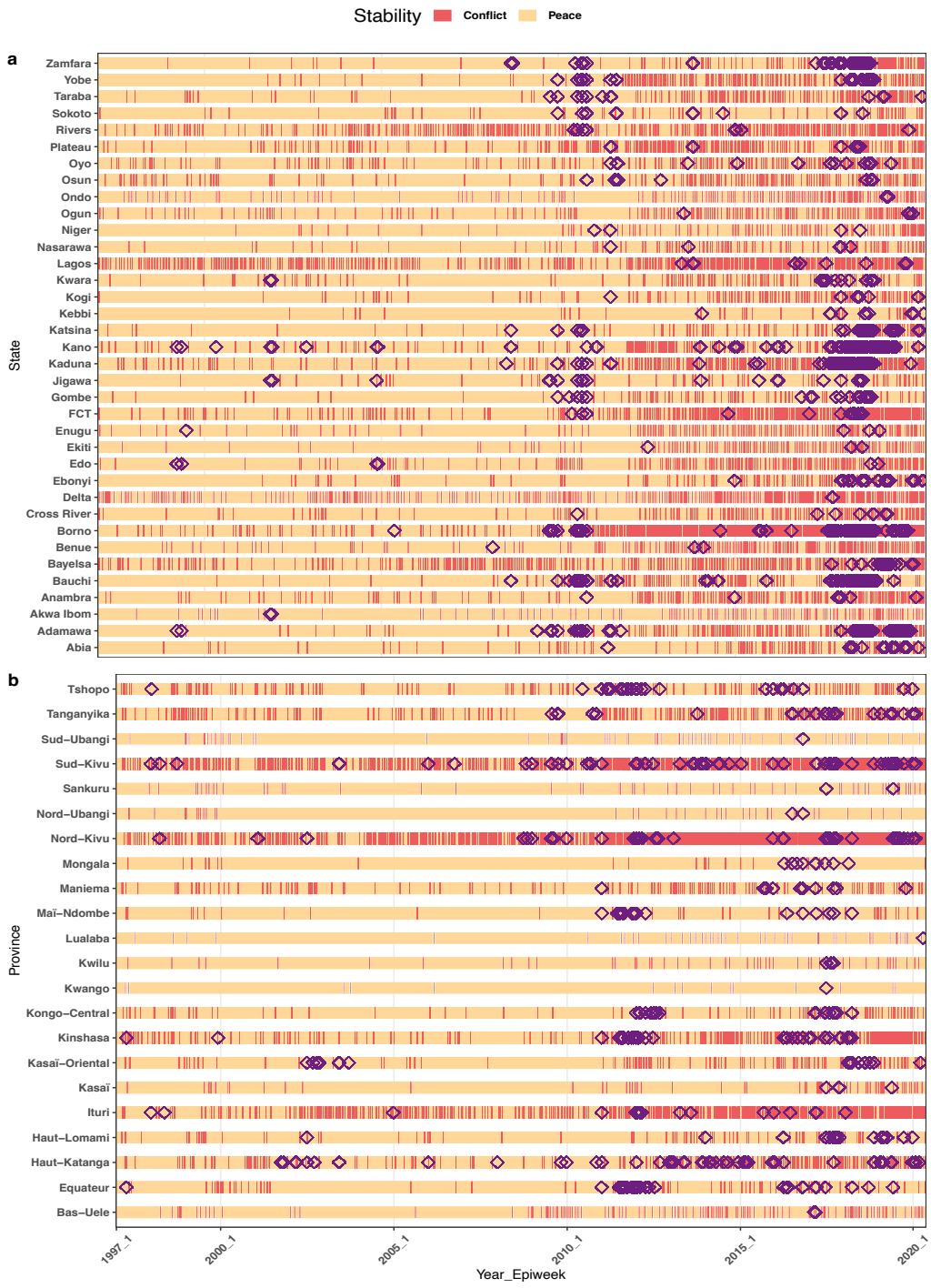


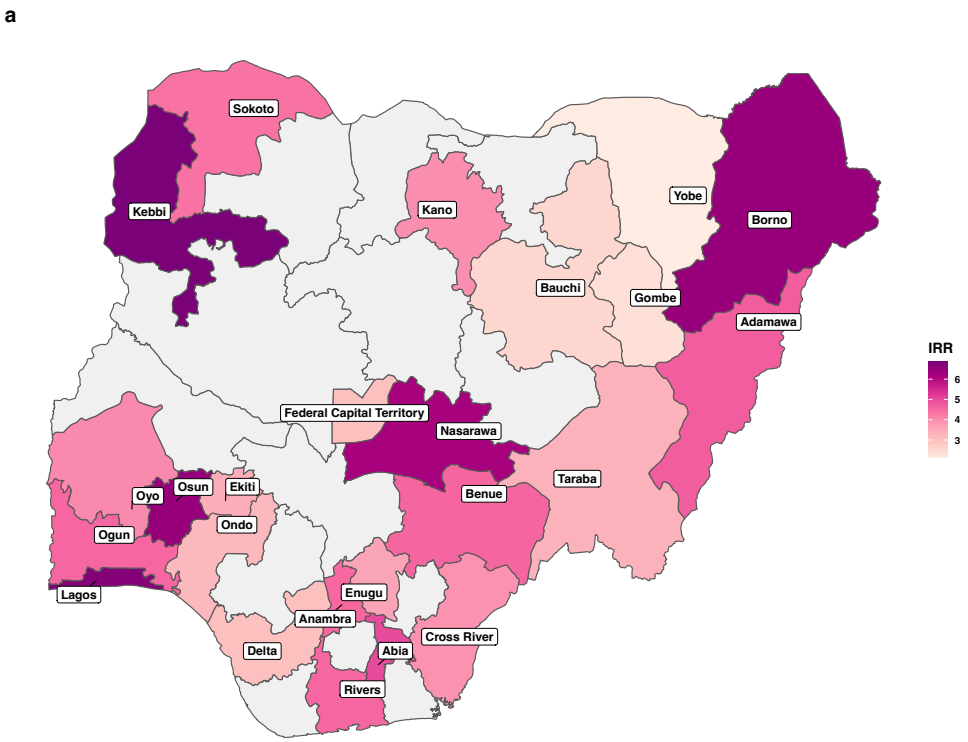
Figure 4.3: Swimmer plots showing the conflict exposure period in the SCCS model (1 week after the onset) and the outbreaks (purple diamonds) for each state/province for **a**, Nigeria and **b**, the Democratic Republic of Congo.

4.3.2 Model Output

Conflict significantly increased the risk of cholera outbreaks (according to IRR) in the past 23 years in Nigeria and the DRC. Nigeria showed an effect of greater magnitude, increasing the risk of cholera outbreaks by 3.6 times (IRR = 3.6, 95%CI = 3.3-3.9). Whereas for the DRC, the risk was increased by 2.6 times (IRR = 2.6, 95%CI = 2.3-2.9).

Of the 36 Nigerian states included in the analysis, 24 showed significant associations between conflict and cholera outbreaks. The strongest effect was found in Kebbi, Lagos, Osun, Borno and Nasarawa, with IRR values ranging from 6.9 to 6.2 (Figure 4.4a).

Ten out of 22 DRC provinces included in the analysis showed a significant relationship between conflict and cholera. Tanganyika, Kasaï-Oriental, Maniema, Nord-Kivu and Kasaï found the strongest values and some were the highest values found in the analysis. In Tanganyika, conflict increased cholera risk by 7.5 times and 3.7 times for Kasaï (Figure 4.4b).



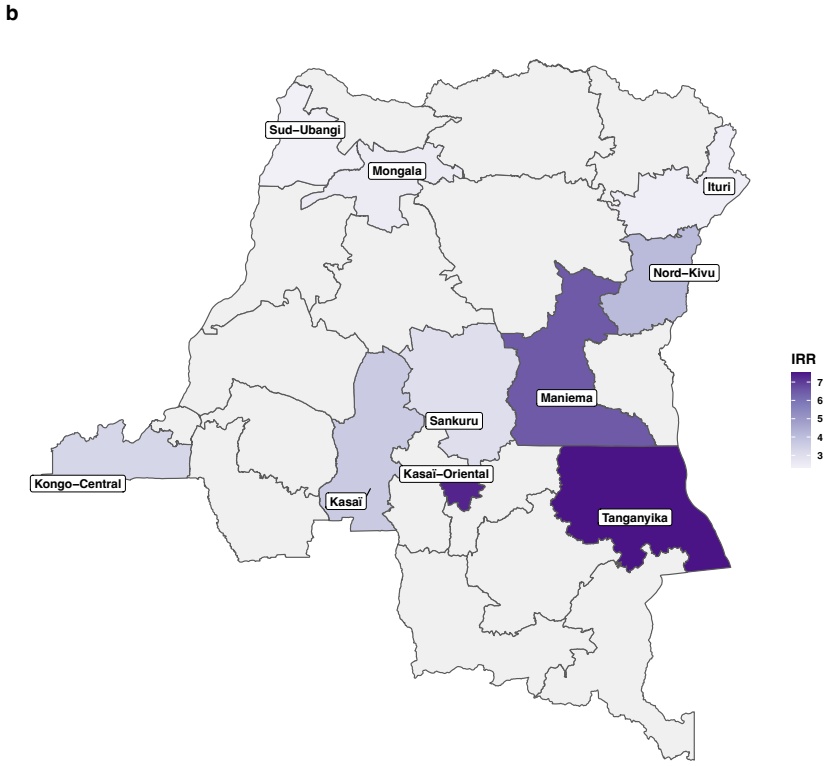


Figure 4.4: Incidence rate ratio (IRR) for the effect of exposure to conflict within one week of the event and cholera at a sub-national level. For **a**, Nigeria and **b**, the Democratic Republic of Congo. Only results that were significant at the threshold p <0.05 are plotted here and labelled.

4.3.3 Sensitivity Analysis

The effect of conflict on cholera outbreaks at a national and sub-national level for both Nigeria and the DRC decreased with increasing exposure period. By week 6 the change was minimal and plateaued or increased. From week 1 to week 10 the risk decreased from 3.6 to 2.08 for Nigeria and from 2.6 to 1.5 for the DRC (Figure 4.5). This suggests that the risk of conflict on cholera is highest soon after the event but a detectable association remains, albeit at a lower level for potentially a long period of time after the event.

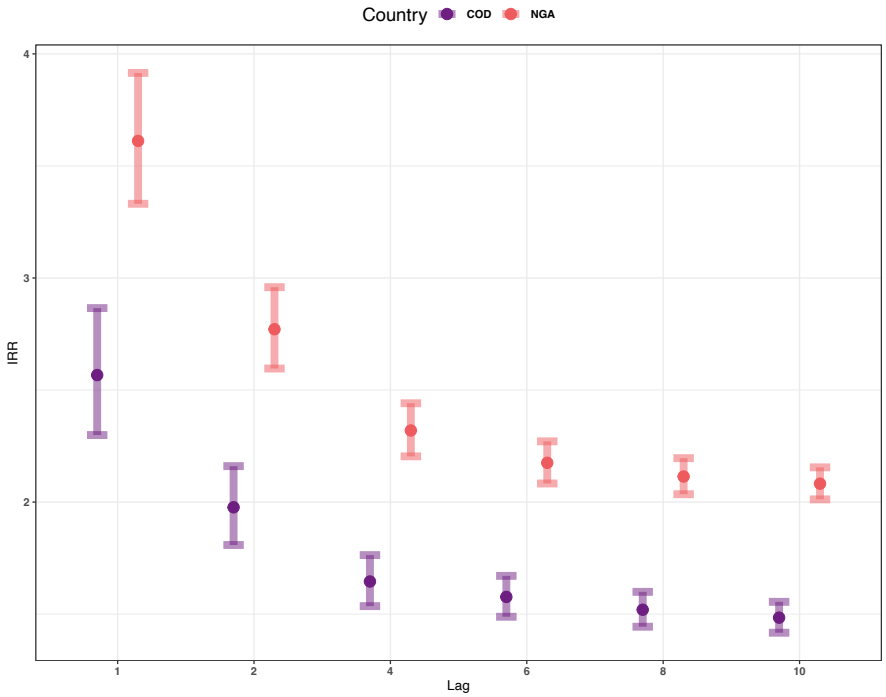


Figure 4.5: Incidence rate ratio (IRR) with 95% confidence intervals for the national sensitivity analysis. The points show the effect of exposure to conflict within 1, 2, 4, 6, 8 and 10 weeks of the event and cholera for Nigeria (NGA) and the Democratic Republic of Congo (COD).

Thirty Nigerian states and 13 DRC provinces were found to be significant for at least one of the lag periods. The most significant states predominately followed the trends of the national analysis, whereas some increased after 6 weeks following the conflict event (Kano & Nassarawa in Nigeria and Tanganyika & Maniema in the DRC). The magnitude of change varied by state, Kebbi saw a decrease in IRR from 6.9 (for the original 1 week exposure period) to 4.0 (for a lag of 10 weeks), whereas Gombe's risk decreased from 2.4 to 1.5 (from 1 week to 10 weeks lag). The IRR values for the sub-national sensitivity analysis are presented in Figure 4.6.

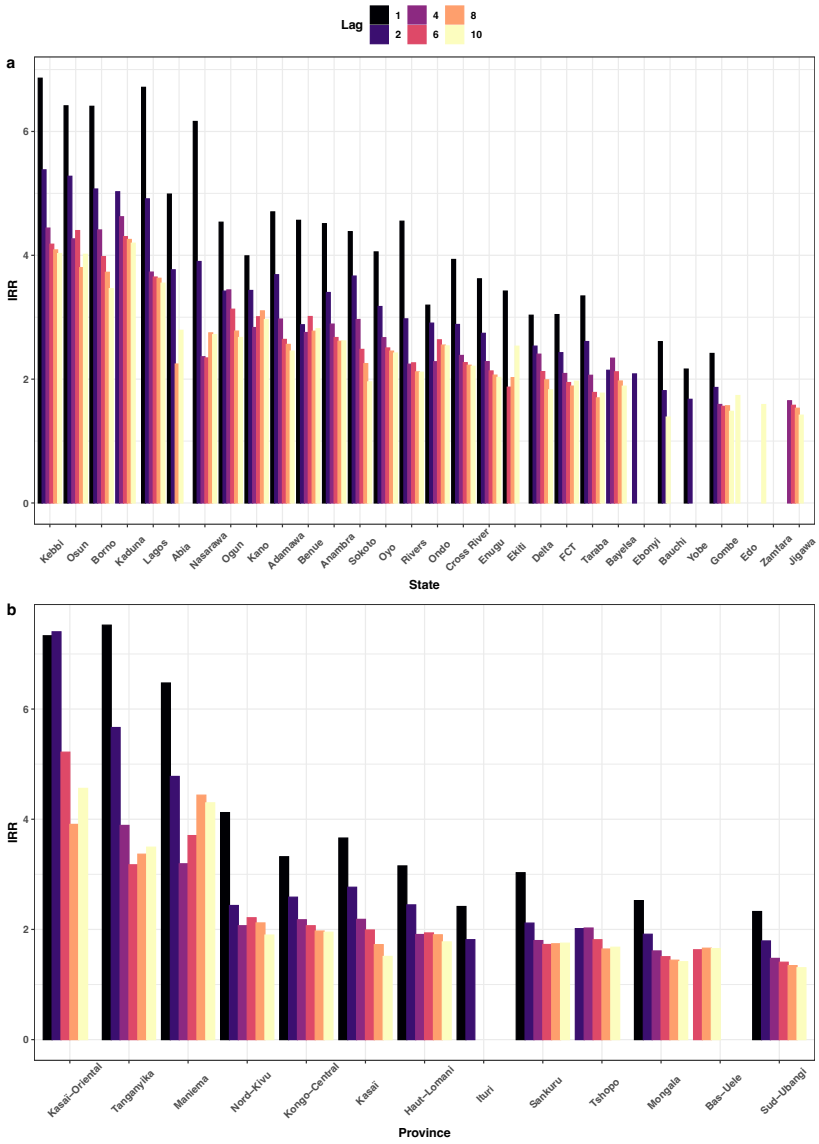


Figure 4.6: Incidence rate ratio (IRR) for the sub-national sensitivity analysis. The bars show the effect of exposure to conflict within 1, 2, 4, 6, 8 and 10 weeks of the event and cholera at administrative level 1. For **a**, Nigeria and **b**, the Democratic Republic of Congo. Only results that were significant at the threshold p <0.05 are plotted here.

Changing the outbreak onset definition yielded similar results to the original analysis. Removing events within both 2 weeks (Scenario 1) and 6 months (Scenario 2) of each other found IRR values within the confidence interval of the initial definition (Figure 4.7). All results remained significant at $p < 0.05$ and provides evidence that temporal autocorrelation did not impact model robustness. How removing these events (cholera outbreaks) affected the data are presented in additional swimmer plots in Supplementary Figures 4.4 & 4.5.

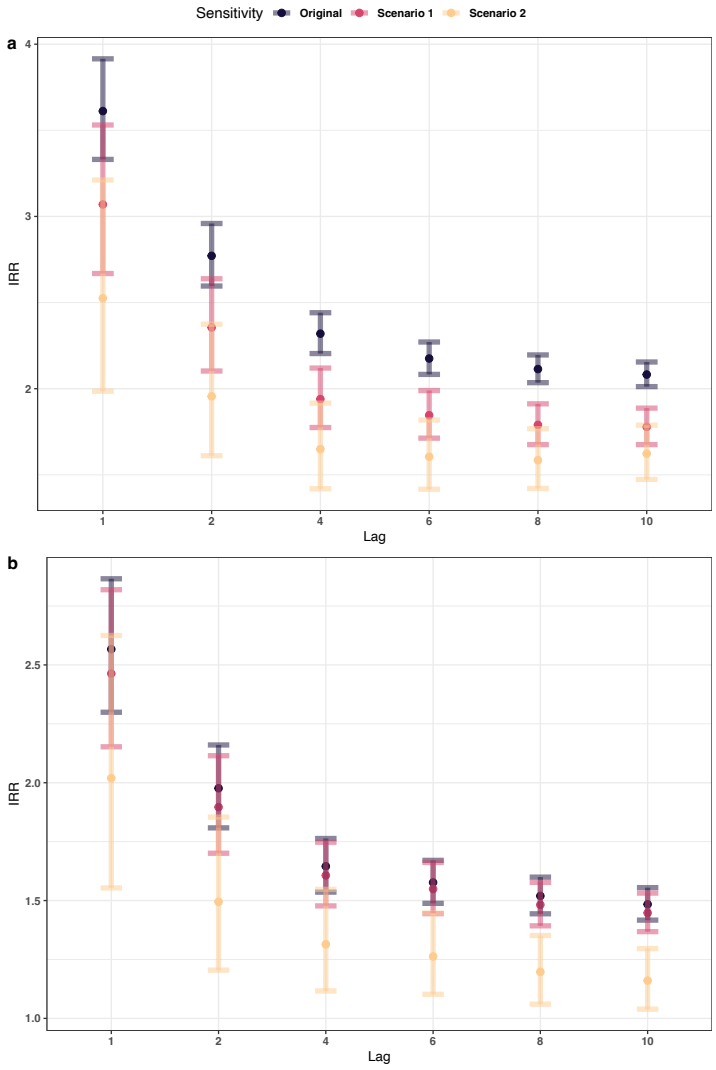


Figure 4.7: Results of the outbreak definition sensitivity analysis. Incidence rate ratio (IRR) values and 95% confidence interval for **a**, Nigeria and **b**, the Democratic Republic of Congo for Scenario 1 (only events > 2 weeks apart) and Scenario 2 (only events > 6 months apart). Both alternative scenarios are compared against the "Original" analysis using all weeks with outbreaks.

4.3.4 Percentage Attributable Fraction

Based on the randomly resampled IRR values (1,000 samples) from the model results (3.6 for Nigeria and 2.6 for the DRC), the onset of a conflict event during epidemiological week 1 in 1997 to week 20 in 2020, was attributable to 19.7% (95%CI = 18.2%-21.2%) of cholera outbreaks in Nigeria and 12.3% (95%CI = 10.2%-14.4%) for the DRC.

4.4 Discussion

Conflict events significantly increased the risk of cholera outbreaks by 3.6 times in Nigeria and 2.6 times in the DRC. The percentage of cholera outbreaks attributable to the conflicts reported here was 19.7% for Nigeria and 12.3% for the DRC. The states/provinces with the highest increased risk were Kebbi, Nigeria at 6.9 times and Kasaï-Oriental, the DRC at 7.3 times. This finding showed that in some states/provinces, the effect of conflict was much greater than the national level.

States/provinces with the greatest increased risk often coincided with areas of high conflict. This provides further evidence to the hypothesis that conflict may be a driver of cholera in Nigeria and the DRC. States/provinces surrounding high conflict areas were also highly significant (e.g., Abia, Ogun, Osun, Maniema, and Tanganyika). The states/provinces here were studied independently but a possible explanation for this may be a spill-over effect. People may flee areas of conflict or a cholera outbreak to neighbouring states and displacement is a known risk factor for disease outbreaks [40]. Displacement is especially important for cholera, as a large proportion of people will be asymptomatic but can still shed the pathogen into local reservoirs, which displaced populations may use as drinking water due to a lack of alternatives [36].

The sensitivity analysis evaluating the effect of lag showed a decrease in effect as the weeks progressed, with some states and provinces seeing a plateau or increase around 6 weeks after the event. The decrease with the lag duration may be a "diluting" effect, as the probability of an outbreak will increase across a longer period (although offsetting the interval helped account for this, see Supplementary Table 4.1). The states/provinces that increased after week 6, were often those with the strongest initial effect, especially in the DRC. This larger initial effect may have a longer lasting impact, potentially due to conflict severity (e.g., more infrastructure damage), while the IRR values of more than 1 (2.08

Nigeria and 1.5 for the DRC) even at 10 weeks after the conflict suggest the impacts are long lasting. However, states/provinces with a large effect in week 1 also had the greatest magnitude of change through the lag periods, potentially because it is not biologically possible to sustain the increased risk for long periods e.g., the number of susceptible individuals would quickly reduce.

Cholera outbreaks can be explosive and self-limiting, due to the high number of asymptomatic individuals, reducing the susceptible pool [36]. This potentially explains why the impacts of conflict on cholera was seen just one week after the event. The incubation period of cholera is relatively short [37], making the effect within the first week found here biologically possible for the pathogen and a realistic timeframe for elevated exposure to manifest in cases. Other examples of cholera cases emerging within the first week after an adverse event include Cyclone Thane hitting the Bay of Bengal [41], water supply interruption in the DRC [42] and Cyclone Aila in West Bengal [43]. This provides further evidence of the need for quick and effective aid in humanitarian crises to avoid outbreaks and reduce mortality [44].

4.4.1 Cholera/Conflict-related Risk Factors

Healthcare facilities can suffer in periods of conflict and cholera outbreaks can overwhelm systems, a potential cause of the relationship between conflict and cholera. Care can be inaccessible because of direct infrastructure damage or difficulties getting to the facilities due to impromptu roadblocks [45]. Supplies may be stolen and/or unable to be delivered, including ORS, pathogen-sensitive antibiotics and OCV, which are important for cholera outbreak control and mortality [46]. Finally, safety is a serious issue, both for healthcare workers and patients and NGOs can withdraw from these areas, citing an inability to ensure the safety of their staff [47]. Steps need to be taken globally to reduce this violence, such as using active clinical management for all patients to enhance the acceptance of pathogen-specific treatment centres [48].

Conflict has the potential to worsen pre-existing vulnerabilities, which can exacerbate poverty, another potential cause of the effect of conflict on cholera. The impacts of poverty can be far reaching and is a known risk factor for cholera [49, 50], along with other diseases [51]. For example, poor urban settlements have faced the brunt of outbreaks including Zika, Ebola, typhoid, and cholera, due to crowding and poor access to WASH [52]. Those in poorer communities may also have more contacts

and greater transmission, creating a vicious cycle [52]. Conflict can result in loss of possessions, habitual residence, and an inability to find employment, reducing income generation, savings and financial backstops [53]. In times of worsening poverty, people may not be able to afford healthcare and basic medical supplies, especially in vulnerable groups. This disruption to daily life can cause many more deaths than direct battlefield fatalities and leads to stagnation in development [54].

A lack of WASH facilities is likely to have contributed to the positive relationship between cholera and conflict found here. Although WASH and poverty were not directly evaluated, their effects are likely to have been important. Conflict events can lead to disruption in sanitation and hygiene and adverse events can act as catalysts in the interaction of contaminated water and the human population [55]. Displacement from conflict can cause issues in accessing WASH (e.g., latrine access, soap availability) and several displacement camps have seen rapid cholera outbreaks, including the DRC after the Rwandan genocide in 1994 [56]. If people are displaced due to conflict, this may result in the use of water contaminated with toxigenic strains of *Vibrio cholerae* because alternatives are lacking, leading to outbreaks.

4.4.2 Limitations

A potential limitation is the plausible existence of multiple causal pathways, leading to misclassification, due to time-variant confounders. One causal pathway could be a conflict event in an adjacent geographic area being causally linked to the conflict event in the current geographic area. This classification bias would underestimate the effect of conflict on cholera. If a cholera outbreak was "imported" from a neighbouring state/province (spatial autocorrelation), this would be classified as a genuine, autochthonous event, which would likely be non-differential (likely to happen during an exposed or non-exposed period). Additionally, by completing the sub-national analysis and recalculating the IRR, this accounted for spatial autocorrelation within the state/province, but not across the country.

An additional causal pathway for cholera, that may have resulted in residual confounding here, could be the presence of waterbodies. Water is considered fundamental in cholera transmission and large basins are present in both study areas e.g., the Lake Chad basin in Nigeria and the African Great Lakes Region in the DRC [57]. However, no study has yet demonstrated a long-term presence of toxigenic

Vibrio cholerae in African lakes [58]. Understanding these additional environmental factors including seasonal weather changes and the pre-existing vulnerabilities discussed is very important, as these are known to impact disaster-related outbreaks and multi-disaster events [55, 59]. Although beyond the scope of the methods used here, as conflict was investigated in isolation, this will be addressed in Chapter 5.

Under-reporting, over-reporting and a reporting lag may have impacted the degree of effect found in Chapter 4. Under-reporting is a significant issue in global cholera and conflict estimates, due to asymptomatic cases, disincentives to report, logistical issues and the ambiguity of classifying events [28, 60]. Cholera surveillance is difficult in conflicts, due to displaced populations and security issues. Alternatively, during times of conflict health surveillance can be enhanced by the government and/or NGOs (e.g., establishing cholera treatment centres (CTC)). Reporting delay is another potential problem and some national disease reporting delays have been found to range from 12 days for meningococcal disease to 40 days for pertussis [27]. Additionally, ACLED (the source of the conflict data used here) have stated that 17.4% of their published conflict data is delayed or lagged, which ranges by event type, location and actor [25].

States which reported conflict but no cholera outbreaks were removed, as the SCCS model is a case-only approach. Analysing cases only, instead of the corresponding complete cohort, translates into a loss of efficiency (by only analysing a sample of the data), but previous work showed that this loss is small, especially when the fraction of the sample experiencing the exposure is high (92% of states/provinces were included in the analysis). Moreover, the loss of efficiency from this method must be weighed against a better control of time-varying confounders (by fixing via stratification) as discussed above. Previous examples illustrated that the SCCS design is likely to produce more trustful results than the corresponding cohort analysis, especially when a strong residual confounding bias is likely [29, 31].

The severity or intensity of both conflicts and cholera outbreaks were not evaluated here, as binary variables were used. It was also assumed that the impacts of conflict were experienced evenly across the state/province. Conflict severity is complex and far-reaching and despite some national indexes being available [61, 62], these often focus more broadly on disasters and without a universal measure of conflict severity, comparing studies remains challenging. Making assessments of how a conflict event impacts a health outcome is difficult and involve assumptions and oversimplifications. Despite these

difficulties and although beyond the scope of this work, it is an important area of future research. The conflict data used here were categorised and future analysis could assess if these different conflict types impacted the effect on cholera. Qualitative studies with those working in a variety of different organisations in conflict-affected areas, may be the best method to answer these questions.

Despite the limitations of conflict and cholera data, the data used here are to the highest standard currently available and has been used by several other studies, making the research comparable [63, 64, 5]. Creating partnerships with those working on the ground and exploring more sensitive data options is investigated in Chapter 5 [65]. Additional methods to account for data limitations included setting both the event and the exposure to a binary outcome to reduce the impacts of severity and using a weekly instead of daily temporal scale to account for delays. Furthermore, several methods of validating the cholera data were used (e.g., comparing the dataset to other data sources, considering multiple accounts of the same outbreak) [24] and the sensitivity analyses helped highlight the duration of the effect and possible lag effects of the conflict data and the effect of altering the cholera outbreak definition.

4.4.3 Conclusion

In summary, this analysis shows a clear relationship between cholera and conflict in both Nigeria and the DRC at both a national and sub-national level. Conflict increased the risk of cholera outbreaks by 7.3 times in some states/provinces and almost 20% of cholera outbreaks were attributable to conflict in Nigeria. The effect of conflict on cholera appears have a rapid and severe onset but should not be considered acute, as it may have a long-lasting effect, if not at a much lower level. This finding potentially holds in other countries and diseases and highlights how the SCCS methodology could be used in different contexts.

Cholera risks are likely multi-factorial in both northeastern Nigeria and eastern DRC and several conditions need to be met for emergencies to lead to cholera outbreaks. Sufficient and rapid support, along with enhanced efforts to build community trust can reduce this increased risk. Finding conflict resolution should be the main priority in fragile states and pre-existing vulnerabilities need to be addressed, such as poverty, expansion of affordable healthcare and improvements in WASH. By reducing these vulnerabilities, communities will have greater resources to adapt to social extremes and could

help to reduce vulnerabilities both in times of conflict and peace.

It is important to try and understand the complexities of the multiple risk factors involved in conflict-related (and natural hazard-related) cholera outbreaks, incorporating as many as possible into modelling studies. Chapter 5 will build on the methodologies from Chapters 3 and 4, using a data source at a finer spatial scale. A continuous, rather than a binary outcome variable will be used to quantify severity and evaluating multiple extremes together, rather than in isolation. Multiple hazard events can occur, and the presence of a current extreme does not prevent further disasters from occurring. Despite this, few studies have investigated the impacts of multiple disasters and even less in the context of pre-existing vulnerabilities. In collaboration with the Nigeria Centre for Disease Control, Chapter 5 aims to address both these research gaps and build on the previous methodological limitations, creating a framework for cholera outbreak risk in fragile settings.

References

[1] F.A. Dureab et al. "Yemen: cholera outbreak and the ongoing armed conflict". en. In: *J. Infect. Dev. Countr.* 12 (2018), pp. 397–403.

[2] C.C. Blackburn, P.E. Lenze, and R.P. Casey. "Conflict and cholera: Yemen's man-made public health crisis and the global implications of weaponizing health". en. In: *Health Secur.* 18 (2020), pp. 125–31.

[3] F. Federspiel and M. Ali. "The cholera outbreak in Yemen: lessons learned and way forward". en. In: *BMC Public Health* 18 (2018), pp. 1–8.

[4] R. Sato. "Effect of armed conflict on vaccination: evidence from the Boko haram insurgency in northeastern Nigeria". en. In: *Confl. Health* 13 (2019), pp. 1–.

[5] C.R. Wells. "The exacerbation of Ebola outbreaks by conflict in the Democratic Republic of the Congo". en. In: *PNAS* 116 (2019), pp. 24366–24372.

[6] S.J. Tijjani and L. Ma. "Is Nigeria prepared and ready to respond to the COVID-19 pandemic in its conflict-affected northeastern states?" en. In: *Int. J. Equity Health* 19 (2020), pp. 1–4.

[7] D. Agbiboa. “The ongoing campaign of terror in Nigeria: Boko Haram versus the state”. en. In: *Stability* (2013), p. 2.

[8] Council Foreign Relations. *Violence in the Democratic Republic of Congo*. en. URL: https://www.cfr.org/global-conflict-tracker/conflict/violence-democratic-republic-congo.

[9] M. Ali et al. “Updated global burden of cholera in endemic countries”. en. In: *PLoS Neglect. Trop. Dis.* 9 (2015), p. 0003832.

[10] D. Bompangue. “Cholera epidemics, war and disasters around Goma and Lake Kivu: an eight-year survey”. en. In: *PLoS Neglect. Trop.* D. 3 (2009).

[11] United Nations Statistical Division. *Millennium Development Goal Indicators*. en. 2015. URL: https://unstats.un.org/unsd/mdg/SeriesDetail.aspx?srid=580.

[12] C.P. Farrington. “Relative incidence estimation from case series for vaccine safety evaluation”. en. In: *Biometrics* (1995), pp. 228–35.

[13] R. Brauer et al. “Antipsychotic drugs and risks of myocardial infarction: a self-controlled case series study”. en. In: *Eur. Heart J.* 36 (2015), pp. 984–92.

[14] I.J. Douglas et al. “The risk of fractures associated with thiazolidinediones: a self-controlled case-series study”. en. In: *PLoS Med.* 6 (2009), p. 1000154.

[15] K. Jean et al. “Assessing the impact of preventive mass vaccination campaigns on yellow fever outbreaks in Africa: A population-level self-controlled case series study”. en. In: *PLoS Med.* 18 (2021), p. 1003523.

[16] World Health Organization. *Disease Outbreak News (DONs)*. en. 2021. URL: https://www.who.int/csr/don/en/.

[17] ProMED. *Search ProMED Posts*. en. 2021. URL: https://promedmail.org/promed-posts.

[18] World Health Organizations Regional Office for Africa. *Weekly bulletins on outbreaks and other emergencies*. en. 2021. URL: https://www.afro.who.int/health-topics/disease-outbreaks/outbreaks-and-other-emergencies-updates.

[19] Cholera Platform. *Regional Updates (Cholera Bulletin)*. en. 2021. URL: http://plateformecholera.info.

[20] CRED/UCLouvain (D. Guha-Sapir) EM-DAT. *EM-DAT Public*. en. 2020. URL: https://public.emdat.be.

[21] Nigeria Centre for Disease Control. *Weekly Epidemiological Report*. en. 2021. URL: https://ncdc.gov.ng/reports/weekly.

[22] Nigeria Centre for Disease Control. *NIGERIA - PREPAREDNESS AND RESPONSE TO ACUTE WATERY DIARRHOEA OUTBREAKS*. en. 2017. URL: http://www.plateformecholera.info/index.php/country-monitoring/nigeria/152-wca/strategic-framework/library-of-national-plans/481-nigeria-preparedness-and-response-to-acute-watery-diarrhoea-outbreaks.

[23] RDC Ministère de la Santé Publique. *Situation épidémiologique du choléra en République Démocratique du Congo en 2011*. 2012. URL: https://reliefweb.int/sites/reliefweb.int/files/resources/RAPPORT%5C%20SUR%5C%20LA%5C%20SITUATION%5C%20DU%5C%20CHOLERA%5C%20EN%5C%20RDC%5C%20EN%5C%202011%5C%20DB-30_06_12.pdf.

[24] G.E.C. Charnley et al. "Accessing sub-national cholera epidemiological data for Nigeria and the Democratic Republic of Congo during the seventh pandemic". en. In: *BMC Infect. Dis.* 22 (2022), pp. 1–7.

[25] H.D.X. *The Humanitarian Data Exchange*. en. 2021. URL: https://data.humdata.org.

[26] R.A. Jajosky and S.L. Groseclose. "Evaluation of reporting timeliness of public health surveillance systems for infectious diseases". en. In: *BMC Public Health* 4 (2004), pp. 1–9.

[27] S. Ri et al. "Attacks on healthcare facilities as an indicator of violence against civilians in Syria: An exploratory analysis of open-source data". en. In: *PloS ONE* 14 (2019), p. 0217905.

[28] N.B. Weidmann. "A closer look at reporting bias in conflict event data". en. In: *Am. J. Pol. Sci.* 60 (2016), pp. 206–18.

[29] I. Petersen, I. Douglas, and H. Whitaker. "Self controlled case series methods: an alternative to standard epidemiological study designs". en. In: *BMJ* (2016), p. 354.

[30] C.P. Farrington, H.J. Whitaker, and M.N. Hocine. "Case series analysis for censored, perturbed, or curtailed post-event exposures". en. In: *Biostatistics* 10 (2009), pp. 3–16.

[31] H.J. Whitaker et al. "Tutorial in biostatistics: the self-controlled case series method". en. In: *Stat. Med.* 25 (2006), pp. 1768–97.

[32] T.M. Therneau. *survival: A Package for Survival Analysis in R*. en. 2021. URL: https://CRAN.R-project.org/package=survival.

[33] H. Turner and D. Firth. *gnm: Generalized nonlinear models in R: An overview of the gnm package*. R package version 1.1-2. 2022. URL: https://cran.r-project.org/package=gnm.

[34] E. Mostofsky, B.A. Coull, and M.A. Mittleman. "Analysis of observational self-matched data to examine acute triggers of outcome events with abrupt onset". In: *Epidemiol.* 29.6 (2018), p. 804.

[35] S. Xu et al. "Use of fixed effects models to analyze self-controlled case series data in vaccine safety studies". In: *J. Biom. Biostat.* (2012), p. 006.

[36] A.A. King et al. "Inapparent infections and cholera dynamics". en. In: *Nature* 454 (2008), pp. 877–80.

[37] A.S. Azman et al. "The incubation period of cholera: a systematic review". en. In: *J. Infect.* 66 (2013), pp. 432–8.

[38] Hadley W. et al. *dplyr: A Grammar of Data Manipulation*. R package version 1.0.7. 2021. URL: https://CRAN.R-project.org/package=dplyr.

[39] Tim H. *resample: Resampling Functions*. R package version 0.6. 2022. URL: https://CRAN.R-project.org/package=resample.

[40] J.T. Watson, M. Gayer, and M.A. Connolly. "Epidemics after natural disasters". en. In: *Emerg. Infect. Dis.* 13, 1 (2007).

[41] T. Fredrick et al. "Cholera outbreak linked with lack of safe water supply following a tropical cyclone in Pondicherry". en. In: *J. Health Popul. Nutr.* 33 (2012), p. 31.

[42] A. Jeandron et al. “Water supply interruptions and suspected cholera incidence: a time-series regression in the Democratic Republic of the Congo”. en. In: *PLoS Med.* 12 (2015), p. 1001893.

[43] R. Bhunia and S. Ghosh. “Waterborne cholera outbreak following cyclone Aila in Sundarban area of West Bengal, India, 2009”. en. In: *Trans. R. Soc. Trop.* 105 (2011), pp. 214–9.

[44] R.V. Tauxe et al. “Epidemic cholera in Mali: high mortality and multiple routes of transmission in a famine area”. en. In: *Epidemiol. Infect.* 100.2 (1988), pp. 279–289.

[45] C. Sousa and A. Hagopian. “Conflict, health care and professional perseverance: a qualitative study in the West Bank”. en. In: *Glob. Public Health* 6 (2011), pp. 520–33.

[46] E.J. Cartwright et al. “Recurrent epidemic cholera with high mortality in Cameroon: persistent challenges 40 years into the seventh pandemic”. en. In: *Epidemiol. Infect.* 141 (2013), pp. 2083–93.

[47] M.S.F. DRC. *Violent attacks against staff force MSF to end projects in Fizi territory, South Kivu*. en. URL: https://www.msf.org/msf-forced-pull-out-eastern-drc-territory-following-violent-attacks.

[48] V.K. Nguyen. “An epidemic of suspicion—Ebola and violence in the DRC”. en. In: *New Engl. J. Med.* 380 (2019), pp. 1298–9.

[49] A. Talavera and E.M. Perez. “Is cholera disease associated with poverty?” en. In: *J. Infect. Dev. Ctries.* 3.06 (2009), pp. 408–411.

[50] K. Penrose et al. “Informal urban settlements and cholera risk in Dar es Salaam, Tanzania”. et. In: *PLoS. Neglect. Trop. Dis.* 4.3 (2010).

[51] M.P. Fallah et al. “Quantifying poverty as a driver of Ebola transmission”. en. In: *PLoS Neglect. Trop. Dis.* 9 (2015), p. 0004260.

[52] M. Eisenstein. “Disease: poverty and pathogens”. no. In: *Nature* 531 (2016), pp. 61–3.

[53] O.C. Okunlola and I.G. Okafor. “Conflict–Poverty Relationship in Africa: A Disaggregated Approach”. en. In: *J. Interdiscip. Econ.* (2020), pp. 1–26.

[54] J.F. Trani et al. “Poverty, vulnerability, and provision of healthcare in Afghanistan”. en. In: *Soc. Sci. Med.* 70 (2010), pp. 1745–55.

[55] A. Jutla, R. Khan, and R. Colwell. “Natural disasters and cholera outbreaks: current understanding and future outlook”. nl. In: *Curr. Environ. Health. Rep.* 4.1 (2017), pp. 99–107.

[56] D.B. Nkoko et al. “Dynamics of cholera outbreaks in Great Lakes region of Africa, 1978–2008”. en. In: *Emerg. Infect. Dis.* 17.11 (2011).

[57] M.E. Birmingham et al. “Epidemic cholera in Burundi: patterns of transmission in the Great Rift Valley Lake region”. en. In: *Lancet* 349 (1997), pp. 981–5.

[58] D. Bompangue et al. “Lakes as source of cholera outbreaks, Democratic Republic of Congo”. pt. In: *Emerg. Infect. Dis.* 14 (2008), p. 798.

[59] G.E.C Charnley et al. “Traits and risk factors of post-disaster infectious disease outbreaks: a systematic review”. en. In: *Sci. Rep.* 11 (2021), pp. 1–4.

[60] K.O. Elimian et al. “Descriptive epidemiology of cholera outbreak in Nigeria, January–November, 2018: implications for the global roadmap strategy”. en. In: *BMC Public Health* 19 (2019), pp. 1–1.

[61] Notre Dame Global Adaptation Initiative. *ND-GAIN Country Index*. en. 2019. URL: https://gain.nd.edu/our-work/country-index/.

[62] European Commission. *INFORM Severity*. en. 2022. URL: https://drmkc.jrc.ec.europa.eu/inform-index.

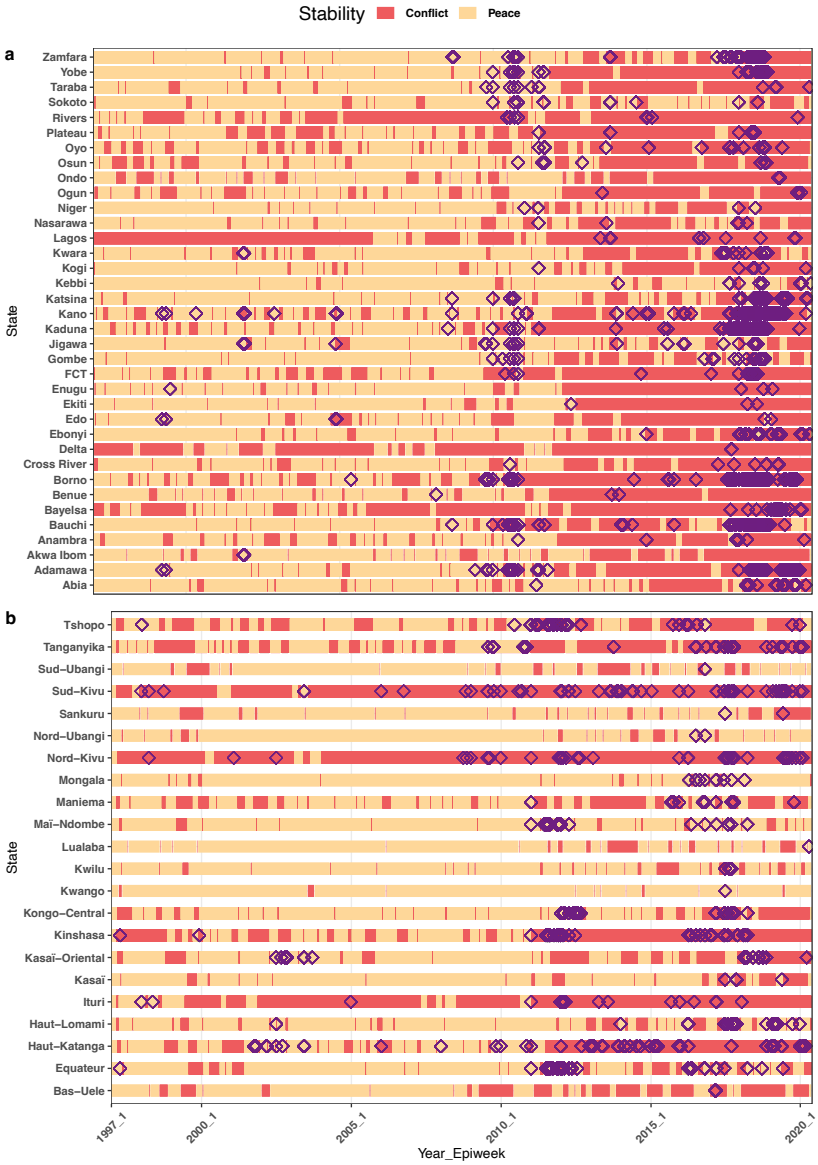
[63] M.U. Kraemer et al. “Dynamics of conflict during the Ebola outbreak in the Democratic Republic of the Congo 2018–2019”. en. In: *BMC Med.* 18 (2020), pp. 1–.

[64] M. Gayer et al. “Conflict and emerging infectious diseases”. en. In: *Emerg. Infect. Dis.* 13 (2007), p. 1625.

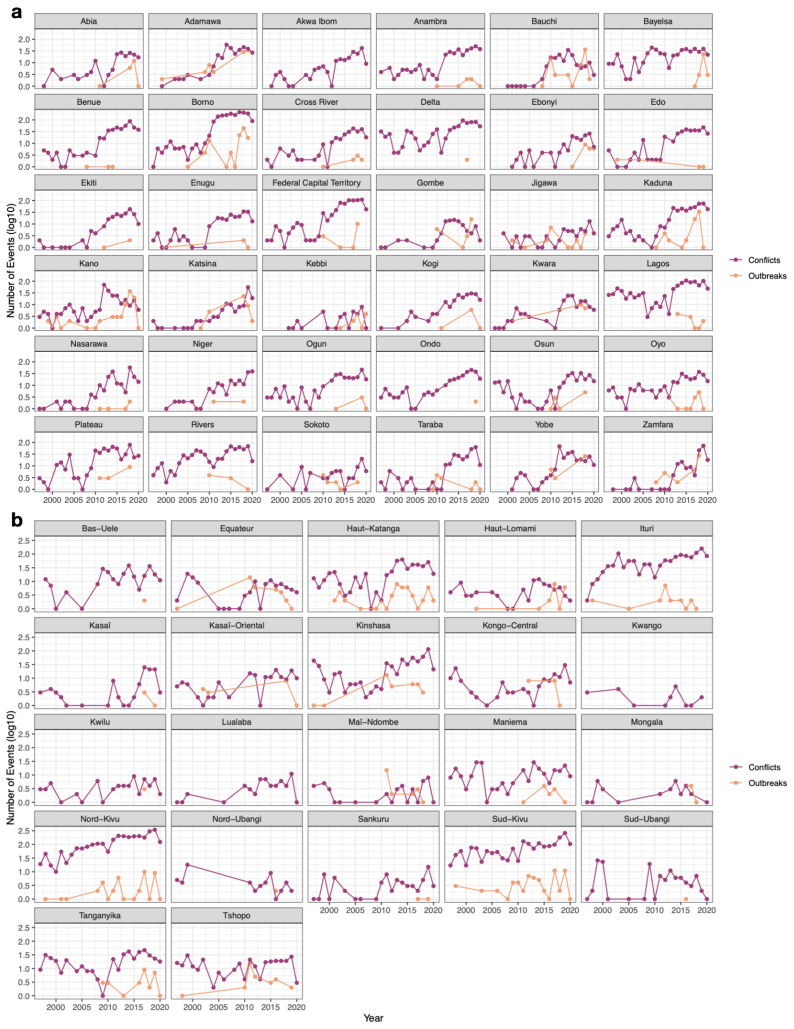
[65] G.E.C. Charnley et al. “Investigating the impact of social and environmental extremes on cholera time varying reproductive number in Nigeria”. en. In: *PLoS Glob. Public Health* 2 (2022), e0000869.

Supplementary Material

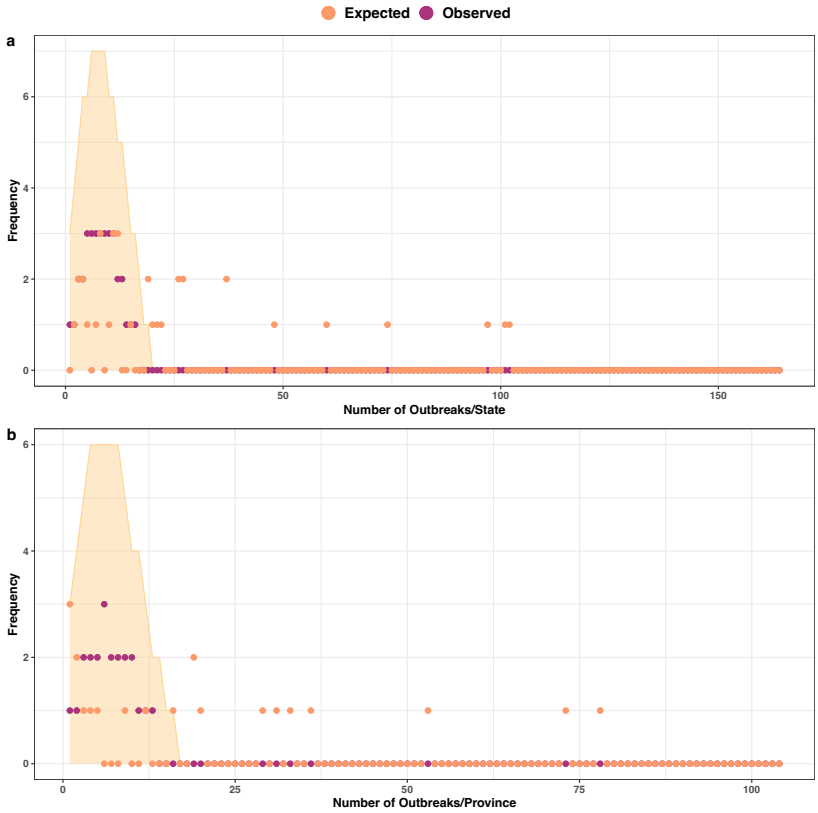
Supplementary Figures



Supplementary Figure 4.1: Swimmer plots showing the conflict dataset for lag 10 in the sensitivity analysis. In relation to outbreaks (purple diamonds) for **a** Nigeria and **b** the Democratic Republic of Congo. FCT - Federal Capital Territory.



Supplementary Figure 4.2: Number of outbreak (orange) and conflict (purple) events by year in **a** Nigeria and **b** the Democratic Republic of Congo over the full study period.



Supplementary Figure 4.3: Poisson probability distribution fit to the outbreak data. The simulated counts were obtained from 10,000 random realizations of a Poisson process of rate λ = number of total national outbreaks/number of states or provinces, for **a**, Nigeria and **b**, the Democratic Republic of Congo. Expected values are the median simulated counts from the distribution with a 95% confidence interval.

$$\lambda_{NGA} = \frac{782}{38} = 20.59$$

$$\lambda_{COD} = \frac{396}{26} = 15.23$$

R function "rpois(n,λ)", with n representing the number of states, was then used to generate Poisson random variables, which could be plotted in Sup. Fig. 4.3 against the data fitted to the model.

Supplementary Figure 4.4: Swimmer plots showing the effect of the outbreak definition sensitivity analysis on distribution of outbreaks and conflicts. Scenario 1, removing all outbreaks < 2 weeks apart, is presented here for **a** Nigeria and **b** the Democratic Republic of Congo.

Supplementary Figure 4.5: Swimmer plots showing the effect of the outbreak definition sensitivity analysis on distribution of outbreaks and conflicts. Scenario 2, removing all outbreaks < 6 months apart, is presented here for **a** Nigeria and **b** the Democratic Republic of Congo.

Supplementary Tables

Supplementary Table 4.1: The layout of the pseudo-dataset dataframe fitted to the model. Each event and exposure are given a reference number (indiv).

indiv	exday	eventday	start	end	event	exgr	interval	loginterval
1	3	374	1	3	0	0	2	0.69314718
1	3	374	3	4	0	1	1	0
1	3	374	4	542	1	0	538	6.287859
2	4	374	1	4	0	0	3	1.09861229
2	4	374	4	5	0	1	1	0
2	4	374	5	542	1	0	537	6.285998

The data (datLong) was fit to the model as follows: clogit(event $\sim$ exgr + strata(indiv) + offset(loginterval), data = datLong).

The data set up follows the work of Heather Whittaker, further code and examples are available at: http://stats-www.open.ac.uk/sccs/r.htm. The data are based on the examples related to multiple risk periods. The aim is to evaluate the likelihood of: event = 1 and exgr = 1, vs event = 1 and exgr = 0. A pre and post exposure period are included to account for the possibility that the event could increase or decrease the probability of an exposure and because exposures can occur after the event. The interval is set up as an offset to account for that fact that a longer interval would increase the chances of the event occurring within it, not because the exposure increased the event but because there was a greater period of time for it to occur by chance.

Additional explanations of these assumptions are available at:

1. Petersen I, et al. Self controlled case series methods: an alternative to standard epidemiological study designs. BMJ 2016;354.

2. Farrington CP, et al. Case series analysis for censored, perturbed, or curtailed post-event exposures. Biostatistics 2009;10(1):3-16.

Supplementary Information

Supplementary Information 4.1: Literature included in the cholera datasets for Chapter 4 (fitted to the SCCS models).

Democratic Republic of Congo

1. Ingelbeen B. et al. Recurrent cholera outbreaks, Democratic Republic of the Congo, 2008–2017. Emerg Infect Dis 2019;25(5):856.
2. Kelvin A A. Cholera outbreak in the Republic of Congo, the Democratic Republic of Congo, and cholera worldwide. J Infect Dev Ctries 2011;5:137-143.
3. D'Mello-Guyett L. et al Distribution of hygiene kits during a cholera outbreak in Kasaï-Oriental, Democratic Republic of Congo: a process evaluation. Conf Health 2020;14(1):1-17.
4. Dongmo NT. et al. Cholera Outbreak in the Kasaï Oriental Province, Democratic Republic of Congo (DRC), 2018: The Case Fatality Rate in Mbuji-Mayi City and in the Rural Area. Science 2020;8(1):8-18.
5. Bompangue D. et al. Cholera ante portas–The re-emergence of cholera in Kinshasa after a ten-year hiatus. PLoS Curr 2012;4.
6. Schyns C. et al. Cholera in Eastern Zaire, 1978. Ann Soc Belg Med Trop 1979;59(4):391-400.

Nigeria

1. Fatiregun A. Epidemiology of an outbreak of cholera in a south-west state of Nigeria. South Afr J Epidemiol Infect 2012;27(4):201-204.
2. Elimian KO. et al. Descriptive epidemiology of cholera outbreak in Nigeria, January–November, 2018: implications for the global roadmap strategy. BMC Public Health 2019;19(1):1264.
3. Dalha, MM. et al. Descriptive characterization of the 2010 cholera outbreak in Nigeria. BMC Public Health 2014;14(1):1167.
4. Umoh JU. et al. Epidemiological features of an outbreak of gastroenteritis/cholera in Katsina, Northern Nigeria. Epidemiol Infect 1983;91(1):101-111.

5. Hutin Y. et al. A large cholera outbreak in Kano City, Nigeria: the importance of hand washing with soap and the danger of street-vended water. J Water Health 2003;1(1):45-52.

6. Sule IB. et al. Descriptive epidemiology of a cholera outbreak in Kaduna State, Northwest Nigeria, 2014. Pan Afr Med J 2017;27.

7. Nnaji RN. et al. Cholera outbreak investigation, Gajala community, Birnin kudu local government area (LGA), Jigawa state, Nigeria, September 2015. Int J of Infect Dis 2016;45:144-145.

8. Lawoyin TO. et al. Outbreak of cholera in Ibadan, Nigeria. Euro J Epidemiol 1999;15(4):365-368.

9. Dan-Nwafor CC. et al. A cholera outbreak in a rural north central Nigerian community: An unmatched case-control study. BMC Public Health 2019;19(1):112.

10. Ndon JA. et al. Vibrio-associated gastroenteritis in the lower Cross-River Basin of Nigeria. J Clin Microbiol 1992;30(10):2730-2732.

11. Ngwa MC. et al. The multi-sectorial emergency response to a cholera outbreak in Internally Displaced Persons camps in Borno State, Nigeria, 2017. BMJ Glob Health 2020;5(1).

12. Denue BA. et al. Low case fatality during 2017 cholera outbreak in Borno State, North Eastern Nigeria. Annals of Afr Med 2018;17(4):203.

13. Adeneye AK. et al. Risk factors associated with cholera outbreak in Bauchi and Gombe states in north East Nigeria. J Pub Health and Epidemiol 2016;8(11):286-296.

14. Akyala Ishaku A. et al. Investigation of cholera outbreak in an urban north central Nigerian community: the Akwanga experience. Public Health Res 2014;4:7-12.

15. Fatiregun AA. et al. Cholera outbreak in a southwest community of Nigeria: investigation of risk factors and evaluation of a district surveillance system. West Afr J Med 2013;32(3):173. [Abstract only]

16. Hunponu-Wusu OO. Epidemiological aspects of an El Tor cholera outbreak in Kaduna, Nigeria. Trop Geograph Med 1973;25(3):277-81. [Abstract only]

17. Abubakar AT. et al. Cholera outbreak-IDP camps in Maiduguri, northern Nigeria, September 2015. Int J Infect Dis 2016;45:132.

18. Gidado S. et al. Cholera outbreak in a naïve rural community in Northern Nigeria: the importance of hand washing with soap, September 2010. Pan Afr Med J 2018;30.

Chapter 5

The Impact of Social and Environmental Extremes on Cholera Time-Varying Reproduction Number in Nigeria: A Machine Learning Approach

Dissemination

A modified version of the full chapter is published at:

Charnley GEC, Yennan S, Ochu C, Kelman I, Gaythorpe KAM, Murray KA. Investigating the impact of social and environmental extremes on cholera time-varying reproduction number in Nigeria. *PLoS Global Public Health* 2022;2(12):e0000869

Abstract

Nigeria currently reports the second highest number of cholera cases in Africa, with numerous socio-economic and environmental risk factors, some of which have been discussed in previous chapters. Less investigated are the role of extreme events, despite recent work showing their potential importance. To address this gap, time-varying reproduction number (R_t) was estimated from cholera incidence in Nigeria and a machine learning approach used to evaluate its association with extreme events (conflict, flood, drought) and pre-existing vulnerabilities (poverty, sanitation, healthcare). Using the best fit model, a traffic-light system for cholera outbreak risk was created, with two scenarios (Red & Green) and used to predict R_t. The system highlighted extreme events and socio-economic thresholds for outbreaks to occur. Chapter 5 found that reducing poverty and increasing access to sanitation lessened vulnerability to increased cholera risk caused by extreme events (monthly conflicts and the Palmers Drought Severity Index). The work presented here shows the need for sustainable development for disaster prevention and mitigation and to improve health.

5.1 Introduction

Pre-existing vulnerabilities, such as poverty and lack of access to clean water and sanitation, have a significant impact on cholera control and the effect of these pre-existing vulnerabilities on disease risk can be exacerbated in times of environmental (drought, floods) and social extremes (conflicts). These risk factors can in turn act as a catalyst for, or exacerbate the impacts of, outbreaks, especially in endemic countries where the disease already circulates. Despite the known catalytic effect of disasters on disease outbreaks, previous research has largely examined extreme events in isolation [1, 2] (Chapters 3 & 4), despite notable examples of multi-hazard outbreaks [3], while others do not include multiple pre-existing socio-economic factors into the methodology [4, 5]. Research linking several social and environmental extremes to diseases, including via risk factor cascades, is a global research gap and is important for predicting cholera transmission and mitigating outbreaks [6].

Nigeria has one of the highest cholera burdens in Africa (and globally) [7, 8] and has experienced many large outbreaks [9, 10, 11, 12]. This is likely due to the presence of many underlying social and environmental risk factors, including a favourable climate [13, 14], poor access to WASH [15,

16] and a high proportion of the population living in poverty (62% at <1.25/day) [17, 18, 19]. It also has a relatively robust reporting system which may correlate with more cases, as cholera is an under-reported disease and cases and deaths are often missed or misattributed. The country has been frequently challenged by both social and environmental extremes such as droughts and floods, which may alter in intensity and frequency with climate change [6, 18], along with ongoing conflict in the northeastern region due to Boko Haram (Islamic State West Africa Province) [20, 6]. Due to the ongoing presence of these issues in Nigeria (conflict and environmental change), it is important to understand their specific effects to protect the health of the population and inform effective infectious disease control policy.

The aim of Chapter 5 is to expand the current understanding of the role of extreme events in causing or contributing to cholera in Nigeria. In collaboration with the Nigeria Centre for Disease Control (NCDC), a range of environmental and social covariates were evaluated by way of machine learning to understand how they influence cholera time-varying reproduction number (R_t). The chapter takes advantage of the predictive capacity of machine learning techniques and uses R_t in a novel application to understand the complexities of disaster-related risk factors on cholera outbreak evolution, rather than case or deaths numbers. The originality of the data used here are important, as modelling and testing cholera assumptions across multiple data sources are important to improve our understanding of cholera dynamics.

A traffic-light system of cholera risk will be created using the model with the best predictive power. The system will help to illustrate how disasters and pre-existing vulnerabilities alter R_t in Nigeria, stating specific quantitative thresholds and triggers. The aim of the traffic-light system is to move beyond stating and discussing risk factors more generally, creating thresholds to help guide policy targets. Cholera predictions using quantitative models and scenarios are a global research gap, and are important in understanding cholera risk factors and outbreak triggers now and in the future. The novel approach presented here aims to use a relatively simple model, based on routinely collected available data, to predict the level of cholera risk for Nigeria as low or high. The framework could be employed by a range of professionals working in fragile settings to target interventions towards key disaster-related risk factors. The objectives for Chapter 5 are as follows:

1. Use cholera surveillance data from Nigeria to calculate cholera incidence and model cholera time-varying reproduction number.

2. Understand the impacts of multiple extreme conditions on cholera outbreaks in Nigeria.

3. Evaluate how the pre-existing vulnerabilities of the population may alter the relationship between cholera and extreme events.

4. Predict a traffic-light system of cholera outbreak risk, to understand triggers and thresholds for outbreaks to occur in Nigeria.

5.2 Methods

5.2.1 Datasets

Cholera data were obtained from NCDC and contained surveillance linelist data for 2018 and 2019. The data included information on patients' age and sex and were on a daily temporal scale, to administrative level 4 (village or settlement). The data also provided information on the outcome of infection and whether the patient was hospitalised. The data were subset to only include cases that were confirmed either by rapid diagnostic tests or by laboratory culture and only these confirmed cases were used in the analyses. To test if removing suspected cases bias the results and to prove model robustness, a sensitivity analysis was completed running the analysis on all the cholera data (confirmed and suspected), further details and the results are shown in Supplementary Information 5.1.

Additionally, NCDC provided OCV data, which were represented by the campaign start and end date, the location (administrative level 1, states) and the coverage. OCV was transformed to an annual binary outcome variable (0-1) for each state (e.g., if coverage was 100% in a specific year and state, the data point was assigned 1, no states that were vaccinated had less than 100% coverage). OCV coverage was assigned annually based on the assumption that vaccine-derived immunity to cholera is relatively short (see 1.3.1 History and Global Burden of Cholera).

A range of covariates were investigated based on previously understood cholera risk factors from the literature and prior chapters. Covariates (and their temporal/spatial scales) included factors related to conflict (monthly, daily/settlement) [21], drought (Palmers Drought Severity Index, Standardised Precipitation Index (SPEI), monthly/state) [22, 23], internally displaced persons (IDPs) (households,

individuals, annual/national) [24], WASH (improved drinking water, piped water, improved sanitation, open defecation, basic hygiene, annual/state) [25], healthcare (total facilities, facilities per 100,000 people, annual/state) [21], population (total, annual/state) [26] and poverty (Multidimensional Poverty Index (MPI), headcount ratio in poverty, intensity of deprivation among the poor, severe poverty and population vulnerable to poverty, annual/state) [21].

PDSI data were taken from the same source as Chapter 3, and additional information on the collection and granularity of these data are available in 3.2.1 Datasets and Study Period. The SPEI data were from the National Centre for Earth Observation's Centre for Environmental Data Analysis archive. The dataset is for the whole of Africa at 1 month to 48 months time scales and calculated based on precipitation estimates from satellite-based and station data. The spatial resolution of the data was 5x5 kilometers at a monthly temporal scale. The dataset was provided as a netCDF file, and similar to the PDSI data, shapefiles were used to attribute the grids to administrative levels and available at administrative level 1 [23].

Here, several drought metrics were used, measured across multiple time windows. The benefits of using multiple metrics when investigating both drought and floods have been suggested in Chapter 3. The drought indices were used to measure relative dryness/wetness, not long-term drought changes, due to the short timescale of the cholera surveillance dataset (2018-2019). Using a drought metric, instead of raw precipitation or temperature data were selected to account for several environmental variables (temperature, precipitation and potential evapotranspiration) and to better present how the raw data translated into drier or wetter environments.

Covariate data were on a range of spatial and temporal scales (as stated above), therefore administrative level one (state) was set as the spatial granularity, as this best captured the data used. All covariate data, other than IDPs (national) and conflict (settlement/village), were at administrative level 1 and data on a finer spatial scale were attributed to administrative level 1 (counting the conflict events) or repeated for national data. The finest temporal scale possible (daily) was used for covariate selection and if two temporal scales were available, both were run through covariate selection. Covariate values were repeated if data were not available at the daily level, which was the case for covariates unlikely to change at a daily rate e.g., poverty, WASH.

For covariate predictions (5.2.4 Nowcasting and 5.2.5 Traffic-Light System for Cholera Outbreak Risk), monthly data were used, as this best captured the temporal scale of the data selected in the best fit

model (monthly or annual). Additionally, values were also repeated if unlikely to change at a daily rate (e.g., MPI), or if the daily change was likely to be incorrect due to lags and reporting issues (e.g., conflict and cholera). The datasets and methods used here were approved by Imperial College Research Ethics Committee and a data sharing agreement between NCDC and the research team.

5.2.2 Incidence and Reproduction Number

The 2018 and 2019 laboratory confirmed linelist data were used to calculate incidence. Incidence was calculated on a daily scale by taking the sum of the cases (not the rates) reported by state and date of onset of symptoms. This created a new dataset with a list of dates and corresponding daily incidence for each state. All analyses were completed in R version 4.1.0. (packages "incidence" [27] & "EpiEstim" [28]).

Rather than using incidence as the outcome variable (which has less implicit assumptions in terms of latency, contacts and population size), R_t was calculated, as it is more descriptive and provides information on epidemic evolution (e.g., $R_t >= 1$ means cases are increasing), instead of new reported disease cases for a single time point. R_t was calculated from the incidence of confirmed cases using the parametric serial interval method, which uses the mean and the standard deviation of the serial interval (SI). The parametric SI method was used (vs non-parametric which uses a discrete distribution), as cholera SI can be adequately modelled by a gamma probability distribution and has a fixed set of parameters e.g., mean and standard deviation.

SI is the time from clinical onset in the primary case to onset in the secondary case and therefore impacts the evolution of the epidemic and speed of transmission. It is defined as $TA + IB$, where TA is the time from clinical onset in the first case to when they infect the second case and IB is the incubation period of the subsequent case. The SI for cholera is well-documented and there are several estimates in the literature [29, 30, 31]. To account for this reported variation in SI, a sensitivity analysis was conducted with SI set at 3, 5 and 8 days and a standard deviation of 8 days.

The mean and the standard deviation of the SI was used in a renewal equation. Renewal theory is a generalisation of a Poisson process for arbitrary holding times. A holding time is the time between events and must have any positive distribution (usually exponential), be independant of each other, have a finite mean and be identically distributed. The equation is shown below, where I_t is

the incidence of symptom onset at time t, I_{t-s} is the past incidence and ω_s is the serial interval distribution. The mean R_t value (with the standard deviation plotted in Figure 5.3) calculated from this process was used as the outcome variable.

$$I_t \sim Pois(R_t \sum_{s=1}^{t} I_{t-s}\omega_s).$$

Estimating R_t too early in an epidemic increases error, as R_t calculations are less accurate when there is lower incidence over a time window. When few or no case counts are available to constrain inference, the method is largely driven by the inherent prior distributions (gamma) and assumptions. A way to understand how much this impacts R_t values is to use the coefficient of variation (CV), which is a measure of how spread out the dataset values are relative to the mean (CV = standard deviation/mean). The lower the value, the lower the degree of variation in the data, which is achieved by having a higher incidence over the time window. A coefficient of variation threshold was set to 0.3 (or less) as standard, based on previous work [28]. To reach the CV threshold, calculation start date for each state was altered until the threshold CV was reached. States with <40 cases were removed, as states with fewer cases did not have high enough incidence across the time window to reach the CV threshold. Additionally, R_t values were calculated over monthly sliding windows, to ensure sufficient cases were available for analysis within the time window.

5.2.3 Model Fitting and Covariate Selection

Supervised machine learning algorithms such as decision-tree based algorithms, are now a widely used method for predicting disease outcomes and risk mapping [32, 33]. They work by choosing data points randomly from a training set and building a decision tree to predict the expected value given the attributes of these points. Transparency is increased by allowing the number of trees (estimators), number of features at each node split and re-sampling method to be specified. Random Forests (RF) then combines several decision trees into one model, which has been shown to increase predictive accuracy over single tree approaches (due to less bias and overfitting), while also dealing well with interactions (by taking hierarchical dependencies into account) and non-linear relationships (by using regression random forests) [34, 35].

The covariates listed above (conflict, drought, IDPs, WASH, healthcare, population and poverty) were first clustered to assist in the selection of covariates for model inclusion and to understand any multicolinearities. Despite RF automatically reducing correlation through subsetting data (re-sampling helps reduce the chances of re-fitting the same patterns) and tuning the number of trees and depth (the more the tree grows, the more likely it is to be overfit, controlling the depth reduced this) [33, 36], the process here lends support that the final model is measuring somewhat independent processes and not purely overfitting the same patterns [32]. The clustering was based on the correction between the covariates meeting an absolute pairwise correlation of above 0.75. A secondary covariate selection process (as used in Chapter 3) was run during preliminary analysis and acted as a method of validation. The process is detailed in Supplementary Information 5.2.

Random forest variable importance was used to rank all 22 clustered covariates. Variable importance provided an additional method of guiding the fitting of the best fit model, by testing the covariates found to be of higher variable importance first, but still testing every possibility. In this context, variable importance is a measure of the cumulative decreasing mean standard error each time a variable is used as a node split in a tree. The remaining error left in predictive accuracy after a node split, based on out-of-bag (OOB) sampling, is known as node impurity and a variable which reduces this impurity is considered more important.

As RF models are a well established method, there are now a number of methods available to fit the models using R software (R Studio). The methods vary in terms of fit and control over model tuning (although differences in model fit were expected to only be marginal). Here, three different packages were tested for model fitting on a subset of the data, these included randomForest [37], ranger [38] and caret [39]. The caret package was selected as it was found superior at fitting the model based on correlation, R^2 and RMSE and also allowed for the greatest degree of control over tuning.

Training (70% of data) and testing (30%) datasets were created to train the model and test the model's predictive performance. A single binary split was set and used on the full dataset to create the training set and testing set. Random forest regression models (as opposed to classification models) were used since the outcome variable (R_t) was continuous. The model was manually tuned using a parameter grid search and estimated an optimal number of predictors at each split of two, based on the lowest OOB error rate with Root-Mean-Square Error (RMSE) used as the evaluation metric. RMSE indicates how well the model can predict the value of a response variable in absolute terms.

(R package "rsample", functions initial_split(), training() and testing() [40])

The parameters for training were set to repeated k-fold cross-validation for the re-sampling method. Cross-validation was chosen over bootstrapping, as this allows the model performance to be assessed based on test metrics. Ten re-sampling interactions (based on model stability) and five complete sets of k-folds to complete were used. There is no one correct number of k-folds to choose, as it is hard to estimate how well the folds represent the data. Generally, the number of k-folds is based on sample size, and four or five are usually selected, unless the sample is particularly large (which was not the case here). Five folds meant that around 20% of the data were used for validation, which has proved robust in the past (R functions trainControl() and train() [39]).

A stepwise analysis was used to fit the models to R_t for each SI (3, 5 & 8 days), taking into consideration the covariate clustering and variable importance. One covariate was selected from each cluster, and all combinations of covariates were tested until the best-fit model was found. Models were assessed against each other in terms of predictive accuracy, based upon R^2 (coefficient of determination) and RMSE. Predictions were then calculated on the testing dataset to compare incidence-based (R_t values calculated using the incidence data) vs covariate-based R_t values (R_t values calculated through model predictions). The terms, actual vs predicted was not used here, as all R_t values were modelled making the term "actual" misleading in this context. Model performance evaluations were built on multiple metrics including correlation (against the testing dataset), R^2 (how well the predictor variable can explain the variation in the response variable) and RMSE.

Despite random forest models being accurate and powerful for prediction, they are easily over-fit (fitting to the testing dataset too closely or exactly), therefore calculating prediction error is important. Little to no error in the predictions are an indication of over-fitting which can occur through predictions based off too small a dataset, more parameters than can be justified by the data (resulting in too many tree/too deep) and multicollinearity. Here, error was calculated using mean absolute error (MAE), which indicates how close the predictions are to the outcomes, in absolute terms. Where y_i is the prediction and x_i is the true value, with the total number of data points as n.

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n}.$$

5.2.4 Nowcasting

The best fit model, in terms of predictive power according to the metrics above, was used to predict R_t for the remaining states which did not have sufficient reported cases (< 40) to calculate R_t using incidence or had missing data for certain dates. Data for the best fit model covariates were collected for the states and missing dates from the sources given above. This created a full monthly dataset for each state for 2018 and 2019, with sliding monthly values of R predicted, using covariate data for the same month.The data for the selected covariates are shown spatially in Supplementary Figure 5.1.

5.2.5 Traffic-Light System for Cholera Outbreak Risk

The best fit model was then used to predict the traffic-light system for cholera outbreak risk, by manipulating the covariate values and using these to predict R_t. A disadvantage of RF models is that they are considered a "black box" modelling technique, meaning the complexity of the model makes it difficult to see how the input produces the output and the relationships and patterns that the model is based upon. The scenarios helps to address this methodological limitations, by further understanding and illustrating the relationships between the covariates and the outcome variable. The traffic-light system was defined as:

- Red - High risk of cholera transmission = Covariate values which predicted R_t over 1
- Green - Low risk of cholera transmission = Covariate values which predicted R_t below 1

By using the scenarios, cholera outbreak triggers were identified based on the conditions of the four selected covariates. The covariate values were manipulated in a similar way to the creation of the marginal effect plots in Chapter 3 (3.3.2 Output from the Best Fit Model). Covariate values were changed in a step-wise pattern through a full range of possible values (Sanitation 30-70, PDSI -4-+4, Conflict 0-40, MPI 0.1-0.5), while the other covariates were kept constant at the mean value for R >= 1 in the full dataset (Sanitation 43.64327, PDSI 0.2192958, Conflict 7.152047, MPI 0.3884532). The mean value for R >= 1 was chosen for the other covariates, to identify the threshold needed to push R values below one, when poor conditions (in terms of cholera risk) were present. To illustrate the historical trends between the best fit model covariates and the R_t thresholds ($R_t >= 1$, $R_t < 1$), the

data were split both spatially (by state) and temporally (by month) in Figure 5.1 & Supplementary Figure 5.2.

5.2.6 Spatial Heterogeneities

To understand spatial differences in the relationship between the selected social and environmental extremes (conflict and PDSI) and cholera outbreak risk and the role pre-exiting vulnerabilities played in altering these relationships, six states were selected for additional analysis. These states were selected because they had either a clear positive or clear negative relationship with conflict or PDSI and R_t (PDSI is hypothesised to increase R_t at either end of the scale, +4 = extreme wetness or -4 = extreme dryness) [41, 4] and similar mean values for the other covariates in the model, to increase comparability. The included states were Borno, Kaduna, Nasarawa, Ekiti, Lagos and Kwara (see Figure 5.1). The process above for predicting R_t under the traffic-light scenarios was repeated for the six states but only PDSI and conflict values were manipulated, keeping the other three covariates at the mean value for $R_t >= 1$ across the full dataset for the state. The spatial analyses identified the thresholds in conflict and PDSI needed to push R_t values below 1.

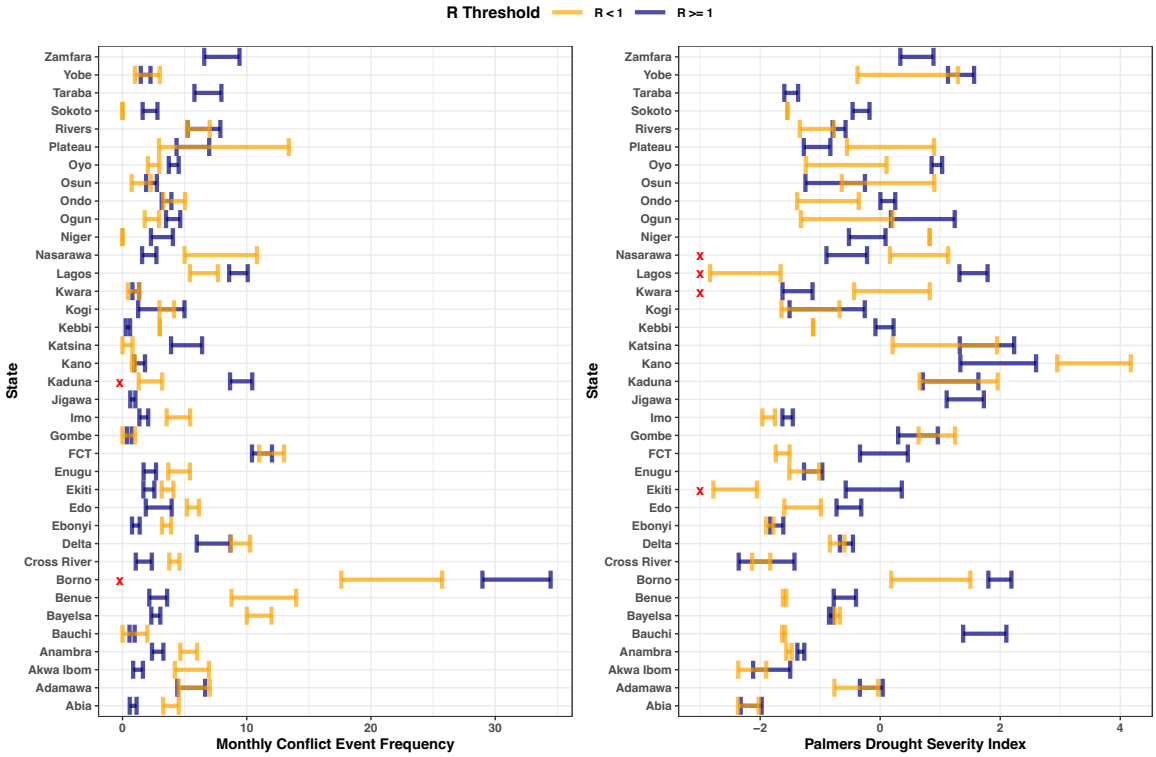


Figure 5.1: Historical spatial trends between the selected social (conflict, **left panel**) and environmental (PDSI, **right panel**) extremes and the R_t thresholds ($R_t >= 1$, $R_t < 1$). Presented as the mean and standard error for the two covariates for the full dataset split by state and R_t threshold. The "x" shows the states which were included in the spatial heterogeneity analysis: Conflict (Borno and Kaduna), extreme wetness (Lagos and Ekiti), extreme dryness (Nasarawa and Kwara).

5.3 Results

5.3.1 Incidence and Reproduction Number

In Nigeria, there were 837 and 564 confirmed cholera cases for 2018 and 2019, respectively, resulting in 279 data points at 70% training across 6 states (out of 44,208 and 2,486 total cases for 2018 and 2019, respectively). The results from the sensitivity analysis (5,627 data points across 16 states) including confirmed and suspected cases, proved model robustness and that the smaller dataset was not biasing the results (see Supplementary Information 5.1). The geographic distribution of confirmed cases is shown in Figure 5.2a and are concentrated in the northeast of the country, with Adamawa, Borno,

Katsina and Yobe having the highest burden. The high number of cases in Katsina may be a product of population size, being the third most populous state in Nigeria, whereas for Adamawa (21/37), Borno (13/37) and Yobe (30/37) this was not the case [42].

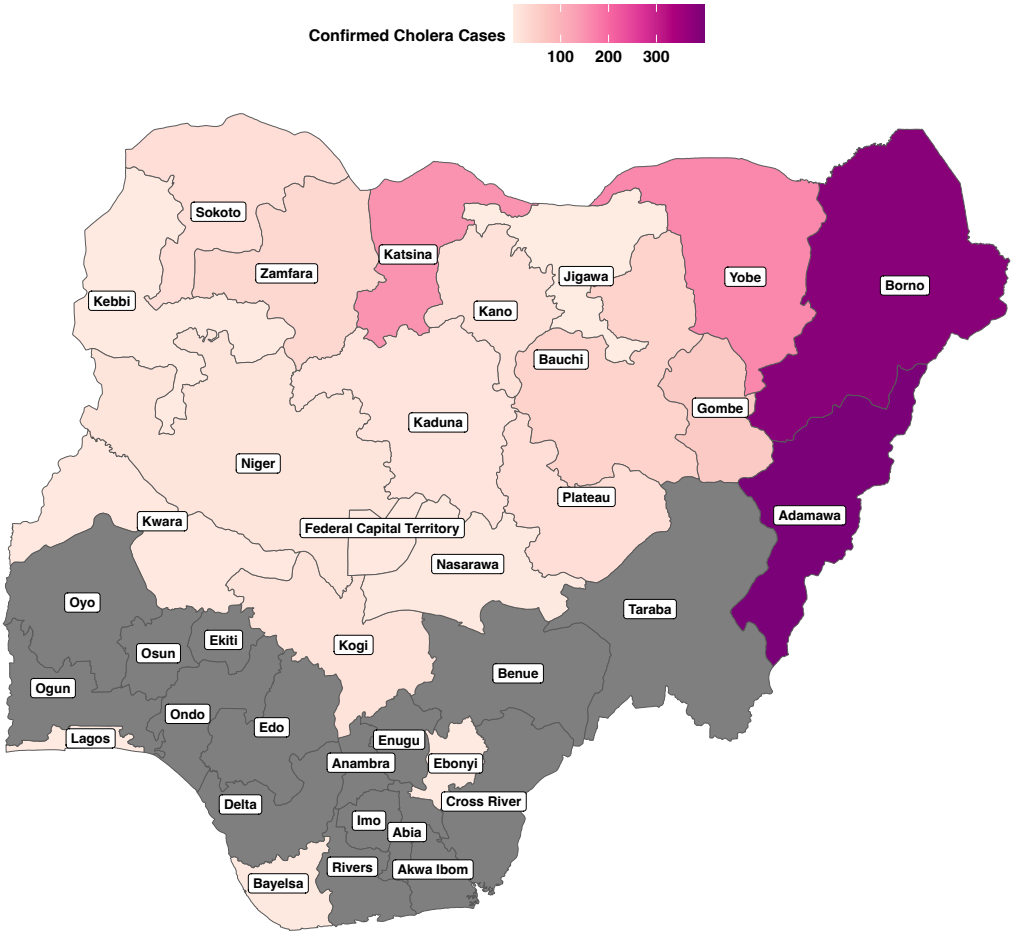


Figure 5.2: Number of confirmed cholera cases in 2018 and 2019 in Nigeria by state, grey indicates states that had no reported confirmed cases.

Six states for 2018 and two states for 2019 had sufficient cases (>40) to have R_t calculated, including Adamawa (2018 & 2019), Bauchi (2018), Borno (2018 & 2019), Gombe (2018), Katsina (2018) and Yobe (2018). Both the R_t values and the incidence data used to calculate R_t are shown temporally in Figure 5.3 for each state and year. Some states appear to have a peak in transmission around June-July (hottest and driest months), whereas others appear later during September to October (rainy season) (see Supplementary Figure 5.3 for average Nigerian climate).

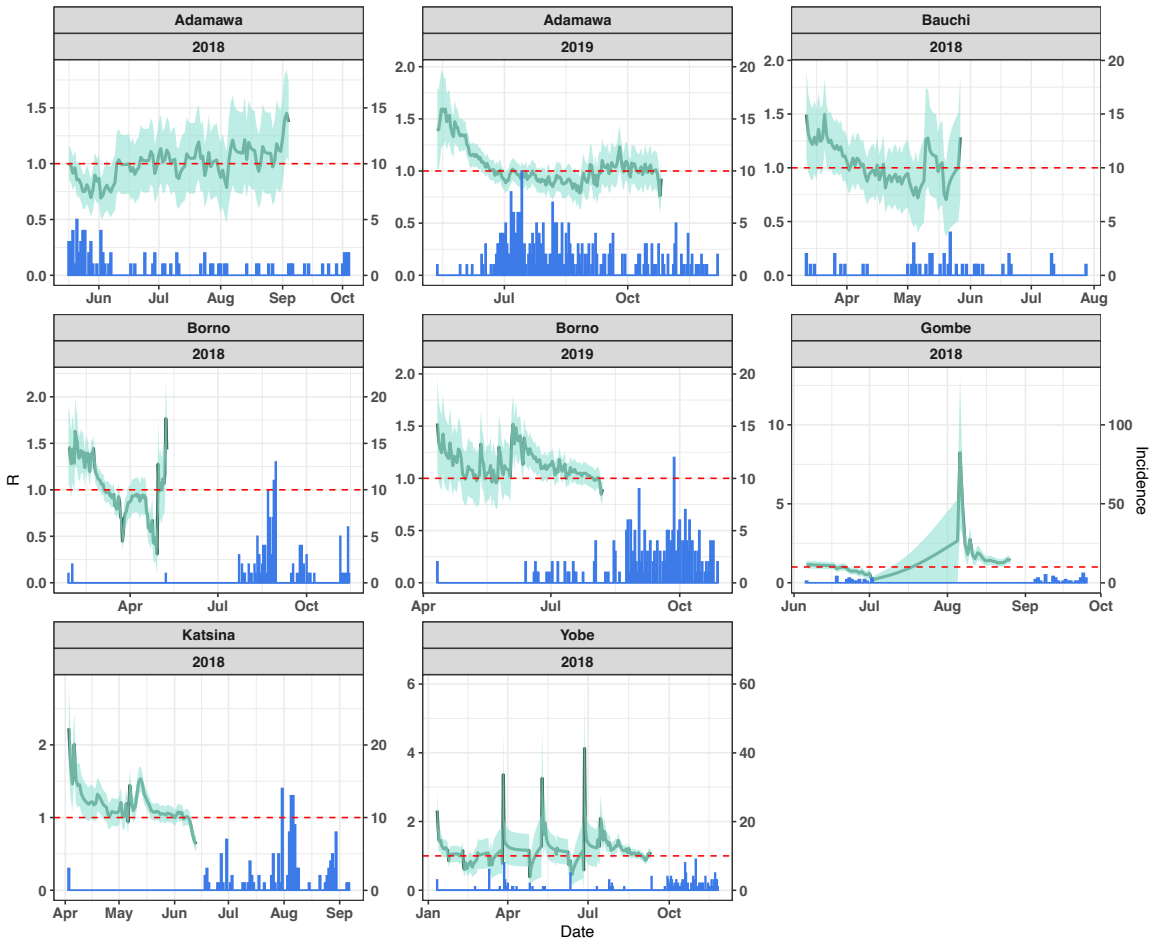


Figure 5.3: Mean R_t values over monthly sliding windows with standard deviation around the mean (line and shading), calculated from the daily incidence (bar) of cholera. The data used were only confirmed cholera cases for 2018 and 2019 of states which met the threshold of >= 40 cases. Presented are R_t values calculated using an SI of 5 days (8 days standard deviation).

5.3.2 Covariate Selection and Model Fitting

Twenty-two covariates were included in the clustering and variable importance analysis and were grouped into nine clusters. The clusters and variable importance (based on reducing node impurity) of each covariate are shown in Figure 5.4. Stepping through different covariate combinations, the best fit model included number of monthly conflict events, MPI (annual), PDSI (monthly) and improved access to sanitation (annual), fitted to R_t values with a serial interval of 5 days (standard deviation: 8 days). The fit of the incidence-based vs covariate-based R_t values (including MAE) are shown in Figure 5.5 and had a correlation of 0.87, with the model RMSE at 0.17 and R^2 of 0.51.

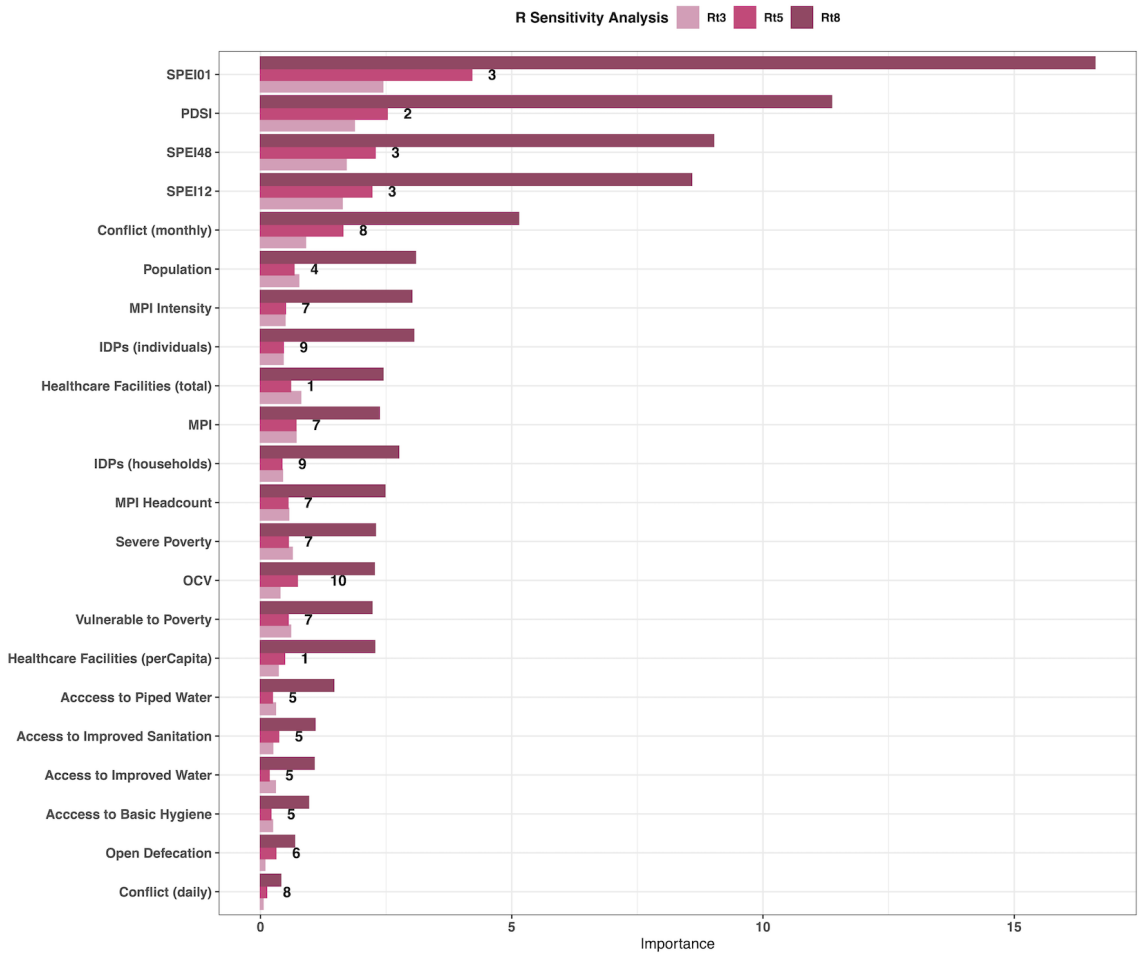


Figure 5.4: The variable importance for the 22 covariates tested for inclusion in the best fit model. All three serial interval values tested are shown (Rt3 - 3 days, Rt5 - 5 days, Rt8 - 8 days) and the numbers represent the clusters. Variable importance is measured through node impurity (see 5.2.3 Methods for details). SPEI01, 12, 48 - Standardised Precipitation Index calculated on 1, 12 and 48 month scale. PDSI - Palmers Drought Severity Index. MPI - Multidimensional Poverty Index. IDP – Internally Displaced Persons. OCV - Oral Cholera Vaccination.

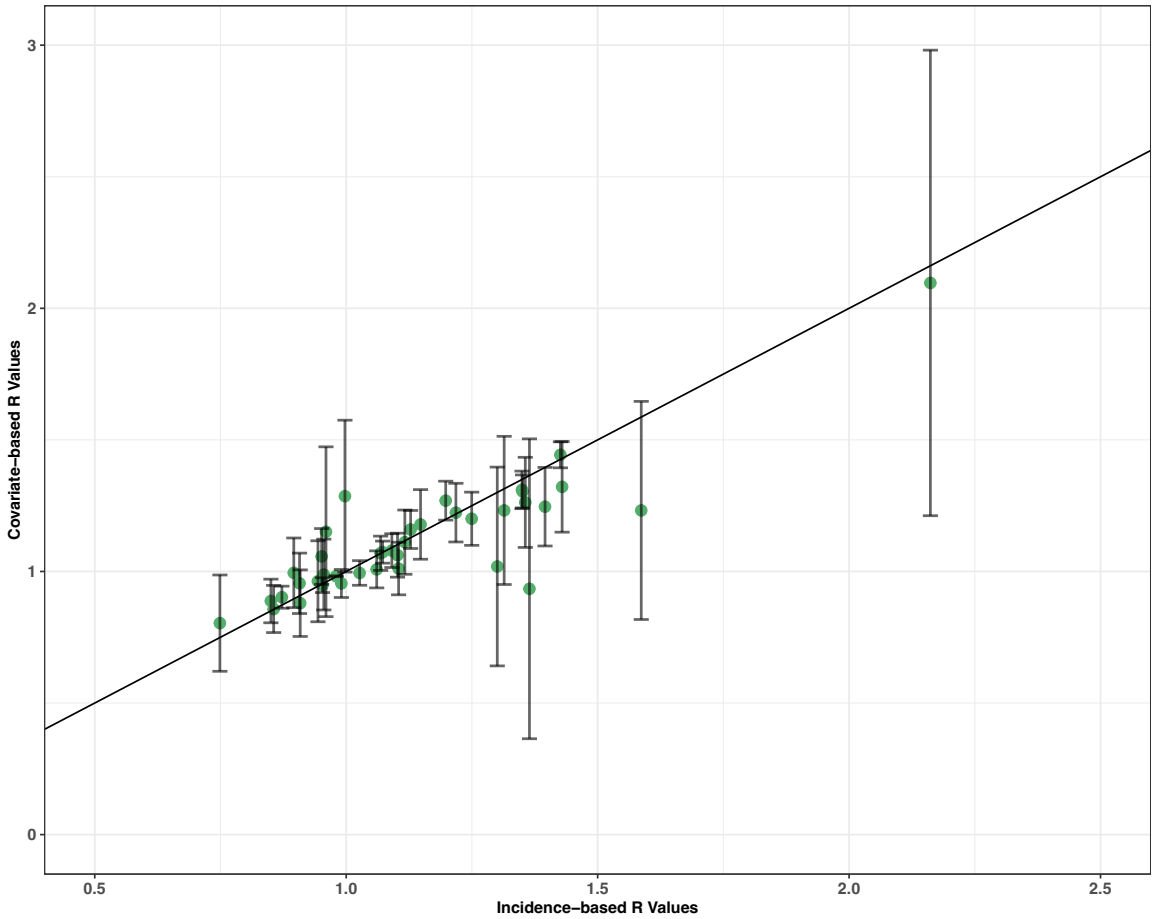


Figure 5.5: Incidence-based vs covariate-based R_t values for the best fit model, fitted to the testing dataset. The error bars show mean absolute error and the line is a linear trend line.

5.3.3 Nowcasting

Using the best fit model, R_t was predicted for the remaining 31 states which did not have sufficient cases to be included in the R_t calculations and any missing dates for the six states which were included. The term "nowcasting" is used to describe this process, which is used throughout epidemiology to understanding the current state of a disease by assessing key epidemiological characteristics of an ongoing outbreak, in this case, using out of sample predictions [43]. This created estimates of R_t for all 37 states on a monthly temporal scale for 2018 and 2019. The predictions provide further evidence that the model accurately predicts R_t, as the higher R_t values were in areas with known elevated cholera burden (northern and northeastern regions) and the states which only marginally fell below the threshold (<40 confirmed cases) for R_t calculations (e.g., Niger, Sokoto, Taraba) (Figure 5.6).

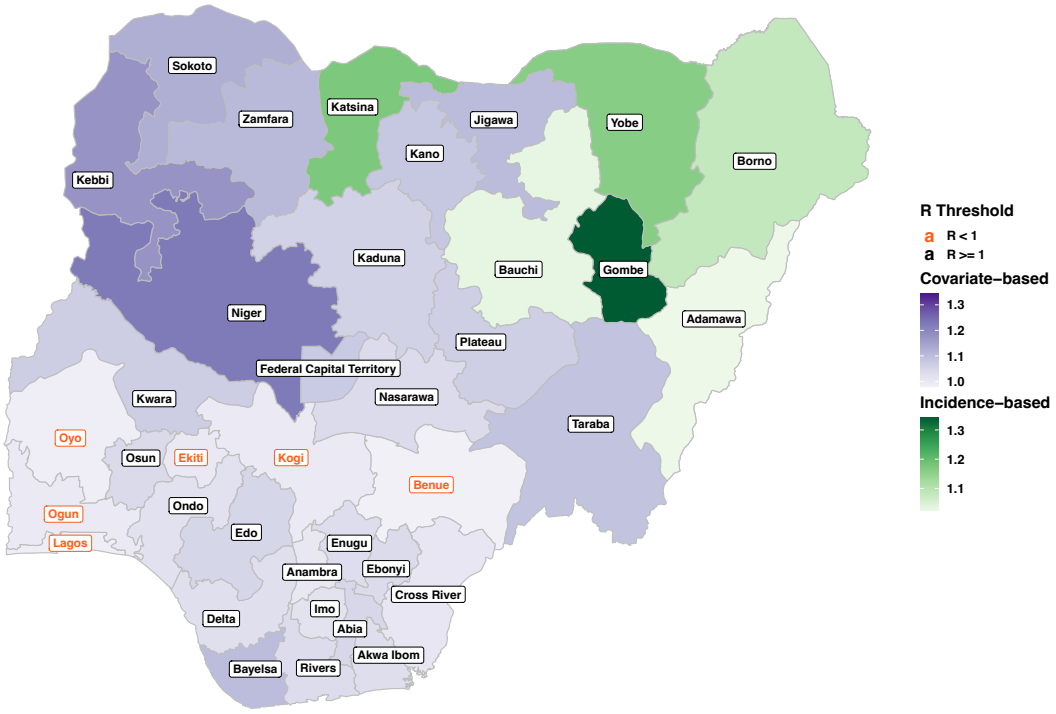


Figure 5.6: Average R_t values for 2018 and 2019 for all 37 Nigerian states. Incidence-based (green) - the six states which met the >= 40 case thresholds. Covariate-based (purple) - the 31 states which did not meet the threshold and had R_t predicted using the best fit model. State label colour shows which states had an average R_t value of $R_t >= 1$ (black) and $R_t < 1$ (orange).

5.3.4 Traffic-Light System for Cholera Outbreak Risk

Figure 5.7 shows the predicted R_t values for the traffic-light scenarios (Red = R_t over 1 and Green = R_t less than 1) of cholera outbreak risk, based on the four selected covariates. Sanitation and MPI had a clear relationship with the R_t threshold, with consistently lower MPI (less poverty) and a higher proportion of people with access to sanitation seeing lower R_t values. R_t increased above 1 at 54% or lower for improved sanitation access and MPI values of above 0.38. The historical average sanitation level for $R_t >= 1$ was 43.6% for the full dataset, whereas for $R_t < 1$ it was 61.2%, for MPI the mean values were 0.38 and 0.13 for $R_t >= 1$ and $R_t < 1$, respectively.

In contrast, PDSI shows a less defined relationship, with Figure 5.1 & 5.7 showing the polarity of the relationship between PDSI and cholera. PDSI increased R_t values above 1 at around -0.25 to -4 and +0.75 to +4, showing that both wetter and drier conditions increased cholera transmission. Similar

to Chapter 3, this may also suggest that drought is more important in Nigeria, with drier conditions (PDSI -4 to 0) having a wider range of values in the Red scenario.

For monthly conflict events, R_t values increased above 1 at 16 events but this varied widely among states (Figure 5.1). Some states had very low conflict event frequency over the study period, which may have resulted in a less defined relationship in several areas. Furthermore, some states appeared to be less effected by conflict, in terms of cholera transmission, even when conflict frequency was high. Potentially due to less pre-existing vulnerability or due to the high conflict in the area resulting in better population preparedness for the disruption caused by conflict and/or the risk of cholera transmission in these fragile settings.

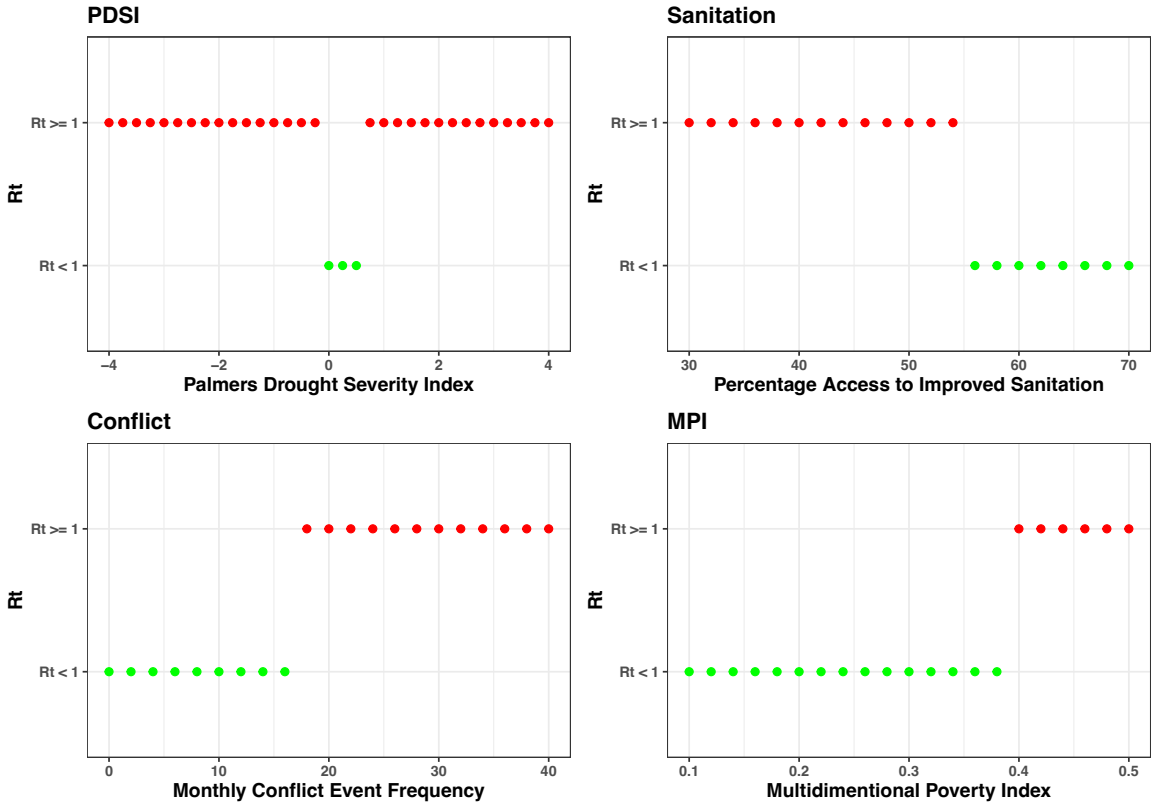


Figure 5.7: Traffic-light system of cholera risk in Nigeria. The traffic-light scenarios (Red = R_t over 1 and Green = R_t less than 1) for each of the four covariates in the best fit model. PDSI - Palmers Drought Severity Index. MPI - Multidimensional Poverty Index.

5.3.5 Spatial Heterogeneities

Conflict

Borno and Kaduna were selected due to their clear positive relationship between conflict and R_t (increased conflict and $R_t >= 1$). The traffic-light scenarios created for conflict in these two states found a consistently high cholera outbreak risk. The Green traffic-light scenario was relatively small, with only a narrow range of conflict values predicting R_t values less than 1. Both Kaduna and Borno have high levels of poverty and low access to sanitation (40-41% access). For Borno, raising monthly conflict events from 2 to 3 increased R_t above 1, but an increase in access to sanitation from 41% to 46% pushed the R_t value back below one. This relationship continued in a stepwise pattern and in a similar way for MPI but to a lesser degree. The results suggest that increasing sanitation and therefore decreasing vulnerability, allowed the states to adapt to increasing conflict and keep the R_t value below 1 (See Supplementary Figure 5.4).

Drought

Four states were investigated to evaluate the differences between extreme wetness (Lagos and Ekiti) and extreme dryness (Nasarawa and Kwara) and R_t values over 1 (Supplementary Figures 5.5 & 5.6). In contrast to Borno and Kaduna, all four states predicted consistently lower R_t values, a potential explanation for this is the high variable importance of PDSI (Figure 5.4) and the high levels of sanitation and low levels of poverty in all four states, contributing to overall lower predicted levels of cholera. Therefore, the model was detecting a signal in only small PDSI changes, that resulted in changing R_t values, which may not have been detected in other states with higher rates of poverty and lower levels of sanitation access. It also helps to highlight the bidirectionally of the relationship between PDSI and cholera transmission, confirming the hypothesis that both extreme wetness and extreme dryness can cause R_t to increase.

5.4 Discussion

The results presented here show the importance of social and environmental extremes on cholera outbreaks in Nigeria, along with the importance of underlying vulnerability and socio-economic factors. Of the 46,694 suspected cases in the full dataset, 1,401 were testing and found positive for cholera, either by rapid diagnostic test or culture in Nigeria in 2018 and 2019. The northeast of the country carried the highest burden of disease, particularly in Adamawa, Yobe, Borno and Katsina.

Six states were used to calculate the R_t values, including Adamawa, Bauchi, Borno, Gombe, Katsina and Yobe. Twenty-two covariates were considered for model inclusion and the best fit model according to the selected model performance measures (variable importance based on node impurity, RMSE, R^2 and correlations) included monthly conflict events, percentage of the population with access to sanitation, MPI and PDSI. Using the best fit model, nowcasting was used to calculate the R_t values for the remaining thirty-one states which did not meet the threshold.

The predicted R_t values from the traffic-light scenarios helped to shed light on the thresholds and triggers for raising R_t values above 1 in Nigeria. MPI and sanitation showed a well-defined relationship with R_t, with consistently higher access to sanitation and less poverty (lower MPI value) when R_t was less than 1. Thresholds which pushed R_t above one included decreasing access to sanitation below 54% and increasing the MPI above 0.38.

The relationship between R_t and conflict events and PDSI varied spatially, with some states showing a negative and some states a positive association. The effect of PDSI and conflict on R_t predictions appeared largely dependent on the access to sanitation and level of poverty within the states, with high levels of sanitation and low poverty resulting in a decreased effect on R_t. Better sustainable development in the state appeared to act as a buffer to social and environmental extremes in the results here, potentially due to people having less pre-existing vulnerability and therefore a greater capacity to adapt to these events.

5.4.1 Environmental & Social Extremes and Cholera in Nigeria

Since 2002, Boko Haram (and Islamic State's West Africa Province) has been gaining a foothold and territory in northeastern Nigeria which has resulted in ongoing conflict, unrest and oppression

of civilians [44]. Currently 5,860,200 people live in Borno state [45], where the fighting has been most concentrated. Millions of people comprise conflict-affected populations globally and there is an increasing proportion of people living in early post conflict areas, which are still fragile and can lack services [46]. In terms of health and disease, conflict has known risk factors for cholera along with several other diseases [20, 2, 47] and can worsen several of the social risk factors discussed above (see 4.4.1 Cholera/Conflict-related Risk Factors).

Here, conflict was included in the best fit model and in some states, highly influential in terms of cholera transmission. These results are the first to highlight the impacts of Boko Haram on a specific infectious disease, whereas previous research has focused more generally on public health [48, 49, 50]. The influence of conflict shows the need to incorporate and include the impacts of conflict in disease control measures in Nigeria and potentially other conflict-affected countries. Providing services and protecting health in conflict zones is especially challenging and coordination across organisations in reporting and operations are needed to streamline resources and prevent duplication of services [51]. The traffic-light system used here helps highlight the need to protect basic services and reduce inequities in conflict situations to protect health and prevent outbreaks.

PDSI and several of the other drought indices tested here showed high variable importance but, in some states, had only marginal influence on R_t predictions when the PDSI values were manipulated. When analysing spatial differences between R_t and PDSI, the relationship appears to be bidirectional, with both extreme wetness (PDSI = +4) and extreme dryness (PDSI = -4) associated with R_t values above 1. Furthermore, access to sanitation and poverty were important in how PDSI impacted R_t, similar to the impacts of conflict. These results should be interpreted with caution though due to the limited timescale of the data used. Drought is a slow-onset disaster, occurring over long timescales and PDSI is generally used to measure this long-term change. However, the insight presented shows that when some areas are impacted by either a relatively wetter or drier environment, extra vigilance may be needed to prevent cholera transmission.

There is significant evidence to show that both droughts [1, 4] and floods [5, 52] can cause cholera outbreaks and elevated transmission and in Nigeria the risks of the dry season and wet season have resulted in cholera outbreaks. This findings was also present in the cholera data here in Figure 5.3 and Supplementary Figure 5.3 Mechanisms through which this can occur includes a lack of water increasing risky drinking water behaviour and floods allowing for the dispersal of the pathogen [41, 53].

Continued work is essential to offset cholera risks related to droughts or floods through sanitation and hygiene, which can take significant time and resources [54]. Additionally, drought and flood mitigation strategies including sustainable water use and management and effective drainage and runoff systems, can reduce the overall impact of the disaster.

5.4.2 Pre-existing Vulnerabilities and Cholera in Nigeria

Currently, 73% of the enteric disease burden in Nigeria is associated with inadequate WASH [55] and here the results show the need for expansion of sanitation to reduce cholera risks and the shocks of extremes on its transmission. The results suggest that the expansion of sanitation would be particularly impactful for cholera control in states with <50% access. In a recent review on the implementation of non-pharmaceutical cholera interventions, there was generally a high acceptance of several WASH interventions. Despite this, education was key and building community relationships is needed to achieve this, such as understanding cultural differences and barriers [56]. This is especially important in areas with conflict, where trust between the government and residents may have been lost [53].

According to the World Bank [57], up to 47.3% (98 million people) of Nigeria's population live in multidimensional poverty. Poverty is a well-known risk factor for cholera [58], despite this, very few studies have suggested quantitative thresholds where poverty leads to disease, which is important for resource allocation. The results here showed that states with an MPI value above 0.38 should be areas for poverty alleviation prioritisation (which includes most of the northern states). Poverty can result in several risk factor cascades, which puts people at risk of not just cholera but several other diseases. Examples of these risks include marginalisation, poor access to WASH [15], inadequate housing [59], malnutrition [53] and overcrowding [60].

The expansion of sustainable development helps to reduce these risks and meeting or exceeding the Sustainable Development Goals would see significant gains in global health [61]. People living in poverty have fewer options and abilities to adapt to new and extreme situations, becoming trapped in the affected area or displaced to areas where their needs are not met [62, 63]. Measuring poverty in monetary terms alone can create issues due to its far reaching impacts and is an advantage of using MPI as a poverty indicator (which takes into account several factors). Nigeria currently has a cash transfer scheme, to increase people's social safety net. The scheme has allowed many Nigerians to

meet the household income limit for poverty elimination but there is a case for turning these funds and attention onto structural reform [64], or more targeted health and WASH programmes [65].

5.4.3 Limitations

A potential limitation may be lagged effects of the covariates on cholera [66, 67]. Both long-term and short-term changes to the population may take time before changes in cholera transmission are evident. Furthermore, some influential factors may be considered slow-onset or rapid-onset and therefore defining their beginning is subjective. The incubation period of cholera is short (<2 hours - 5 days) and Chapter 4 found that acute events are likely to cause an increase in cholera cases within the first week of the event [68, 69, 70]. Calculating R_t on monthly sliding windows and using monthly covariate data helped to reduce potential lagged effects on the R_t values, which would be captured if the one-week lag estimate is applicable here.

As previously discussed, cholera is considered an under-reported disease, and the lack of symptomatic cases means that many are likely to be missed. The data used in Chapter 5 are also on a relatively short timescale and therefore is more accurate at presenting cholera at the current time in Nigeria, rather than historically. Consequently, caution is needed when making generalisable conclusions. There are also incentives not to report cholera cases, due to travel restrictions and isolation and implications for trade and tourism [71]. While during times of crisis, cholera may be over-reported or more accurately represent the cholera burden in the area. This is due to the presence of CTCs, increased awareness among the population and healthcare workers and external assistance from non-governmental organization, detecting cases that may have been missed previously [20].

Despite the temporal (2 years) and spatial (6 states meeting the case threshold) limitations of the surveillance data, data of this detail is time consuming and difficult to collect in fragile settings and is the best data currently available to quantify cholera in Nigeria. Using confirmed cases only is necessary for modelling disease accurately, as in resource poor settings (such as outbreaks and conflicts) only a certain number of cases are confirmed, while it is very likely that several other intestinal pathogens could be causing disease, even in confirmed cholera cases. Therefore, the results and conclusions here are valid, if not more so, than models fit to longer but less accurate data sources.

Calculating R_t can have wide-ranging uncertainty (green shading in Figure 5.3), particularly when

data are lacking. Therefore, uncertainty is particularly high when there is a lull in reported cholera cases e.g., in January to July 2018 in Borno. Calculating R_t over monthly sliding windows and altering the R_t calculation start date for each state helped to reduce this uncertainty and in future studies, using other metrics of disease transmission, such as growth rates, would help reduce uncertainty. R_t calculations assumes that if cases occur, they are detected, which with cholera may be unlikely. Additional reporting assumptions for R_t include a constant reporting rate over the epidemic (CTCs and increased awareness make this unlikely), no cases are imported and therefore each case can be attributed to a previous case and the SI remains constant over the outbreak [72].

Using accurate data is particularly important when fitting RF models as they have relatively powerful predictive capacity. The performance metrics such as the correlation between covariate and incidence-based R_t values, along with the predictions of R_t replicating the reality of cholera in Nigeria (e.g., southern states predicted lower R_t) suggest that the model accurately predicts cholera R_t across the country. However, a limitation of RF models is that they are a predictive, rather than a descriptive tool and therefore are not as effective as other methods in highlighting patterns and relationships in data, which the traffic-light scenarios aimed to address.

5.4.4 Conclusion

The GTFCC 2030 target of reducing cholera deaths by 90% [73] will require acceleration of current efforts and significant commitment, particularly in high burden countries. Increasing cholera research and data are important in achieving this and the traffic-light system for cholera risk presented here sheds light on ways to reduce cholera outbreaks in fragile settings. The results use cholera R_t to highlight the importance of extreme events on cholera transmission in Nigeria, specifically droughts, floods and conflict and how reducing pre-existing vulnerability could offset the resultant cholera risk. The traffic-light system has identified specific targets and thresholds to avoid cholera outbreaks and will hopefully enable targeted and therefore more successful policy strategies.

The research presented in this chapter is the first time several disaster types and measures of population vulnerability have been evaluated together quantitatively in terms of cholera and shows the importance of doing so to gain a more accurate understanding of disease outbreaks in complex emergencies. The work helps to further quantify the impacts of Boko Haram in Nigeria and expand the understanding

of the extent of the conflict. Nigeria is currently working towards its ambitious goal of lifting 100 million people out of poverty by 2030 [54]. If it is successful, this could significantly improve health, increase quality of life and decrease the risks posed by social and environmental extremes.

The previous three chapters, along with the work here in Chapter 5 have presented a number of extremes and risk factors that influence cholera outbreaks. Conclusions from each chapter have drawn on similar themes, that reducing pre-existing vulnerabilities lessens the impacts of extreme conditions and offsets the health risks in fragile settings. It is important to put these findings and conclusions into a policy context and understand if global cholera targets can be met and what may be needed to do so.

The rate at which sustainable development is advancing is arguably too slow and despite several gains and increased attention in recent decades, whether this will be at a pace fast enough to reach the GTFCC 2030 targets mentioned above is very uncertain. It is also important to consider that global shocks and events (e.g., COVID-19), has the potential to erode this development. In Chapter 6, the evidence from the thesis so far, along with historical data and future projections will be used used to investigate if, based on the current evidence, the 2030 goals are likely to be achieved in Nigeria.

References

[1] G.E.C Charnley et al. "Exploring relationships between drought and epidemic cholera in Africa using generalised linear models". en. In: *BMC Infect. Dis.* 21 (2021), pp. 1–2.

[2] C.R. Wells. "The exacerbation of Ebola outbreaks by conflict in the Democratic Republic of the Congo". en. In: *PNAS* 116 (2019), pp. 24366–24372.

[3] V.A. Mugabe et al. "Natural disasters, population displacement and health emergencies: multiple public health threats in Mozambique". In: *BMJ Glob Health* 6.9 (2021), e006778.

[4] A. Rieckmann et al. "Exploring droughts and floods and their association with cholera outbreaks in sub-Saharan Africa: a register-based ecological study from 1990 to 2010". en. In: *Am. J. Trop. Med. Hyg.* 98.5 (2018), pp. 1269–1274.

[5] A. Jutla. “Environmental factors influencing epidemic cholera”. es. In: *Am. J. Trop. Med. Hyg.* 89 (2013), p. 597.

[6] K.O. Elimian. “What are the drivers of recurrent cholera transmission in Nigeria? Evidence from a scoping review”. en. In: *BMC Public Health* 20 (2020), pp. 1–3.

[7] M. Ali et al. “Updated global burden of cholera in endemic countries”. en. In: *PLoS Neglect. Trop. Dis.* 9 (2015), p. 0003832.

[8] J. Lessler. “Mapping the burden of cholera in sub-Saharan Africa and implications for control: an analysis of data across geographical scales”. en. In: *Lancet* 391 (2018), pp. 1908–1915.

[9] M.M. Dalhat. “Descriptive characterization of the 2010 cholera outbreak in Nigeria”. en. In: *BMC Public Health* 14 (2014), pp. 1–7.

[10] M.C. Ngwa. “The multi-sectorial emergency response to a cholera outbreak in internally displaced persons camps in Borno state”. en. In: *Health* 5 (2020), p. 002000.

[11] I.B. Sule et al. “Descriptive epidemiology of a cholera outbreak in Kaduna State, Northwest Nigeria, 2014”. en. In: *Pan Afr. Med. J.* 27 (2017).

[12] A.K. Adeneye. “Risk factors associated with cholera outbreak in Bauchi and Gombe States in North East Nigeria”. en. In: *J. Public Health Epidemiol.* 8 (2016), pp. 286–296.

[13] G.C. De Magny et al. “Regional-scale climate-variability synchrony of cholera epidemics in West Africa”. en. In: *BMC Infect. Dis.* 7 (2007), pp. 1–9.

[14] A.F. Abdussalam. “Modelling the Climatic Drivers of Cholera Dynamics in Northern Nigeria Using Generalised Additive Models”. en. In: *Int. J. Geog. Environ. Manag.* 2.1 (2016), pp. 84–97.

[15] S. Gidado. “Cholera outbreak in a naïve rural community in Northern Nigeria: the importance of hand washing with soap, September 2010”. en. In: *Pan Afr. Med. J.* 30 (2018).

[16] Y. Hutin, S. Luby, and C. Paquet. “A large cholera outbreak in Kano City, Nigeria: the importance of hand washing with soap and the danger of street-vended water”. en. In: *J. Water Health* 1 (2003), pp. 45–52.

[17] C.C. Dan-Nwafor. “A cholera outbreak in a rural north central Nigerian community: an unmatched case-control study”. es. In: *BMC Public Health* 19 (2019), pp. 1–7.

[18] G.C. Leckebusch and A.F. Abdussalam. “Climate and socioeconomic influences on interannual variability of cholera in Nigeria”. it. In: *Health Place* 34 (2015), pp. 107–17.

[19] United Nations Statistical Division. *Millennium Development Goal Indicators*. en. 2015. URL: https://unstats.un.org/unsd/mdg/SeriesDetail.aspx?srid=580.

[20] G.E.C. Charnley et al. “Using self-controlled case series to understand the relationship between conflict and cholera in Nigeria and the Democratic Republic of Congo”. en. In: *Emerg. Infect. Dis.* 28 (2022), pp. 2472–2481.

[21] H.D.X. *The Humanitarian Data Exchange*. en. 2021. URL: https://data.humdata.org.

[22] N.C.A.R. *Dai Global Palmer Drought Severity Index (PDSI)*. en. 2020. URL: https://rda.ucar.edu/datasets/ds299.0/index.html#!sfol-wl-/data/ds299.0.

[23] C.E.D.A. *High resolution Standardized Precipitation Evapotranspiration Index (SPEI) dataset for Africa*. en. 2019. URL: https://catalogue.ceda.ac.uk.

[24] I.O.M. *DTM Nigeria*. pt. 2021. URL: https://displacement.iom.int/nigeria.

[25] Joint Monitoring Programme. *Nigeria*. 2020. URL: https://washdata.org.

[26] WorldBank. *Data Bank Subnational Population*. fr. 2021. URL: https://databank.worldbank.org/source/subnational-population.

[27] Z.N. Kamvar et al. *incidence: Epidemic curves made easy using the R package incidence*. en. 2019. URL: https://doi.org/10.12688/f1000research.18002.1.

[28] A. Cori. *EpiEstim: Estimate Time Varying Reproduction Numbers from Epidemic Curves*. en. R package version 2.2-4. 2021. URL: https://CRAN.R-project.org/package=EpiEstim.

[29] A.S. Azman. "Urban cholera transmission hotspots and their implications for reactive vaccination: evidence from Bissau City, Guinea Bissau". en. In: *PLoS Neglect. Trop. Dis.* 6 (2012), p. 1901.

[30] A.S. Azman. "Population-level effect of cholera vaccine on displaced populations, South Sudan". en. In: *Emerg. Infect. Dis.* 22 (2016), p. 1067.

[31] R. Kahn. "Incubation periods impact the spatial predictability of cholera and Ebola outbreaks in Sierra Leone". en. In: *PNAS* 117 (2020), pp. 5067–73.

[32] A. Hamlet et al. "Seasonality of agricultural exposure as an important predictor of seasonal yellow fever spillover in Brazil". en. In: *Nature Commun.* 12 (2021), pp. 1–1.

[33] T. Kapwata and M.T. Gebreslasie. "Random forest variable selection in spatial malaria transmission modelling in Mpumalanga Province, South Africa". en. In: *Geospat. Health* 11 (2016), pp. 251–262.

[34] L. Breiman. "Random forests". en. In: *Mach. Learn.* 45 (2001), pp. 5–32.

[35] G. Biau. "Analysis of a random forests model". en. In: *J. Mach. Learn. Res.* 13 (2012), pp. 1063–95.

[36] R. Genuer, J.M. Poggi, and C. Tuleau-Malot. "Variable selection using random forests". en. In: *Pattern Recognit. Lett.* 31 (2010), pp. 2225–36.

[37] A. Liaw and M. Wiener. *randomForest: Classification and Regression by randomForest.* 2002. URL: https://CRAN.R-project.org/doc/Rnews/.

[38] M. N. Wright and A. Ziegler. *ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R.* 2017. DOI: 10.18637/jss.v077.i01.

[39] M. Kuhn. *caret: Classification and Regression Training.* en. 2021. URL: https://CRAN.R-project.org/package=caret.

[40] J. Silge et al. *rsample: General Resampling Infrastructure.* R package version 0.1.1. 2021. URL: https://CRAN.R-project.org/package=rsample.

[41] R.V. Tauxe et al. "Epidemic cholera in Mali: high mortality and multiple routes of transmission in a famine area". en. In: *Epidemiol. Infect.* 100.2 (1988), pp. 279–289.

[42] Nigerian National Bureau of Statistics. *Demographic Statistics Bulletin 2021*. 2021. URL: https://nigerianstat.gov.ng/elibrary/read/1241207.

[43] J.T. Wu et al. "Nowcasting epidemics of novel pathogens: lessons from COVID-19". In: *Nat. Med.* 27.3 (2021), pp. 388–395.

[44] J.A. Falode. "The nature of Nigeria's Boko Haram war, 2010-2015: A strategic analysis". en. In: *Perspect. Terror.* 10 (2016), pp. 41–52.

[45] Borno State Government. *Population*. en. 2016. URL: https://bornostate.gov.ng/population/.

[46] R.M. Garfield, J. Polonsky, and F.M. Burkle. "Changes in size of populations and level of conflict since World War II: implications for health and health services". en. In: *Disaster Med. Public Health Prep.* 6 (2012), pp. 241–6.

[47] F. Federspiel and M. Ali. "The cholera outbreak in Yemen: lessons learned and way forward". en. In: *BMC Public Health* 18 (2018), pp. 1–8.

[48] A. Chukwuma and U. E. Ekhator-Mobayode. "Armed conflict and maternal health care utilization: evidence from the Boko Haram Insurgency in Nigeria". In: *Soc. Sci. Med.* 226 (2019), pp. 104–12.

[49] O. Omole, H. Welye, and S. Abimbola. "Boko Haram insurgency: implications for public health". In: *Lancet* 385 (2015), p. 941.

[50] U. E. Ekhator-Mobayode and A. Abebe Asfaw. "The child health effects of terrorism: evidence from the Boko Haram Insurgency in Nigeria". In: *Appl. Econ.* 51 (2019), pp. 624–38.

[51] M. Ricau et al. "Evaluation of monitoring tools for WASH response in a cholera outbreak in northeast Nigeria". en. In: *J. Water Sanit. Hyg. Dev.* 11 (2021), pp. 972–82.

[52] P. Sidley. "Floods in southern Africa result in cholera outbreak and displacement". en. In: *BMJ* 336 (2008), p. 471.

[53] G.E.C. Charnley, I. Kelman, and K.A. Murray. "Drought-related cholera outbreaks in Africa and the implications for climate change: a narrative review". en. In: *Pathog. Glob. Health* (2021), pp. 1–10.

[54] F.I. Onwe et al. "Factors responsible for the 2015 Cholera outbreak and spread in Ebonyi state, Nigeria". en. In: *J. Epidemiol. Soc. Nigeria* 2 (2018), pp. 53–58.

[55] World Bank Group. "A Wake Up Call: Nigeria Water Supply, Sanitation, and Hygiene Poverty Diagnostic". en. In: *World Bank* (2017).

[56] J.A. Polonsky. "Feasibility, acceptability, and effectiveness of non-pharmaceutical interventions against infectious diseases among crisis-affected populations: a scoping review". en. In: *Infect. Dis. Poverty* 11 (2022), pp. 1–9.

[57] World Bank. *Tackling poverty in multiple dimensions: A proving ground in Nigeria.* en. 2021. URL: https://blogs.worldbank.org/opendata/tackling-poverty-multiple-dimensions-proving-ground-nigeria.

[58] A. Talavera and E.M. Perez. "Is cholera disease associated with poverty?" en. In: *J. Infect. Dev. Ctries.* 3.06 (2009), pp. 408–411.

[59] K. Penrose et al. "Informal urban settlements and cholera risk in Dar es Salaam, Tanzania". et. In: *PLoS. Neglect. Trop. Dis.* 4.3 (2010).

[60] M. Ververs and R. Narra. "Treating cholera in severely malnourished children in the Horn of Africa and Yemen". en. In: *Lancet* 390 (2017), pp. 1945–6.

[61] Y. Schirnding. "Health and sustainable development: can we rise to the challenge?" en. In: *Lancet* 360 (2002), pp. 632–7.

[62] M. Masozera, M. Bailey, and C. Kerchner. "Distribution of impacts of natural disasters across income groups: A case study of New Orleans". en. In: *Ecol. Econ.* 63 (2007), pp. 299–306.

[63] M. Lahsen and J. Ribot. "Politics of attributing extreme events and disasters to climate change". en. In: *Wiley Interdiscip. Rev. Clim. Change* 13 (2022), p. 750.

[64] S. Onyeiwu. *Nigeria's poverty profile is grim. It's time to move beyond handouts*. en. 2021. URL: https://theconversation.com/nigerias-poverty-profile-is-grim-its-time-to-move-beyond-handouts-163302.

[65] B. Ajisegiri. "Geo-spatial modeling of access to water and sanitation in Nigeria". da. In: *J. Water Sanit. Hyg. Dev.* 9 (2019), pp. 258–80.

[66] R. Reyburn et al. "Climate variability and the outbreaks of cholera in Zanzibar, East Africa: a time series analysis". en. In: *Am. J. Trop. Med. Hyg.* 84.6 (2011), pp. 862–869.

[67] M. Emch et al. "Local environmental predictors of cholera in Bangladesh and Vietnam". en. In: *Am. J. Trop Med. Hyg* 78 (2008), pp. 823–32.

[68] T. Fredrick et al. "Cholera outbreak linked with lack of safe water supply following a tropical cyclone in Pondicherry". en. In: *J. Health Popul. Nutr.* 33 (2012), p. 31.

[69] R. Bhunia and S. Ghosh. "Waterborne cholera outbreak following cyclone Aila in Sundarban area of West Bengal, India, 2009". en. In: *Trans. R. Soc. Trop.* 105 (2011), pp. 214–9.

[70] A. Jeandron et al. "Water supply interruptions and suspected cholera incidence: a time-series regression in the Democratic Republic of the Congo". en. In: *PLoS Med.* 12 (2015), p. 1001893.

[71] D. Ganesan, S.S. Gupta, and D. Legros. "Cholera surveillance and estimation of burden of cholera". fr. In: *Vaccine* 38 (2020), pp. 13–7.

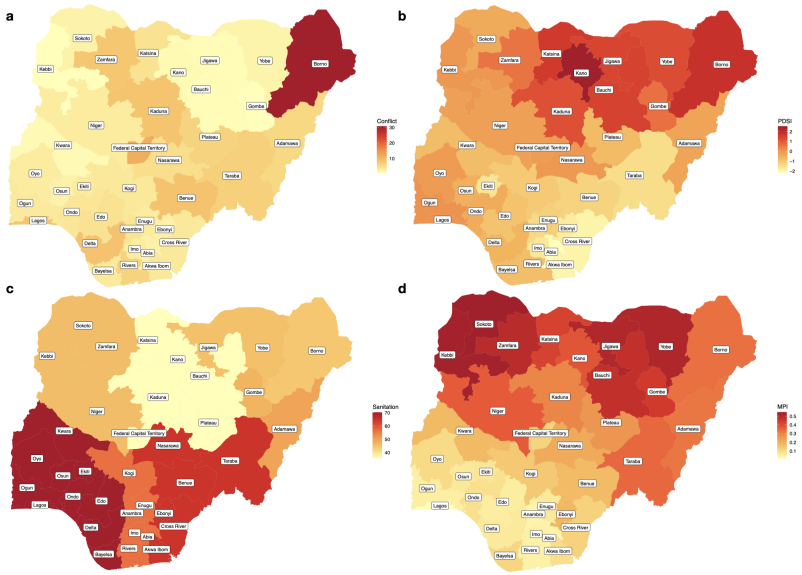
[72] A. Cori et al. "A new framework and software to estimate time-varying reproduction numbers during epidemics". In: *Am. J. Epidemiol.* 178.9 (2013), pp. 1505–1512.

[73] Global Task Force on Cholera Control. *Roadmap 2030*. en. 2020. URL: https://www.gtfcc.org/about-gtfcc/roadmap-2030/.

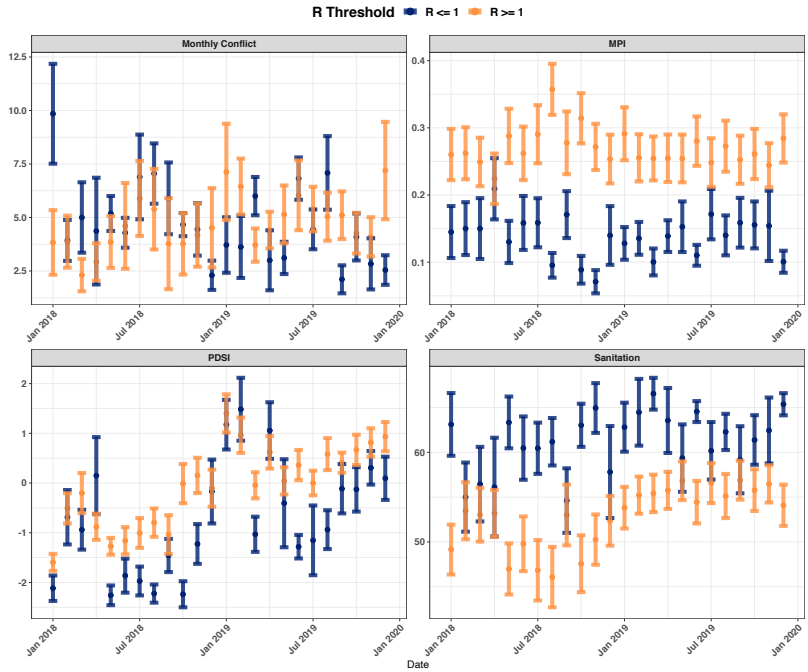
[74] WorldBank Climate Change Knowledge Portal. *Nigeria - Climatology*. 2020. URL: https://climateknowledgeportal.worldbank.org/country/nigeria/climate-data-historical.

Supplementary Material

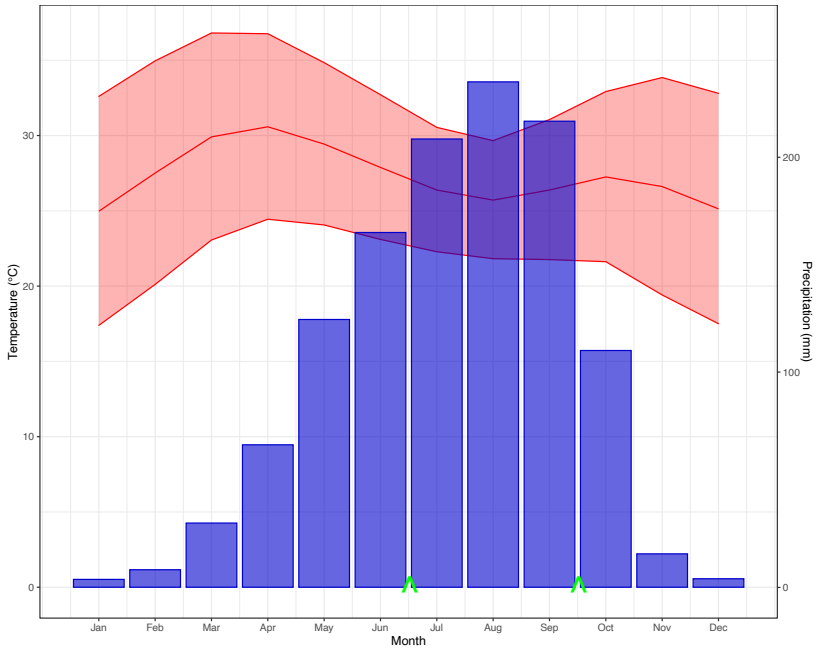
Supplementary Figures



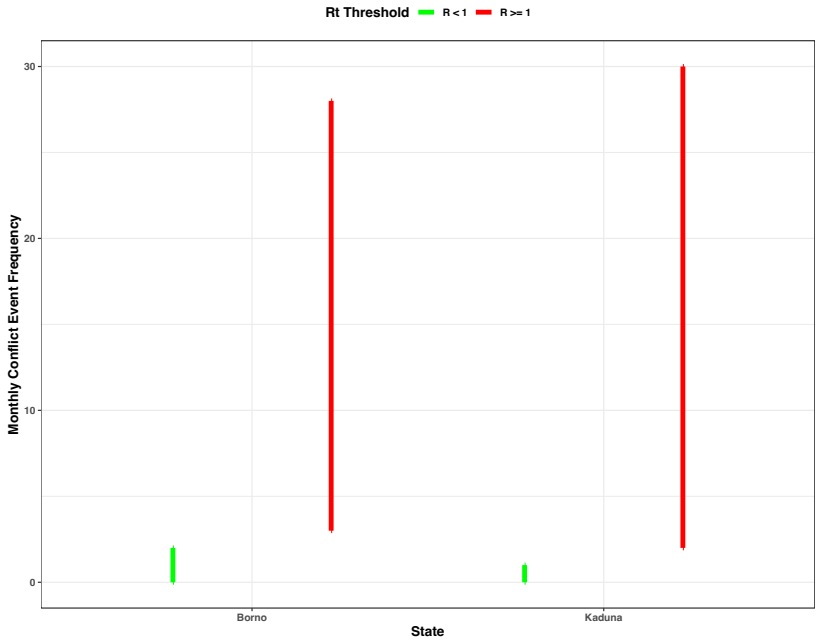
Supplementary Figure 5.1: Average values of the four covariates included in the best fit model. By state, covariates included: **a**, monthly conflict events, **b**, Palmers Drought Severity Index (PDSI), **c**, percentage access to sanitation and **d**, Multidimensional Poverty Index (MPI).



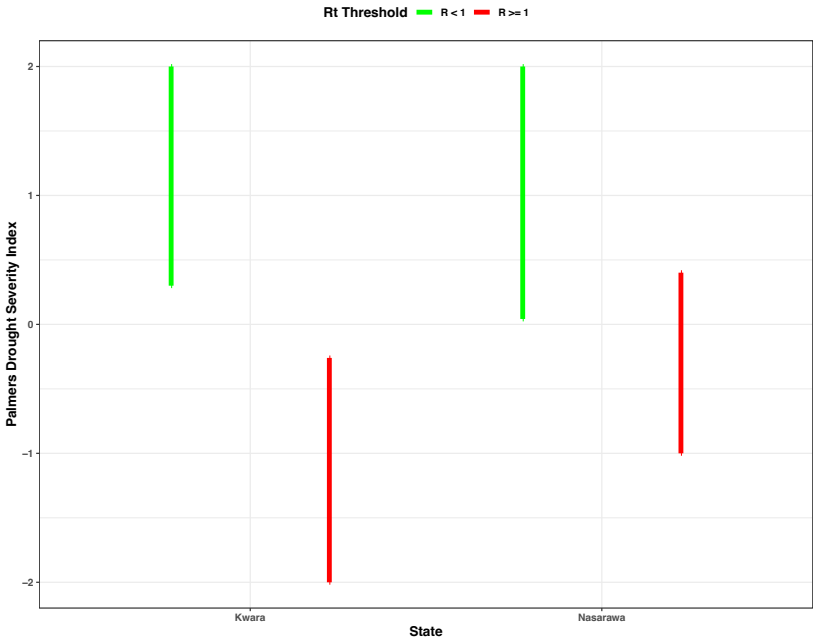
Supplementary Figure 5.2: Historical temporal trends between the best fit model covariates and the R_t thresholds ($R_t >= 1$, $R_t < 1$). The mean and standard error for the four covariates included in the best fit model for the full dataset split by month and R_t threshold.



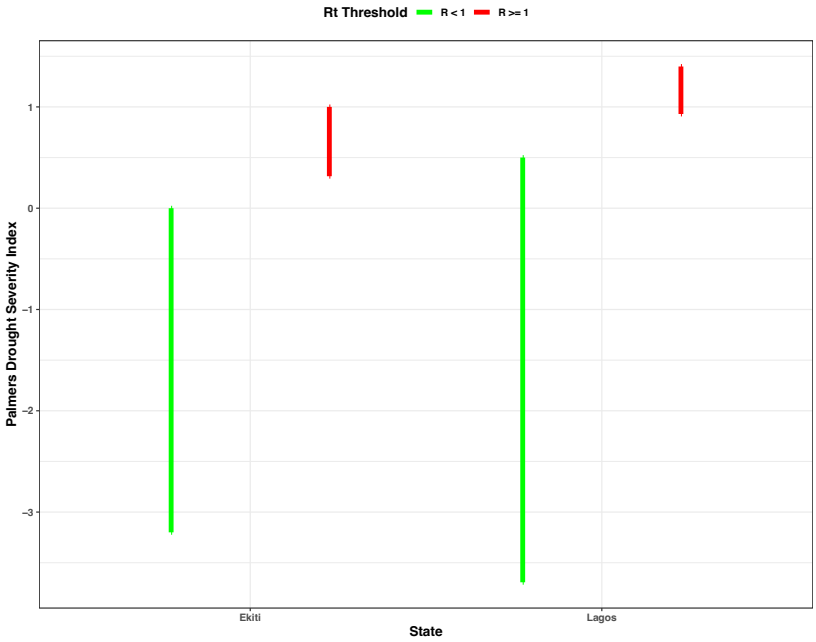
Supplementary Figure 5.3: Monthly Nigerian climatology of minimum, mean and maximum temperature (red lines and shading) and precipitation (blue bars) based on averages from 1991-2020, with the cholera peaks found in the NCDC dataset used here (green arrows) [74].



Supplementary Figure 5.4: Traffic-light system of cholera risk for conflict only for Borno and Kaduna. The other three (PDSI, Sanitation and MPI) covariate values were retained at the mean value for $R_t >= 1$ for the full dataset: Sanitation 41.1 and 40.4, MPI 0.33 and 0.31 and PDSI 1.95 and 1.49 for Borno and Kaduna, respectively.



Supplementary Figure 5.5: Traffic-light system of cholera risk for PDSI (drier conditions) only for Kwara and Nasarawa. The other three (Conflict, Sanitation and MPI) covariate values were retained at the mean value for $R_t >= 1$ for the full dataset: Sanitation 69.9 and 68, MPI 0.14 and 0.27 and Conflict 1 and 2 for Kwara and Nasarawa, respectively.

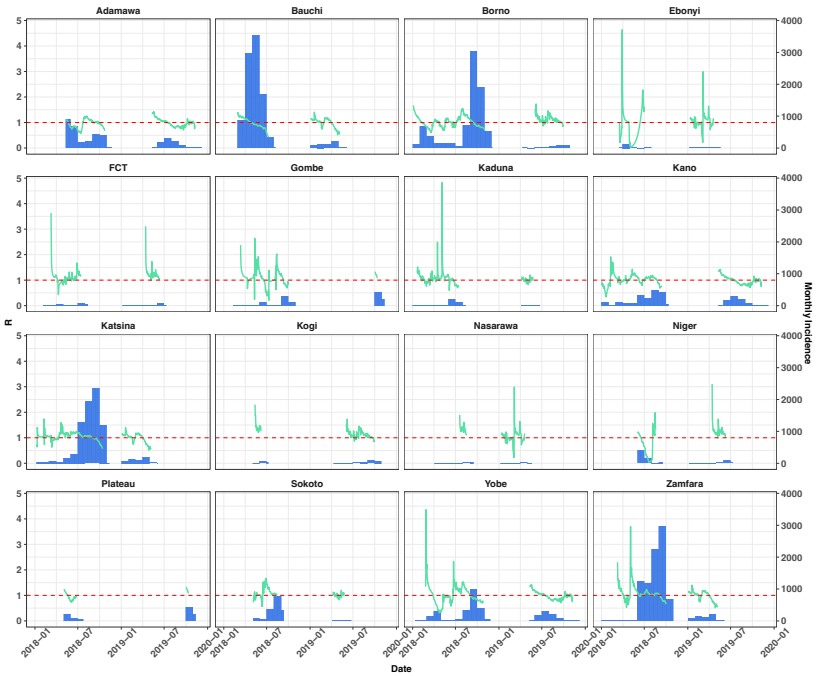


Supplementary Figure 5.6: Traffic-light system of cholera risk for PDSI (wetter conditions) only for Ekiti and Lagos. The other three (Conflict, Sanitation and MPI) covariate values were retained at the mean value for $R_t >= 1$ for the full dataset: Sanitation 70.5 and 70.5, MPI 0.086 and 0.016 and Conflict 2 and 10 for Ekiti and Lagos, respectively.

Supplementary Information

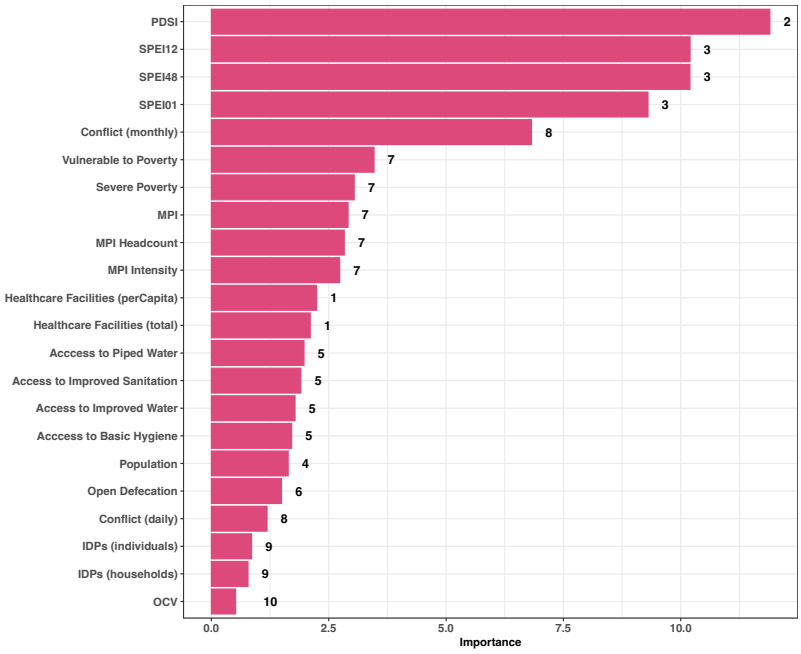
Supplementary Information 5.1: Sensitivity analysis using confirmed and suspected cholera cases. The analysis includes R_t calculations, variable importance and model fitting for the full dataset.

The data for the confirmed and suspected cholera cases had R_t calculated for 16 states (compared to 6 in the original model), which met the >40 cases thresholds for inclusion. The incidence and R_t calculations for the included states are shown below:



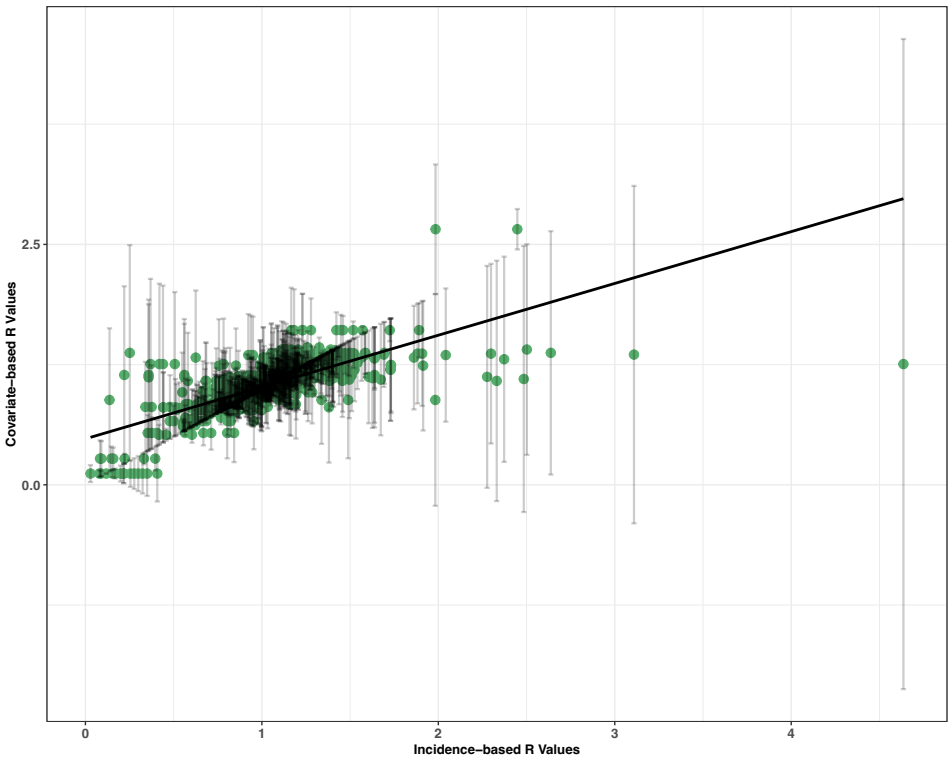
R_t values over monthly sliding windows (line) calculated from the daily incidence (bar) of cholera. The data used were suspected and confirmed cholera cases for 2018 and 2019 of states which met the threshold equal to or more than 40 cases.

The new dataset consisted of 5,627 data-points for variable importance and model fitting (compared to 279 in the original model). The variable importance plot (shown below) was similar to the variable importance for the original model (Figure 5.4), with only minimal changes in covariate importance order. In summary, IDPs, OCV and population were ranked much lower and some of the poverty metrics were much higher. This suggested that only small changes would be found in terms of the best fit model to the new dataset.



The variable importance for the twenty-two covariates considered for model inclusion. A serial interval of 5 days (with 8 days SD) was used and the numbers represent the clusters. SPEI01, 12, 48 - Standardised Precipitation Index calculated on 1, 12 and 48 month scale. PDSI - Palmers Drought Severity Index. MPI - Multidimensional Poverty Index. OCV – Oral cholera vaccination.

The new model did not improve model fit in terms of predictive power (shown below) and the same best fit model was selected. Any changes in the performance metrics were negligible (0.001 difference) and there was a slight decrease in correlation (0.71 in the new model), potentially due to the larger dataset creating greater variation. The sensitivity analysis using all the data, proved that the original model was robust and that the smaller dataset did not bias the results.



Incidence-based vs covariate-based R_t values for the best fit model fitted to the testing dataset. The error bars show mean absolute error and the line is a linear trend line.

Supplementary Information 5.2: Additional covariate selection for Chapter 5 using linear regression.

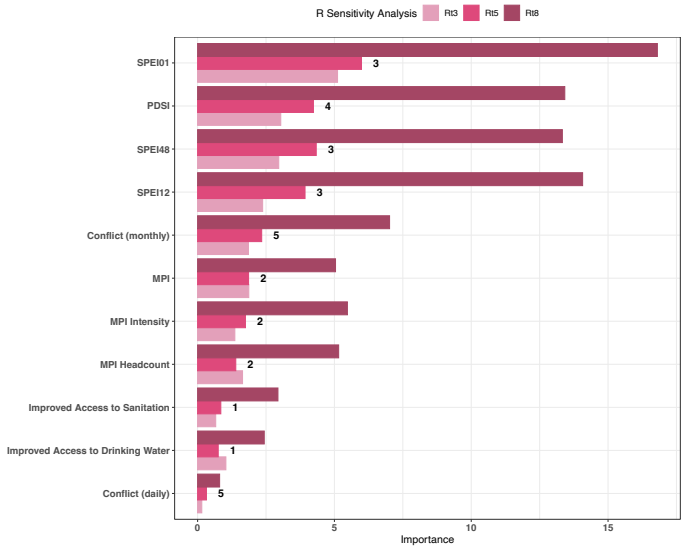
The same 21 covariates (conflict, drought IDPs, WASH, healthcare, population and poverty) analysed using variable importance were also run through an additional covariate selection process and stepwise analysis as developed by (and used in Chapter 3):

1. Garske, T. *et al.* Yellow fever in Africa: estimating the burden of disease and impact of mass vaccination from outbreak and serological data. PLoS Med. 11, e1001638 (2014).

2. Gaythorpe, K. A. M. *et al.* The global burden of yellow fever. Elife 10, e64670 (2021).

The selection process removes covariates that are not significantly associated with the outcome variable (Rt3, Rt5, Rt8) at p <0.1 using linear regression. It then clusters the remaining covariates based on the correction between them at an absolute pairwise correlation of above 0.75.

Ten were removed, either because they were not significantly associated with the outcome variable (R_t) or because they were too highly correlated with other covariates (healthcare facilities, piped water, open defecation, population, IDPs, severe poverty, vulnerable to poverty, basic hygiene). Eleven covariates remained and were grouped into five clusters, the clusters and variable importance of each covariate are shown below



The variable importance for the eleven remaining covariates after variable selection. All three serial interval values tested are shown (Rt3 - 3 days, Rt5 - 5 days, Rt8 - 8 days) and the numbers represent the clusters. SPEI01, 12, 48 - Standardised Precipitation Index calculated on 1, 12 and 48 month scale. PDSI - Palmers Drought Severity Index. MPI - Multidimensional Poverty Index.

Chapter 6

Cholera Past and Future in Nigeria: are the GTFCC 2030 Roadmap Targets Achievable?

Dissemination

A modified version of the full chapter is available as a pre-print at:

Charnley GEC, Yennan S, Ochu C, Kelman I, Gaythorpe KAM, Murray KA. Cholera past and future in Nigeria: are the Global Task Force on Cholera Control's 2030 targets achievable? *medRxiv* 2022;https://doi.org/10.1101/2022.12.06.22283154 [pre-print].

Abstract

Understanding and continually assessing the achievability of global health targets is very important in reducing the burden of disease and subsequent mortality. The GTFCC Roadmap aims to reduce cholera deaths by 90% and eliminate the disease from twenty countries by 2030. The Roadmap has three axes which focus on reporting, response and coordination. Chapter 6 aims to assess the likelihood that the GTFCC targets will be reached by 2030 in Nigeria and how the three axes could be strengthened to reach and exceed these goals. By analysing the historical data and creating future scenario projections, a 2050 target appears more realistic in Nigeria based on the results, although for the more urban and developed southern states, the 2030 targets could be reached. Improving reporting capacity and the monitoring of risk factors will help in continually assessing these targets and whatever target year is set, cholera elimination and control should aim to be achieved as quickly as possible. Additionally, long-term investments in WASH services, poverty alleviation and conflict resolution is particularly important in the northern states. A global coordinated effort is essential to control not just cholera but also to prevent pandemics in the future.

6.1 Introduction

Global target setting for both health and development are widely used strategies, examples include the Global Fund Strategy [1] and the United Nations SDGs [2]. The aim of these strategies are often to set a target in which all countries can work towards a common goal, with the hope that this will encourage commitment and knowledge and resource sharing. Development and health are interconnected and several global development goals will be fundamental to reaching health targets. Many countries and regions with the lowest levels of development in terms of poverty, education and health, also overlap with areas of high disease burden, particularly infectious diseases [3].

Despite the commitment of many governments and organisations to global scale strategies and targets, several fail to produce significant gains in development and health. Some notable examples include an inability to curb carbon emissions at a pace that will prevent a 2°C rise in temperatures (despite the signing of the Paris Agreement and Conferences of the Parties) [4] and shortcoming in the SDGs that have evoked only limited transformative policy impact towards the goals [5]. To reach these

ambitious targets, it is essential to continually assess both the successes and shortcomings to ensure that progress is continually made towards and beyond the target.

In 1992, the GTFCC was established as a global partnership of more than 50 institutions. The aim was to increase the capacity, tools and assistance for developing and implementing National Cholera Plans (NCPs). In 2017, the GTFCC launched "Ending Cholera: A Global Roadmap to 2030", encouraging partner organisations to sign the Declaration to End Cholera. The Roadmap focuses on three axes: early detection and response, interventions in cholera hotspots and effective coordination at all levels. The GTFCC believes that through these three axes there will be no more country-wide uncontrolled cholera outbreaks by 2030 and the disease will be eliminated from twenty countries, resulting in a 90% reduction in cholera deaths (Figure 6.1) [6].

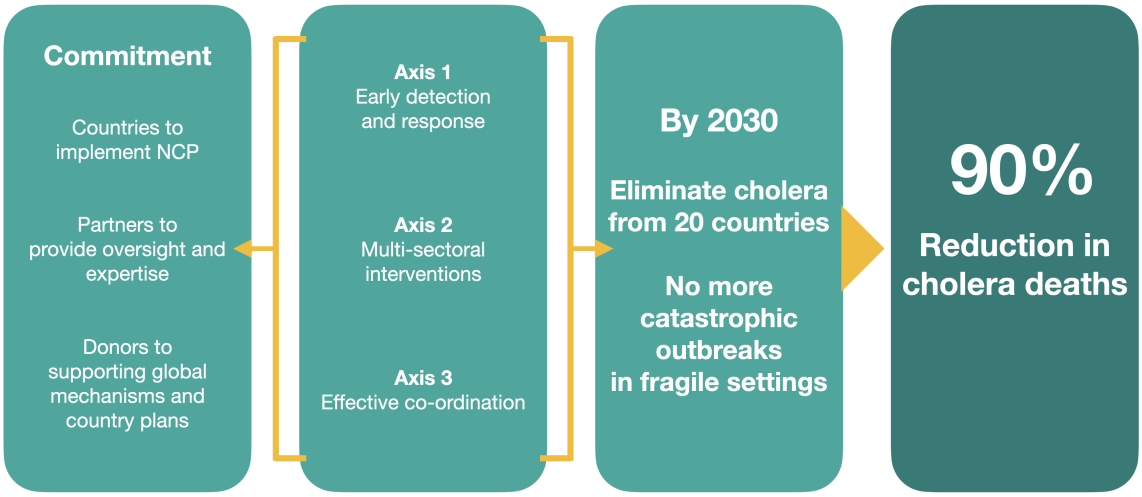


Figure 6.1: The GTFCC theory of change for cholera elimination. NCP - National Cholera Plan. Adapted from: [6].

As the 2030 goal approaches, whether these targets will, and can be achieved and what resources and investments are needed, are essential questions to answer. Understanding the achievability of these goals (Figure 6.1) will help countries and global partnerships to plan for 2030 and beyond, continuing to make gains in cholera prevention and control. Goals should be ambitious and encourage partners to strive for the best outcome possible but they also need to be clear and have significant commitment and motivation from governments, non-governmental organisations and the population.

There is an estimated 1.3 billion people at risk of cholera globally and approximately 2.86 million

annual cases (1.3-4.0 million) [7, 8], the majority of which are in sub-Saharan Africa and the Indian Subcontinent. The global overlap between poverty and cholera is well established and cholera control successes will heavily depend upon improving quality of life. There have been gains in cholera control at a local level, through socio-economic development, particularly access to water and sanitation, although this has been minimal at the global level [9]. Increased access to treatment has also been very important in controlling outbreaks such as ORS and OCV [10]. However, whether global improvements will be fast enough is highly uncertain as the 2030 target nears, while accounting for the fact that global crisis may cause regression of development and progress. For example, COVID-19 is estimated to have erased four years of progress against poverty and caused disaster-related deaths to rise sixfold [2].

Cholera forecasting is a useful tool in understanding the achievability of the 2030 targets. However, relatively few studies have used cholera projections, most of which taking a climate change focus [11, 12, 13] and even fewer studies have evaluated the likelihood of meeting the 2030 GTFCC targets using forecasting tools [14, 15]. Chapter 6 aims to address this research gap and to understand if the current pace of development will be enough to reach the 2030 targets in Nigeria. A number of future scenarios will be created and used to project cholera to 2070 in Nigeria. The scenarios will range from "best-case" to "worst-case", with both progress and regression from the current figures. Both national and sub-national scenario data will be be used to project cholera outbreak occurrence and transmission with the best fit models from Chapter 3 and Chapter 5.

The projections, along with historical cholera data and the environmental and socio-economic covariate data that were found most significant in this thesis, will be used to evaluate the achievability of the 2030 targets in Nigeria and make policy suggestions going forward. The research will highlight both successes and areas for improvement in the current strategy and where development needs accelerating. Incorporating the results of cholera research and modelling into policy is important to improve the relevance of epidemiological research and Chapter 6 aims to highlight how this can be achieved. The objectives for Chapter 6 are as follows:

1. Evaluate the historical data for cholera and the environmental and social risk factors found important in the models.

2. Project cholera transmission and outbreak occurrence to 2070 both nationally and sub-nationally

with varying degrees of global change.

3. Use the historical data and projections to assess the achievability of the 2030 targets in Nigeria based on the three axes outlined in the Roadmap.

6.2 Methods

6.2.1 Datasets

Historical

The number of reported cholera deaths (from suspected or confirmed cholera cases) were used in the historical analysis. Two data sources were chosen, as cholera data can vary widely and as previously stated has several limitation including over-reporting, under-reporting and reporting lags. The sources included the WHO's Global Health Observatory [16], which was used to fit the models in Chapter 3 (1970-2016) and the Global Health Data Exchange (GHDx) (1990-2016) [17]. The temporal and spatial scale of both sources were annual and national. Cholera deaths were transformed to deaths per 100,000, to make the analysis more comparable, as Nigeria is currently the most populous country in Africa at 211 million (2021), significantly higher than the next largest population, which is Ethiopia at 118 million.

The historical environmental and social data were taken from the previously used data sources of selected covariates. These included PDSI (national, annual, 1895-2016) [18], average temperature (°C, national, annual, 2000-2016) [19], water withdrawal per capita (national, annual, 1985-2010) [20], poverty headcount at <$1.25/day (national, annual, 1985-2018) [21], proportion of the population in extreme poverty (national, annual, 1981-2019) [22], MPI (administrative 1, annual, 2017-2018) [23], conflict events and fatalities (administrative 3, daily, 1997-2020) [23], percentage access to improved sanitation (administrative 1, annual, 2003-2017) [24].

Projected

WorldClim [25] was used for the projected environmental data at a monthly temporal granularity. WorldClim data are gridded and were transformed here to administrative level 1, as this spatial scale best captured the range of the other data sources. The data included minimum temperature and maximum temperature measured in degrees Celsius and precipitation (in mm). Projections were single values for 2050 and 2070 at three different RCP pathways (RCP4.5, 6.0 and 8.5, further explanation of the RCPs is available in Supplementary Information 3.2).

6.2.2 Historical Analysis

Pearson correlation coefficients (r) was used to understand the strength and direction of the linear relationship between the covariates analysed. The correlation coefficient is a ratio (0 to +1/-1) between variance and standard deviation and is expressed algebraically below. x_i and y_i are the values of the x and y variable in the sample and $\bar{x}$ and $\bar{y}$ are the means of the values of the x and y variables. One data source for cholera and poverty were selected for the correlations based on the length of the timescale (WHO cholera data and the proportion of the population in extreme poverty), to increase data completeness.

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

Time series of the historical data were used to illustrate how the covariates changed through time. To visualise and analyse the trends, linear regression trend lines and loess curves, ACF and AutoRegressive Integrated Moving Average (ARIMA) were used for the historical cholera death data. ARIMA is based upon three terms where p is the order (number of time lags) of the autoregressive model, d is the degree of differencing (the number of times the data have had past values subtracted), and q is the order of the moving-average model. ACF is a measure of autocorrelation and describes how well the present value of the series is related with its past values. ACF helps to illustrate the effect of time on the data and the appropriateness of ARIMA forecasting. The forecasting was based on a variation of the Hyndman-Khandakar algorithm, which combines unit root tests, minimisation of AIC and MLE, to return the best ARIMA forecast model according the AIC (R package "forecast", function auto.arima() [26])

6.2.3 Projections

The national projections for Nigeria used the best fit model from Chapter 3. In summary, this was a generalised linear model, with a binary outcome variable of cholera outbreak occurrence. The selected covariates included PDSI, average temperature, poverty headcount at <$1.25/day and freshwater withdrawal per capita. The sub-national projections used the best fit model from Chapter 5, this was a random forest model using cholera time-varying reproduction number as the outcome variable. The model included PDSI, MPI, percentage access to improved sanitation and monthly conflict event frequency.

Bootstrap resampling (10,000 samples) was used to obtain 95% confidence intervals for all projections. The projections were applied to the 2030 Roadmap targets and a new 2050 target delineated, based on the results. Twenty-fifty was chosen as the target year, due to it being halfway from the current target (2030) to the end of the projection period (2070), taking into consideration the full scope of the projections. Additionally, the results indicate 2050 as a realistic target, at the current pace of progress, and the two global initiatives the scenarios are based upon, the RCPs and SDGs, have newly developed 2050 targets [27, 28].

6.2.4 Projection Scenarios

Five projection scenarios were created to project cholera to 2070 using the two models from Chapter 3 and 5. Despite 2030 being the year of interest here, 2070 was selected as the projection end point, as this was the furthest time point the climate projections provided and allowed for further analysis and discussion beyond 2030. Single year projections were used, rather than averages, to take advantage of the single year environmental projections that were available and have proved accurate (see 1.2.1 Climate Change and Natural Hazards) [29].

The scenarios are based on several degrees of global change and are based upon the RCP scenarios and attainment of the SDGs. The RCPs and SDGs have been used throughout the thesis and cover a wide range of environmental and social scenarios, accounting for varying degrees of emissions reductions and socio-economic development. Several of the scenarios, especially the national scenarios are based on those created in Chapter 3. The scenarios are defined as follows:

- Scenario 1 (S1) - Best-case scenario meeting RCP4.5 and the SDGs
- Scenario 2 (S2) - Intermediate scenario between S1 and S3
- Scenario 3 (S3) - Minimal development and emissions reductions but progress is still made towards to SDGs and RCP8.5 is met
- Scenario 4 (S4) - Some regression from the current levels of sustainable development and increased emissions
- Scenario 5 (S5) - Worst-case scenario with significant regression in development and emissions increases

National

The national projection scenarios were taken from Chapter 3 for S1, S2 and S3, which were more optimistic and saw improvements to varying degrees (see 3.2.4 Projection Scenarios for more details). In summary, PDSI projections related to dataset averages and future conditions following the historical linear trend (PDSI $\sim$ Year, coefficient -0.014961, p-value 0.000512). Temperature was based on the WorldClim projections (RCP4.5, 6.0 and 8.5 for 2050 and 2070) and water withdrawal on varying degrees of meeting SDG6.4 (increase sustainable water-use efficiency across all sectors to address water scarcity). Nigeria is a high water resource and low withdraw country (Figure 3.2), therefore its more optimistic scenarios saw increased water use. Poverty scenarios were set at varying degrees of achieving SDG1.1 and 1.2, which state a 50% reduction in extreme (<\$1.25/day) poverty by 2030 and poverty eliminated by 2070.

For S4 and S5, temperature values for RCP6.0 and 8.5 were reached 20 years earlier (2050 by 2030 and 2070 by 2050). This was based on the assumption that emissions reductions would be even slower and therefore the radiative forcing threshold would be met sooner. For PDSI, previous literature was used to understand the full extent of drought changes across Nigeria, due to the limitations suggested in Chapter 3 (3.2.4). As previously stated, Africa is not projected to be significantly drier and historical studies suggest relative stability [30, 31, 32, 33]. However, projected PDSI changes are spatial heterogeneous and for Nigeria, drought projections suggest both stability [34] and drying [35, 36]. To account for the full range of projected changes and uncertainty stated in the literature,

the scenarios include both stability and significant drying to 2070. The poverty projections represent a 30% and 50% increase by 2070 for S4 and S5, respectively and for water withdrawal, instead of increasing from 2050 as seen in S1-S3, withdrawal would decrease by 10% (S4) and 20% (S5). Both poverty and water withdrawal scenarios were based upon the assumption that in S4 and S5 conditions would move in the opposite direction to the target. A summary of the national scenarios is shown in Table 6.1.

Table 6.1: National cholera projection scenarios for 2020-2070 at decadal intervals.

Scenario	Year	PDSI	Temperature	Poverty	Water withdrawal
Scenario 1	2020	Historical average	2016 value	2016 value	2016 value
Scenario 1	2030	Historical average	2016 value	50% decrease	2016 value
Scenario 1	2040	Historical average	2016 value	50% decrease	2016 value
Scenario 1	2050	Historical average	RCP4.5 2050	Median 2040-2070	20% increase
Scenario 1	2060	Historical average	RCP4.5 2050	Median 2040-2070	20% increase
Scenario 1	2070	Historical average	RCP4.5 2070	Elimination	20% increase
Scenario 2	2020	Median value (S1-S3)	2016 value	2016 value	2016 value
Scenario 2	2030	Median value (S1-S3)	2016 value	2016 value	2016 value
Scenario 2	2040	Median value (S1-S3)	2016 value	2016 value	2016 value
Scenario 2	2050	Median value (S1-S3)	RCP6.0 2050	50% decrease	10% increase
Scenario 2	2060	Median value (S1-S3)	RCP6.0 2050	Median 2050-2070	10% increase
Scenario 2	2070	Median value (S1-S3)	RCP6.0 2070	Elimination	10% increase
Scenario 3	2020	((Coefficient*4) + 2016 value)	2016 value	2016 value	2016 value
Scenario 3	2030	((Coefficient*10) + 2020 value)	2016 value	2016 value	2016 value
Scenario 3	2040	((Coefficient*10) + 2030 value)	2016 value	2016 value	2016 value
Scenario 3	2050	((Coefficient*10) + 2040 value)	RCP8.5 2050	2016 value	5% increase
Scenario 3	2060	((Coefficient*10) + 2050 value)	RCP8.5 2050	Median 2050-2070	5% increase
Scenario 3	2070	((Coefficient*10) + 2060 value)	RCP8.5 2070	50% decrease	5% increase
Scenario 4	2020	Historical average	2016 value	2016 value	2016 value
Scenario 4	2030	Historical average	RCP6.0 2050	2016 value	2016 value
Scenario 4	2040	Historical average	RCP6.0 2050	Median 2030-2070	2016 value
Scenario 4	2050	Median 2040-2070	RCP6.0 2070	Median 2030-2070	10% decrease
Scenario 4	2060	Median 2040-2070	RCP6.0 2070	Median 2030-2070	10% decrease
Scenario 4	2070	50% drier	RCP6.0 2070	30% increase	10% decrease
Scenario 5	2020	Historical average	2016 value	2016 value	2016 value
Scenario 5	2030	Median 2040-2070	RCP8.5 2050	2016 value	2016 value
Scenario 5	2040	Median 2040-2070	RCP8.5 2050	Median 2030-2070	2016 value
Scenario 5	2050	50% drier	RCP8.5 2070	Median 2030-2070	20% decrease
Scenario 5	2060	50% drier	RCP8.5 2070	Median 2030-2070	20% decrease
Scenario 5	2070	50% drier	RCP8.5 2070	50% increase	20% decrease

Sub-national

As the sub-national projections could account for the known spatial heterogeneity in PDSI projections stated above. Sub-national monthly PDSI projections were calculated using the projected 2050 and 2070 environmental data. First, PET (mm/day) was calculated with the temperature data and latitude using the Hargreaves method [37], where R_a is the mean extra-terrestrial radiation in mm/day, which

is a function of latitude and T represents daily air temperature in °C.

$$PET_{hargreaves} = 0.0023 * R_a * (T_{max} - T_{min})^{0.5} * (T_{mean} + 17.8)$$

PDSI is the output of a supply-and-demand model of soil moisture, which includes supply in the form of precipitation, demand from PET and fluxes, which represent how energy and water change in time and space, to give soil moisture. Soil moisture models can then be calibrated using station and satellite observations. A common critique of PDSI is that the behaviour of the index varies by location, making spatial comparisons difficult. The self-calibrated (scPDSI) method accounts for this by automatically calibrating the behaviour of the index at any location by replacing empirical constants in the index computation with dynamically calculated values [38]. Therefore, scPDSI was used to provide PDSI values for 2050 and 2070 for the three RCP scenarios (packages "SPEI" [39] & "scPDSI" [40]).

The scenarios for temperature and poverty followed the same pattern and targets as those in the national scenarios. Sanitation was based on SDG6.2 (achieve access to adequate and equitable sanitation and hygiene for all and end open defecation by 2030) and conflict scenarios were guided by SDG16.1 (Significantly reduce all forms of violence and related death rates everywhere). The SDGs for both sanitation and conflict are particularly ambiguous, regardless of this difficulty, the sanitation and conflict targets are based on a similar pattern to MPI, achieving universal access to sanitation and conflict elimination by 2070 in S1 and a 50% decrease in sanitation access and 50% increase in conflict events by 2070 in S5. A summary of the sub-national scenarios is shown in Table 6.2.

Table 6.2: Sub-national cholera projection scenarios for 2020-2070 at decadal intervals.

Scenario	Year	PDSI	MPI	Sanitation	Conflict
Scenario 1	2020	2020 value	2020 value	2020 value	2020 value
Scenario 1	2030	2020 value	50% decrease	50% increase	50% decrease
Scenario 1	2040	2020 value	50% decrease	50% increase	50% decrease
Scenario 1	2050	RCP4.5 2050	Median 2040-2070	Median 2040-2070	Median 2040-2070
Scenario 1	2060	Median 2050-2070	Median 2040-2070	Median 2040-2070	Median 2040-2070
Scenario 1	2070	RCP4.5 2070	Elimination	100% access	Elimination
Scenario 2	2020	2020 value	2020 value	2020 value	2020 value
Scenario 2	2030	2020 value	2020 value	Median 2020-2050	2020 value
Scenario 2	2040	2020 value	2020 value	Median 2020-2050	2020 value
Scenario 2	2050	RCP6.0 2050	50% decrease	50% increase	50% decrease
Scenario 2	2060	Median 2050-2070	Median 2050-2070	50% increase	Median 2050-2070
Scenario 2	2070	RCP6.0 2070	Elimination	50% increase	Elimination
Scenario 3	2020	2020 value	2020 value	2020 value	2020 value
Scenario 3	2030	2020 value	2020 value	2020 value	2020 value
Scenario 3	2040	2020 value	2020 value	2020 value	2020 value
Scenario 3	2050	RCP8.5 2050	2020 value	Median 2040-2070	2020 value
Scenario 3	2060	Median 2050-2070	Median 2050-2070	Median 2040-2070	Median 2050-2070
Scenario 3	2070	RCP8.5 2070	50% decrease	30% increase	50% decrease
Scenario 4	2020	2020 value	2020 value	2020 value	2020 value
Scenario 4	2030	RCP6.0 2050	2020 value	2020 value	2020 value
Scenario 4	2040	RCP6.0 2050	Median 2030-2070	2020 value	Median 2030-2070
Scenario 4	2050	RCP6.0 2070	Median 2030-2070	Median 2040-2070	Median 2030-2070
Scenario 4	2060	RCP6.0 2070	Median 2030-2070	Median 2040-2070	Median 2030-2070
Scenario 4	2070	RCP6.0 2070	30% increase	30% decrease	30% increase
Scenario 5	2020	2020 value	2020 value	2020 value	2020 value
Scenario 5	2030	RCP8.5 2050	2020 value	2020 value	2020 value
Scenario 5	2040	RCP8.5 2050	Median 2030-2070	2020 value	Median 2030-2070
Scenario 5	2050	RCP8.5 2070	Median 2030-2070	Median 2040-2070	Median 2030-2070
Scenario 5	2060	RCP8.5 2070	Median 2030-2070	Median 2040-2070	Median 2030-2070
Scenario 5	2070	RCP8.5 2070	50% increase	50% decrease	50% increase

6.3 Results

6.3.1 Historical Analysis

Figure 6.2 shows the time series for the WHO and GHDx cholera deaths data for 1970-2016, compared to the mean values for Africa (average of 0.1 cases/100,000). The time series highlights the historically high cholera burden in Nigeria, which is particularly evident in the GHDx data. The differences between the two datasets shows the importance of considering multiple cholera data sources for analyses. The GHDx data illustrates a steep decline from 16.0 cases/100,000 in 1991 to 1.8 cases/100,000 in 2017, with some plateauing periods. Whereas for the WHO data, cholera appears relatively stable over time (average of 0.5 cases/100,000), instead witnessing large peaks, particularly in 1971 (5.1

cases/100,000), 1991 (7.8 cases/100,000) and 1999 (1.7 cases/100,000). Furthermore, the difference in magnitude of the death rate should be noted here, with the GHDx data reporting much higher deaths per 100,000.

Figure 6.2: Time series of historical total annual cholera deaths per 100,000 of the population. GHDx [17] and WHO [16] are the total deaths rates (according to each source) per 100,000 for Nigeria. Africa is the mean annual death rates per 100,000 of the African population. Africa cholera data were from the WHO source [16] and population data for Nigeria and Africa from the UN Department of Economic and Social Affairs [41].

The linear trends and loess curves for the cholera data illustrates a flat trend in the WHO data, whereas the GHDx data shows a steady decrease (Supplementary Figure 6.1). ACF for both datasets show a slow decay (gradual decrease) to within the confidence interval bands (Supplementary Figure 6.2). The ACF decay suggests that the effect of time is not particularly significant in either of the cholera datasets. The weak effect of time was also illustrated in the ARIMA analysis, which showed a flat forecast, predicting the current rate of cholera deaths to continue.

For the environmental and social risk factors analysed here (meteorological drought, temperature, water withdrawal, poverty, conflict and sanitation), data incompleteness meant that calculating cor-

relations were difficult (Supplementary Figure 6.3). Assumptions and averages of the data had to be taken when fitting the models used (3.4.4 Limitations). Regardless of data limitations, cholera deaths had a strong negative correlation between sanitation access and a slight positive correlation with temperature from 1970 to 2016 (Figure 6.3).

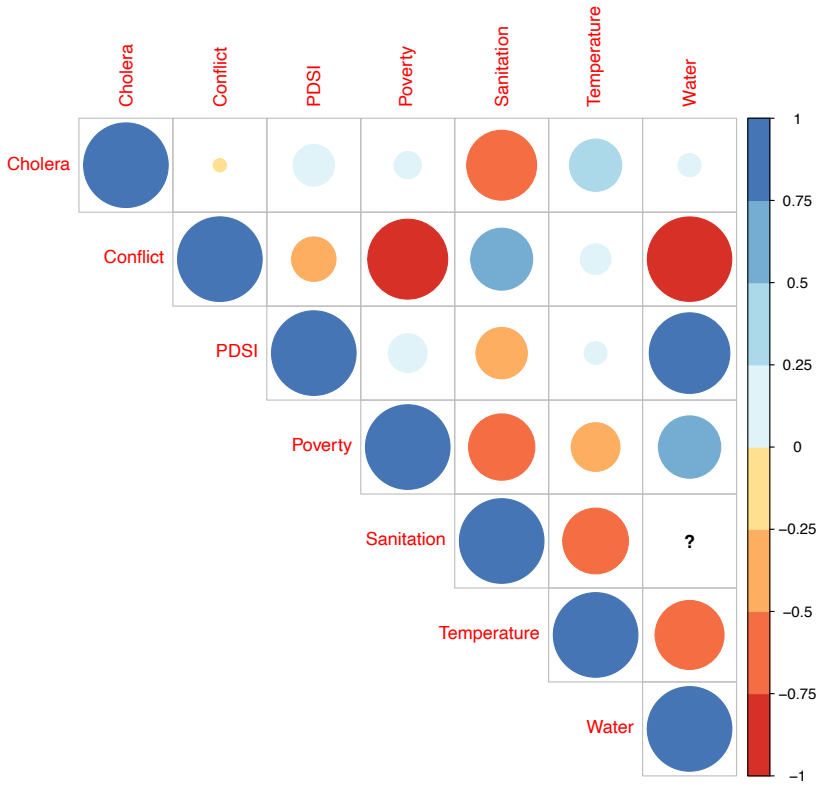


Figure 6.3: Correlation plot for the Pearson correlation coefficient of the six commonly selected covariates analysed here against the WHO cholera deaths data. Positive coefficients are blue suggesting a strong positive association between the corresponding row and column and negative coefficients are in red, suggesting a negative association. The '**?**' represents a negligible value due to data incompleteness.

6.3.2 Scenario Projections

National

Figure 6.4 shows the national cholera projections (in cholera outbreak occurrence, 0-1) to 2070 for the five scenarios, with 95% confidence intervals. The trends for Nigeria are similar to the continental projections presented in Figure 3.6, national cholera occurrence decreased where conditions improved in S1 to S3, starting at 0.95 for all scenarios to 0.83 for S1 and 0.92 for S3 by 2070. These changes were

minimal, especially considering the levels of development achieved in the S1 "best-case" scenario, and several of the confidence intervals overlap (the linear relationship, with standard error is presented in Supplementary Figure 6.4). For S4 and S5, where socio-economic and environmental conditions regressed, there was an increase in cholera outbreak occurrence from 0.95 to 0.98 for S4 and 0.99 for S5.

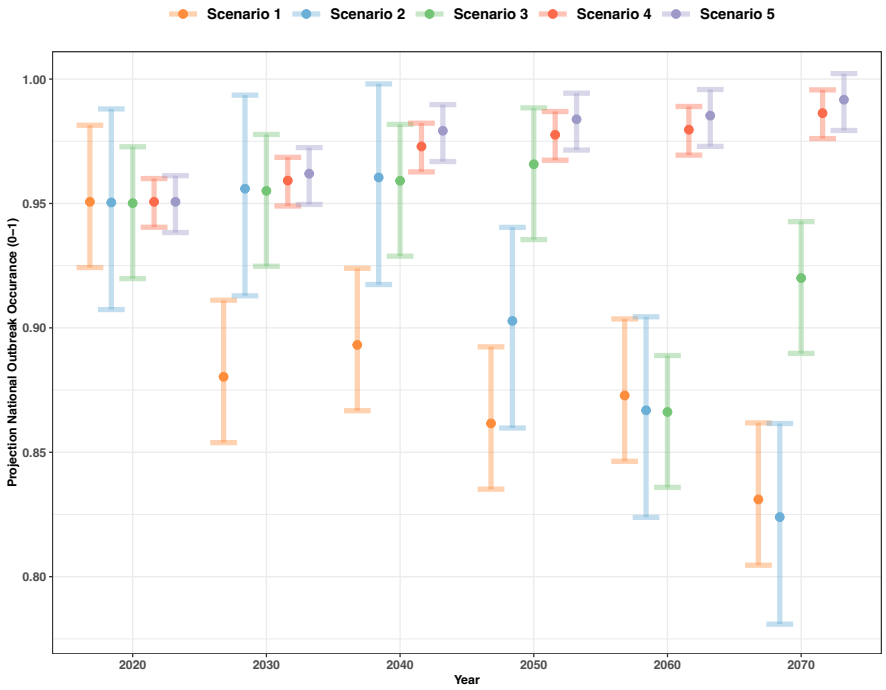


Figure 6.4: National cholera projections for Nigeria (with 95% confidence intervals), in cholera outbreak occurrence (0-1) to 2070, for the five scenarios. The scenarios were from most optimistic with strong progress towards emissions reductions and sustainable development (Scenario 1) to least optimistic, with regression in the current conditions (Scenario 5).

Sub-national

For the sub-national projections, measured in terms of cholera R_t, there were several spatial heterogeneities, resulting in much wider uncertainty than the national projections. Supplementary Figure 6.5 shows a national average R_t value with 95% confidence intervals for comparison with Figure 6.4. Generally, R_t values decreased through the three time points shown and the number of states with R_t values over 1 decreased (Figure 6.5 & Supplementary Figure 6.6) for S1, S2 and S3, where current conditions improved. For S4 and S5 (conditions regressed), the changes appear more complex, with some states faring better than others when faced with worsening social and environmental conditions.

The south of the country had particularly high R_t values in these less optimistic scenarios, whereas the north saw little change, and in some cases, a slight improvement. Despite the heterogeneity, by 2050 average projected R_t values for most regions (based on Nigeria's six geopolitical zones) of the country were less than 1 in Scenario 1, lending to the new proposed 2050 target (Supplementary Figure 6.6).

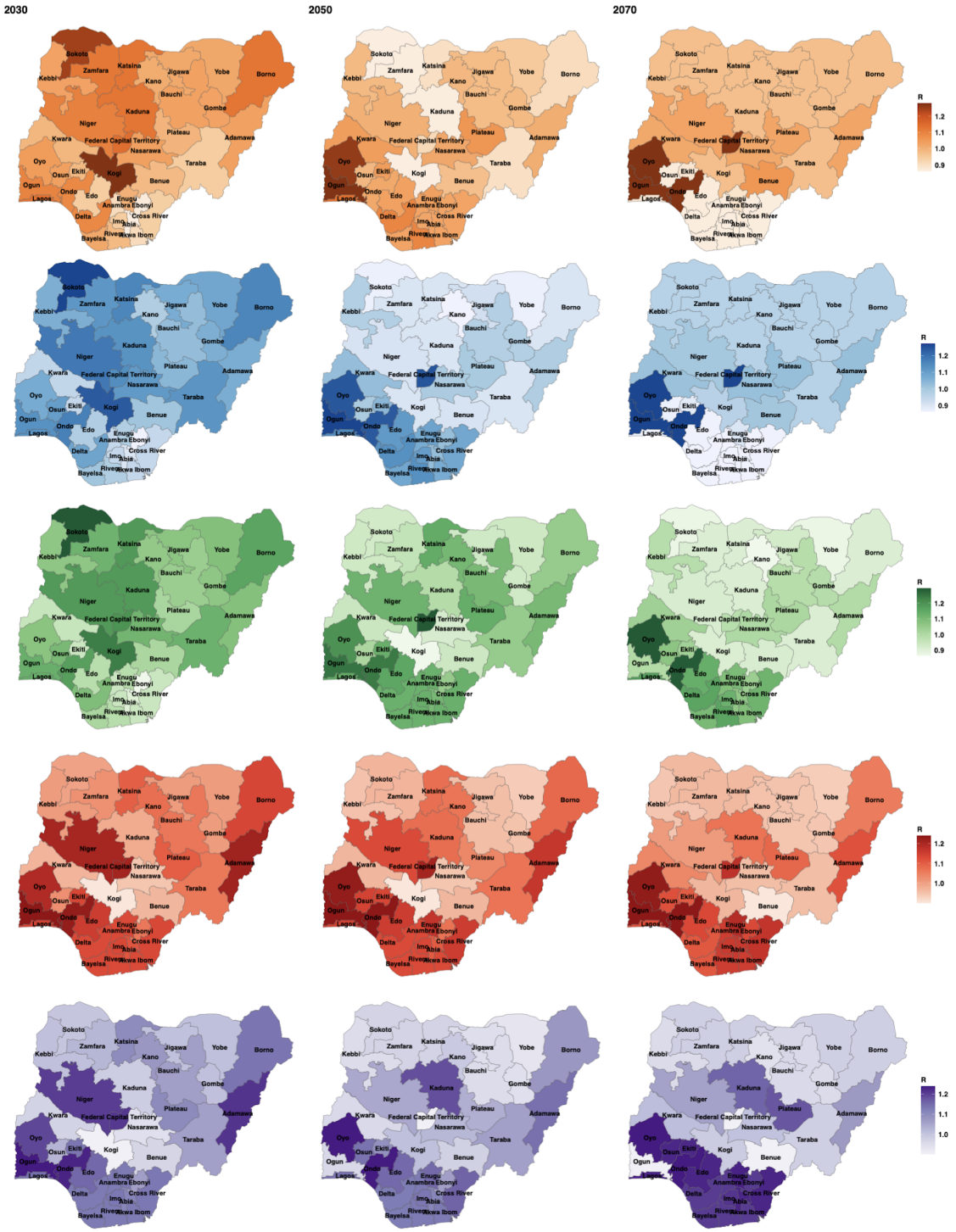


Figure 6.5: Sub-national cholera projections for Nigeria, in cholera reproduction number (R_t), for the five scenarios from most optimistic with strong progress towards emissions reductions and sustainable development (S1) to least optimistic, with regression in the current conditions (S5) (S1 - orange, S2 - blue, S3 - green, S4 - red and S5 - purple) at 3 of the decadal time points (2030, 2050 & 2070).

6.4 Discussion

Using the historical evidence and future projections, Chapter 6 has shed light on the future of cholera in Nigeria to 2030 and beyond. The historical data highlighted the historically high cholera burden in Nigeria, compared to the rest of Africa. The data also showed the wide range of cholera values that are reported and the value of evaluating multiple data sources. The cholera projections provided a more detailed understanding of future trends and if socio-economic development and climate change mitigation could reduce cholera in Nigeria to 2070. Both the national and sub-national projections showed decreases in cholera burden with the more optimistic scenarios (S1-S3). Under S4 and S5, the cholera burden worsened, despite the already high cholera outbreak occurrence, showing the need for continued development. Further discussion of the historical and future trends of cholera in Nigeria and how this could inform the GTFCC Roadmap are provided below.

6.4.1 Evidence from the Historical Data

The WHO data presented large outbreaks and peaks through the instrumental period, whereas the GHDx data steadily decreases but with higher cholera deaths overall. A potential explanation for this is that more sources are considered in the GHDx dataset, compared to the WHO source. The historical cholera trends do not suggest a significant increase or decrease based on previous burden. The ARIMA forecasting and ACF suggested a weak relationship with time in the two datasets analysed. The results show no clear trend to suggest whether cholera is increasing or decreasing and whether the 2030 targets can be met from these data alone.

The lack of relationship between cholera and time found here further highlights the importance of understanding cholera risk factors and their temporal changes over time. As stated in Chapter 1, there has been progress in terms of disease burden and sustainable development, but there are concerns over the pace of this progress. Understanding these risk factors going forward, in terms of monitored and collecting accurate data, will help inform cholera burden and identify hotspots and interventions.

Sanitation had the strongest correlation with the WHO dataset used in Chapter 6 and was selected in several of the models here over other metrics of WASH, including access to clean water and hand washing facilities. Progress has been made in the last 20 years (2000-2020) in terms of expanding access

to WASH services in Nigeria, with the national average percentage access to improved sanitation increasing from 52% to 62% and an expansion in access to improved drinking water by 31% [9]. However, if access to sanitation continues to increase by 10% every 20 years, this would result in 87,120,000 Nigerians without access by 2030 and therefore at a high risk of cholera (based on a 5% increase in access and a projected population of 264 million by 2030 [41]).

In Nigeria, there is a divide in terms of WASH and development between the northern and southern regions of the country (See Supplementary Figure 5.1) [42]. Northern states are generally more rural, and less development. For example, there is a 40% deficit in access to sanitation in the rural compared to the urban population. Nigeria's rural population comprises 47% (99,895,289, 2021) of the population, putting millions of people at risk of cholera in these potentially less developed areas [43]. However, in the last sixty years the rural population has decreased by 38% and with effective urban planning, this continued trend could significantly reduce cholera in Nigeria.

6.4.2 Evidence from the Scenario Projections

Similar to Chapter 3, the projections showed that continued progress towards and beyond the SDGs (in particular SDG1 and 6) and emissions reductions (contributing to PDSI and temperature) could help to improve global health and particularly cholera. The national scenarios showed clear trends in terms of both improvements from S1 (lowest cholera occurrence) to S3 and regression in S4 and S5 (highest cholera occurrence). However, by 2030 none of the national cholera outbreak occurrence projections were close to the 2030 target, achieving only a 12.6% decrease (vs 90% decrease in deaths needed). By 2070, with significant improvements in development and environmental protection, cholera was far from eliminated. Cholera eradication will likely take time, even with development improvements, due to the pathogen circulating in the population and environmental reservoirs [44, 45].

The sub-national projections were more optimistic in regards to the achievability of the GTFCC targets. Overall, for the most optimistic sub-national projections (S1) there was a decrease in R_t values to less than 1 for most states by 2050 and for most southern states by 2030 (Supplementary Figure 6.6). An explanation for this is that more time will be needed in the northern states to reach the required development for significant transmission reductions. However, the projections here and Chapter 5 (5.3.5 Spatial Heterogeneities) suggest that only marginal changes in development could have

a significant impact on transmission. The projections suggest that southern Nigeria could potentially reach the 2030 targets and eradication may be possible in the future, while northern Nigeria must be an area of prioritisation for development, cholera response and conflict resolution.

The sub-national projections additionally highlighted how vital it will be for the development and peace achieved in the south to continue or at a minimum, remain the same. As previously stated in the Results (6.3.2), the worsened conditions of S4 and S5 had a large impact on the southern states in terms of increasing cholera transmission. Whereas, for the northern states the changes were minimal, potentially due to the socio-economic development being already poorer and conflict already high in northern regions, resulting in a smaller change to cholera risk. Decreasing levels of peace and development would potentially be catastrophic in the southern states and to the overall cholera burden in the country.

6.4.3 An Update on the 2030 Targets and Roadmap

Overall, the achievability of the GTFCC targets at the current pace of cholera control and development appears unlikely to be met by 2030 in Nigeria. However, there has been progress towards the GTFCC goals and the southern states appear far more likely to meet these goals in the near term. A new proposed 2050 target, building on the GTFCC Roadmap and expanding on the three axes will now be delineated. Despite the new target being twenty years later, this is not encouraging complacency and a lack of urgency, and effort should still be made to reach the GTFCC goals as quickly as possible, no matter the target year set. The new targets will aim at bringing the northern states to the same levels of development and peace achieved in the south. The 2050 target was set based on the sub-national projection results showing that for the S1 scenario, most states R_t values were less than 1 by 2050.

Axis 1

Axis 1 largely focuses on surveillance and data, which are vital in target setting, allocation of resources and response. Improved surveillance may also help to reduce the cholera data inconsistencies found here. The axis focuses on early detection, but with cholera this can be difficult. Large numbers of infections are often mild or asymptomatic, meaning people do not know they have the disease, and therefore do not seek testing or treatment [8]. While in regions where cholera is circulating, diarrhoeal

disease burden caused by many other pathogens is also high (e.g., shigella, typhoid, dysentery) [46]. Other causative agents can complicate testing, as a positive test does not always mean that cholera is the causative agent of the diarrhoeal disease symptoms. Furthermore, there is reluctance to report cholera at both an individual and national level, due to restrictions on movement and trade and stigmatisation [7].

A method to help improve reporting in Nigeria, would be to offer incentives to test and report. Financial incentives have proved effective at improving health outcomes in Nigeria and are often cost-effective in the long-term as they prevent serious disease and morbidity [47, 48]. To reduce nosocomial transmission, modification and improvements to the cholera rapid diagnostic tests [49], allowing them to be used at home, may be helpful. Furthermore, the rapid diagnostic test for cholera has recently provided highly effective in Nigeria, outperforming laboratory culture [50]. The tests would need to be easy to use and report, inexpensive and widely available. At home testing may also reduce testing hesitancy, to avoid restrictions and stigmatisation [48], while allowing people to make informed decisions about behaviour modifications to limit transmission.

Emphasis is needed both at a government and academic level on improving data quantity and quality. Understanding reporting effort and the accuracy and precision of data are key areas of future research in order to fully understand how well the current data are representing cholera burden. At a global level, a metric of reporting effort would help when comparing disease data that have been collected across multiple countries and therefore with different methods and uncertainty.

Furthermore, risk factor data are needed to fully understand disease dynamics and plan for effective response and interventions. The results here suggest that improvements in tracking poverty and sanitation would be beneficial areas of prioritisation. In Nigeria, benefit could also be gained from testing environmental reservoirs, such as major lakes and rivers, which are known to be used for washing and drinking and can be fundamental in cholera transmission [51, 52]. Water testing would increase the understanding of the environmental burden and would be useful for both research and to understand local risk factors.

Axis 2

Axis 2 (cholera interventions) is arguably the most important area for reaching the GTFCC goals and several other health targets. The Roadmap highlights the need for long-term sustainable WASH implementation and strengthening of healthcare systems to anticipate cholera outbreaks (e.g., capacity building of staff, resources, diagnostics, education and societal engagement and emergency WASH intervention). However, the GTFCC Roadmap and previous research on cholera interventions heavily focuses on outbreak response [15], rather than socio-economic development.

The Roadmap suggests that interventions should target states most at risk, with the analysis presented here suggesting northern states as a priority. Additionally, healthcare should be strengthened more generally, with greater resources and service availability, making healthcare an attractive career option (e.g., fair pay and benefits) to ensure sufficient human resources [53, 54]. For example, Sierra Leone, a country similar to Nigeria in terms of geography and challenges, began their Free Health Care Initiative in 2010, which has helped bring about important health system gains that have particularly benefits vulnerable people [55]. Development planning and targets must also consider that global crises can cause regression of progress, increasing the need to strive beyond health targets.

Designating significant financial resources on outbreak response, is not a cost-effective way of reaching cholera targets, although fundamental to reducing mortality in outbreaks. More emphasis needs to be placed on improving peoples' quality of life, lifting them out of poverty, providing them with basic services and empowering them to improve their own health through resources and education. In the absence of this development, outbreaks will continue to occur, and financial resources will be spent in a reactionary way, rather than with careful planning towards continued progress.

Axis 3

Axis 3 of the Roadmap involves commitment and coordination on a global level, across many sectors. NCDC currently works across multiple levels of the national system and has a detailed response plan for diarrhoeal (including cholera) outbreaks titled, "Preparedness and Response to Acute Watery Diarrhoea Outbreaks" [56]. However, for NCDC to work effectively, there must be a functioning national system and currently bureaucratic and corruption challenges threaten this. Nigeria has seen regression in its anti-corruption progress since 2016, currently scoring 154/180 countries in the

Corruption Perceptions Index [57]. At a global level, a "One World - One Health" approach is needed to prevent pandemics and achieve the GTFCC targets at this level. Recent pandemics and global outbreaks (e.g., COVID-19 and monkeypox) have shown the catastrophic results of countries not working together in a global effort to control disease [58].

NCDC have a designated team working on cholera elimination as a priority within the country. Continued and increased funding to NCDC will be vital for them to continue their work towards cholera control. A barrier to achieving this is healthcare spending, which is comparatively low in Nigeria. Health expenditure is currently at a level not seen since 2002, at 3.03% of Gross Domestic Product. Only fourteen countries globally and five countries in Africa spend less on healthcare than Nigeria [59]. Health needs to be a greater priority in terms of policy and government spending, in order to tackle not just cholera but several other diseases and health challenges.

Nigeria has made several gains in weakening the Boko Haram stronghold in the northeastern states, both in terms of territory and numbers. However, the conflict continues to threaten Nigeria's national security and several previous studies have suggested the negative impacts of conflict on health [60, 61, 62, 63]. Bottom-up stabilisation efforts are working to address local level drivers of insecurity, including strengthening local conflict prevention, restoring governance and services and fostering social cohesion. Reducing regional inequities in Nigeria will not only help reduce the cholera burden, but also increase trust in the government, reducing population vulnerability to extremist recruitment. While at a regional level, the Lake Chad Basin Commission and African Union Commission have highlighted short-, medium- and long-term stabilisation, resilience and recovery needs [64, 65, 66].

In summary, Figure 6.6 illustrates the current Roadmap, and summarises the suggestions made here to improve cholera control beyond 2030 and achieve the GTFCC targets in Nigeria by 2050, while still striving to achieve them as close to 2030 as possible. Here, even with the most optimistic scenarios (Scenario 1 and 2), the uncertainty of the projected changes in cholera at a national and sub-national level (Supplementary Figure 6.5), make it difficult to say with certainty when a 90% reduction in deaths can be achieved in Nigeria. However, with the suggestions and improvements stated above and summarised below, cholera deaths will undoubtedly reduce and therefore with time, the 90% reduction target met.

Axis 1 Early detection and response	Axis 2 Multisectoral approach	Axis 3 Mechanism of coordination
• Early warning surveillance systems • Pre-positioning stocks • Preparedness of WASH systems • Preparedness of the healthcare system • Community engagement	• Identify hotspots and priority areas • Control measures (surveillance, WASH, health care systems, OCV, community engagement and collaboration)	• Nationally-led cross-sectoral programs • GTFCC as a strong coordination platform
• Incentives to report • Rapid at-home tests • Monitoring of the environment and known risk factors	• Shift from outbreak response to long-term sustainable development • Empower local populations to care for their health through poverty alleviation, services and education • Effective urban planning	• Global commitment and coordination for a "One World - One Health" approach to pandemics • Conflict resolution • Increase health expenditure
2030	**90% reduction in deaths and no large-scale outbreaks**	
2050	**Targeting the northern states and continued development in the south**	

Figure 6.6: The 2030 GTFCC Roadmap for cholera elimination (black) with additional suggestions and areas of prioritisation (blue) for 2050 in Nigeria.

6.4.4 Limitations

Data incompleteness and inconsistencies were issues when trying to evaluate the historical data. As stated above, improving surveillance and a greater effort to collect data on cholera risk factors will be very important for target setting and resource allocation and prevent duplication of services [67]. More effort needs to be made globally to collect and collate socio-economic data not just for research purposes but to understand where people need assistance. Multiple data sources were used here to try and account for this issue, and in future cholera research, using sensitivity analysis and testing cholera assumptions across multiple data sources is one method to understand these differences.

All scenario projections have limitations, due to the uncertainty in trying to predict future conditions (particularly human behaviour), along with the limitations of the models (3.4.4 Limitations & 5.4.3 Limitations). The wide range of future scenarios helps to account for some of this uncertainty but will still not be sufficient in capturing all potential future environments. For example, the scenarios here are uni-directional, either getting better or worse from current conditions. All social and environmental drivers either getting better or worse is unlikely, with some metrics improving and some worsening. To add further complexity, these changes could be spatially heterogeneous. Regardless of these limitations, this should not discourage scenario projection analysis, as it is still useful and valid in terms of

understanding future changes and helping to inform cholera prevention and policy.

A further limitation is the different outcome variables and methodology of the two projections. It could be argued that two model types (GLM and RF) and two outcome variables (outbreak occurrence and R_t), make the results here difficult to compare. The GTFCC targets largely focus on reducing cholera deaths, which is why deaths (death rate) were chosen as the cholera metric for the historical data analyses. Although the discrepancies create difficulties in interpretation, in order to reach the 2030 targets and subsequently reduce deaths by 90%, global burden will have to substantially decrease, regardless of the metric used. Therefore, the projections are still useful in presenting the required decrease in burden.

6.4.5 Conclusion

In conclusion, Chapter 6 highlights the importance of and how modelling studies can be used to inform cholera policy. Using the knowledge gained through the work in this thesis, the GTFCC targets look difficult to achieve by 2030 in Nigeria. There is a vital need for continued investment in long-term development, especially in northern Nigeria. The progress and achievements already made in Nigeria in terms of development have not just improved the quality of life of the population but also had a positive effect on cholera and several other diseases. Despite the financial capital needed to improve healthcare, WASH and education, these interventions are cost-effective [68, 69, 70] due to their wide-reaching impacts.

Nigeria currently has one of the largest cholera burdens globally, making it a critical area of study and several of the policy suggestions here could be applied in other countries and regions. Nigeria is also one of the largest African economies, has the largest populations and is arguably set to see the greatest levels of development in the coming decades (due to its vast natural and human resources). If the GTFCC targets are met in Nigeria, this will reduce the risk of cholera for hundreds of millions of people and greatly reduce the global burden of diarrhoeal disease mortality in children under 5 years.

This work has shown the threat of disasters to impact disease and their associated risk factors. Disasters have the ability to erode health and development gains and improvements and progress to go beyond the goals and targets set would be highly beneficial in combating this. Continued progress in development (especially at an accelerated pace) and sustainable urban planning in Nigeria, would

see great improvements in its cholera burden and health status in the coming decades and help the country withstand the shocks of disasters.

References

[1] The Global Fund. *Strategy*. en. https://www.theglobalfund.org/en/strategy/. 2022.

[2] United Nations. *The 17 Goals*. en. https://sdgs.un.org/goals. 2015.

[3] Z. A. Bhutta et al. "Global burden, distribution, and interventions for infectious diseases of poverty". In: *Infect. Dis. Poverty* 3.1 (2014), pp. 1–7.

[4] E. Masood and J. Tollefson. *'COP26 hasn't solved the problem': scientists react to UN climate deal*. en. https://www.nature.com/articles/d41586-021-03431-4. 2021.

[5] F. Biermann et al. "Scientific evidence on the political impact of the Sustainable Development Goals". In: *Nature Sustain.* (2022), pp. 1–6.

[6] Global Task Force on Cholera Control. *Roadmap 2030*. en. 2020. URL: https://www.gtfcc.org/about-gtfcc/roadmap-2030/.

[7] M. Ali et al. "The global burden of cholera". en. In: *Bull. World. Health Organ.* 90 (2012), pp. 209–218.

[8] M. Ali et al. "Updated global burden of cholera in endemic countries". en. In: *PLoS Neglect. Trop. Dis.* 9 (2015), p. 0003832.

[9] H. Richie and M. Roser. *Clean Water and Sanitation*. en. https://ourworldindata.org/clean-water-sanitation. 2021.

[10] J. Ebob. "Cholera Prevention and Control Strategies; A Global Overview". In: *WHL* (2020).

[11] A. F. Abdussalam. "Potential future risk of cholera due to climate change in northern Nigeria". In: *Afr. Res. Rev.* 11 (1 2017), pp. 205–18.

[12] E. C. Lee et al. "The projected impact of geographic targeting of oral cholera vaccination in sub-Saharan Africa: a modeling study". In: *PLoS Med.* 16 (12 2019), e1003003.

[13] S. L. Trærup, R. A. Ortiz, and A. Markandya. "The costs of climate change: a study of cholera in Tanzania". In: *Int. J. Environ. Res. Public Health* 8 (12 2011), pp. 4386–405.

[14] D. Legros. "Global cholera epidemiology: opportunities to reduce the burden of cholera by 2030". en. In: *J. Infect. Dis.* 218.Suppl 3 (2018).

[15] M. T. Islam et al. "A blueprint for eliminating cholera by 2030". In: *Nature Med.* (2022), pp. 1–3.

[16] World Health Organization. *The Global Health Observatory*. en. 2020. URL: https://www.who.int/data/gho.

[17] IHME GHDx. *Global Health Data Exchange*. en. 2016. URL: https://ghdx.healthdata.org.

[18] N.C.A.R. *Dai Global Palmer Drought Severity Index (PDSI)*. en. 2020. URL: https://rda.ucar.edu/datasets/ds299.0/index.html#!sfol-wl-/data/ds299.0.

[19] E.C.M.W.F. *ERA5*. it. 2020. URL: https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5.

[20] H. Ritchie. *Water Use and Stress*. en. 2017. URL: https://ourworldindata.org/water-use-stress.

[21] United Nations Development Programme. *Human Development Data (1990-2018)*. en. 2018. URL: http://hdr.undp.org/en/data#.

[22] M. Roser and E. Ortiz-Ospina. *Global Extreme Poverty*. en. 2019. URL: https://ourworldindata.org/extreme-poverty.

[23] H.D.X. *The Humanitarian Data Exchange*. en. 2021. URL: https://data.humdata.org.

[24] Joint Monitoring Programme. *Nigeria*. 2020. URL: https://washdata.org.

[25] WorldClim. *Future climate data*. it. 2020. URL: https://worldclim.org/data/cmip6/cmip6climate.html.

[26] R. Hyndman et al. *forecast: Forecasting functions for time series and linear models*. R package version 0.1.1. 2021. URL: https://pkg.robjhyndman.com/forecast/.

[27] United Nations. *Vision 2045*. en. 2022. URL: https://vision2045.com.

[28] Intergovernmental Panel on Climate Change. *Global Warming of 1.5℃*. en. 2018. URL: https://www.ipcc.ch/sr15/.

[29] A. Buis. *Study Confirms Climate Models are Getting Future Warming Projections Right*. 2020. URL: https://climate.nasa.gov/news/2943/study-confirms-climate-models-are-getting-future-warming-projections-right/.

[30] Y. Yang et al. "Little change in Palmer Drought Severity Index across global land under warming in climate projections". en. In: *Hydrol. Earth Syst. Sci. Discuss.* (2020).

[31] J. Sheffield, E.F. Wood, and M.L. Roderick. "Little change in global drought over the past 60 years". en. In: *Nature* 491.7424 (2012), pp. 435–438.

[32] R. Touchan et al. "Long term context for recent drought in northwestern Africa". en. In: *Geophys. Res. Lett* 35.13 (2008).

[33] D. Verschuren, K.R. Laird, and B.F. Cumming. "Rainfall and drought in equatorial east Africa during the past 1,100 years". en. In: *Nature* 403.6768 (2000), pp. 410–414.

[34] A. J. Oloruntade et al. "Analysis of meteorological and hydrological droughts in the Niger-South Basin, Nigeria". In: *Glob. Planet. Change* 155 (2017), pp. 225–233.

[35] M. S. Shiru et al. "Projection of meteorological droughts in Nigeria during growing seasons under climate change scenarios". In: *Sci. Rep.* 10.1 (2020), pp. 1–18.

[36] A. T. Ogunrinde et al. "Evaluation of the impact of climate change on the characteristics of drought in Sahel Region of Nigeria: 1971–2060". In: *Afr. Geogr. Rev.* 40.2 (2021), pp. 192–210.

[37] G. H. Hargreaves. "Defining and using reference evapotranspiration". In: *J. Irrig. Drain. Eng.* 120 (6 1994), pp. 1132–9.

[38] N. Wells, S. Goddard, and M. J. Hayes. "A self-calibrating Palmer drought severity index". In: *J. Clim.* 17.12 (2004), pp. 2335–2351.

[39] S Beguería and S. M. Vicente-Serrano. *SPEI: Calculation of the Standardised Precipitation-Evapotranspiration Index*. R package version 1.7. 2017. URL: https://CRAN.R-project.org/package=SPEI.

[40] R. Zhong et al. *scPDSI: Calculation of the Conventional and Self-Calibrating Palmer Drought Severity Index*. R package version 0.1.3. 2018. URL: https://CRAN.R-project.org/package=scPDSI.

[41] United Nations Department for Economic and Social Affairs. *Population Dynamics. World Population Prospectus*. en. 2019. URL: https://population.un.org/wpp/.

[42] A. Ojo and O. Ojewale. *Nigeria's Urbanisation History, Trends, Drivers and Implications*. Springer, 2019, pp. 13–58.

[43] World Bank. *Urban population (% of total population) - Nigeria*. en. 2021. URL: https://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS?locations=NG.

[44] S. Almagro-Moreno and R. K. Taylor. "Cholera: environmental reservoirs and impact on disease transmission". In: *Microbiol. Spectr.* 1.2 (2013), pp. 1–2.

[45] M. S. Islam et al. "Environmental reservoirs of Vibrio cholerae". In: *Vaccine* 38 (2020), A52–A62.

[46] A.A. King et al. "Inapparent infections and cholera dynamics". en. In: *Nature* 454 (2008), pp. 877–80.

[47] E. N. Okeke, I. S. Abubakar, and Z. Wagner. *Reducing Child Deaths in Nigeria Through Contingent Cash Transfers to Expectant Mothers*. en. 2020. URL: https://www.rand.org/pubs/research_briefs/RB10114.html.

[48] D. Ganesan, S.S. Gupta, and D. Legros. "Cholera surveillance and estimation of burden of cholera". fr. In: *Vaccine* 38 (2020), pp. 13–7.

[49] Centres for Disease Control and Prevention. *Crystal® VC Rapid Diagnostic Test (RDT) Procedure*. en. 2021. URL: https://www.cdc.gov/cholera/crystal-vc.html.

[50] K. Elimian et al. "Epidemiology, diagnostics and factors associated with mortality during a cholera epidemic in Nigeria, October 2020–October 2021: a retrospective analysis of national surveillance data". In: *BMJ open* 12.9 (2022), e063703.

[51] G. Bwire et al. “Environmental surveillance of Vibrio cholerae O1/O139 in the five african great lakes and other major surface water sources in Uganda”. In: *Front. Microbiol.* 9 (2018), p. 1560.

[52] R.L. Shapiro et al. “Transmission of epidemic Vibrio cholerae O1 in rural western Kenya associated with drinking water from Lake Victoria: an environmental reservoir for cholera?” In: *Am. J. Trop. Med.* 60.2 (1999), pp. 271–276.

[53] M. D. Ughasoro, D. O. Esangbedo, and I. M. Udorah. “Health-care workers’ perspectives on preparedness of health-care facilities for outbreak of communicable diseases in Nigeria: a qualitative study”. In: *Am. J. Trop. Med. Hyg.* 100.4 (2019), p. 1022.

[54] A. K. Ager et al. “Health service resilience in Yobe state, Nigeria in the context of the Boko Haram insurgency: a systems dynamics analysis using group model building”. In: *Confl. Health* 9.1 (2015), pp. 1–14.

[55] S. Witter et al. “The free healthcare initiative in Sierra Leone: evaluating a health system reform, 2010–2015”. In: *Int. J. Health Plan. Manag.* 33.2 (2018), pp. 434–448.

[56] Nigeria Centre for Disease Control. *NIGERIA - PREPAREDNESS AND RESPONSE TO ACUTE WATERY DIARRHOEA OUTBREAKS*. en. 2017. URL: http://www.plateformecholera.info/index.php/country-monitoring/nigeria/152-wca/strategic-framework/library-of-national-plans/481-nigeria-preparedness-and-response-to-acute-watery-diarrhoea-outbreaks.

[57] Transparency International. *CORRUPTION PERCEPTIONS INDEX*. en. 2021. URL: https://www.transparency.org/en/cpi/2021/index/ssd.

[58] P. Calistri et al. “The components of ‘one world–one health’ approach”. In: *Transbound. Emerg. Dis.* 60 (2013), pp. 4–13.

[59] World Bank. *Current health expenditure (% of GDP) - Sub-Saharan Africa*. en. 2019. URL: https://data.worldbank.org/indicator/SH.XPD.CHEX.GD.ZS?end=2019&locations=ZG&most_recent_value_desc=true&start=2019&view=bar.

[60] C.R. Wells. “The exacerbation of Ebola outbreaks by conflict in the Democratic Republic of the Congo”. en. In: *PNAS* 116 (2019), pp. 24366–24372.

[61] M. Gayer et al. “Conflict and emerging infectious diseases”. en. In: *Emerg. Infect. Dis.* 13 (2007), p. 1625.

[62] G.E.C. Charnley et al. “Investigating the impact of social and environmental extremes on cholera time varying reproductive number in Nigeria”. en. In: *PLoS Glob. Public Health* 2 (2022), e0000869.

[63] G.E.C. Charnley et al. “Using self-controlled case series to understand the relationship between conflict and cholera in Nigeria and the Democratic Republic of Congo”. en. In: *Emerg. Infect. Dis.* 28 (2022), pp. 2472–2481.

[64] S Brechenmacher. *Stabilizing northeast Nigeria after Boko haram*. Vol. 3. Carnegie Endowment for International Peace Washington, DC, 2019.

[65] M. Bello. “The Terror Campaign of Boko Haram: Its Transformation and Challenges to Nigeria’s Security”. In: *Golden Ratio of Social Science and Education* 1.2 (2021), pp. 85–94.

[66] C. Kinsey and A. Krieg. “Assembling a Force to Defeat Boko Haram: How Nigeria Integrated the Market into its Counterinsurgency Strategy”. In: *Def. Secur. Anal.* 37.2 (2021), pp. 232–249.

[67] J. Edwards. *Médecins Sans Frontières. Case Study North-east Nigeria*. en. 2020. URL: https://arhp.msf.es/sites/default/files/MSF-Emergency%5C%20gap-North-east%5C%20Nigeria-case%5C%20study-april%5C%202017.pdf.

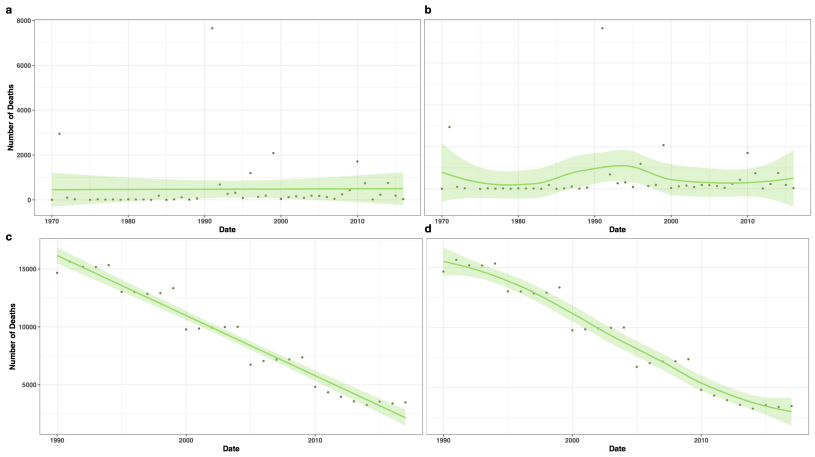
[68] T.O. Tengs et al. “Five-hundred life-saving interventions and their cost-effectiveness”. In: *Risk Anal.* 15.3 (1995), pp. 369–390.

[69] P.J. McEwan. “Cost-effectiveness analysis of education and health interventions in developing countries”. In: *J. Dev. Eff.* 4.2 (2012), pp. 189–213.

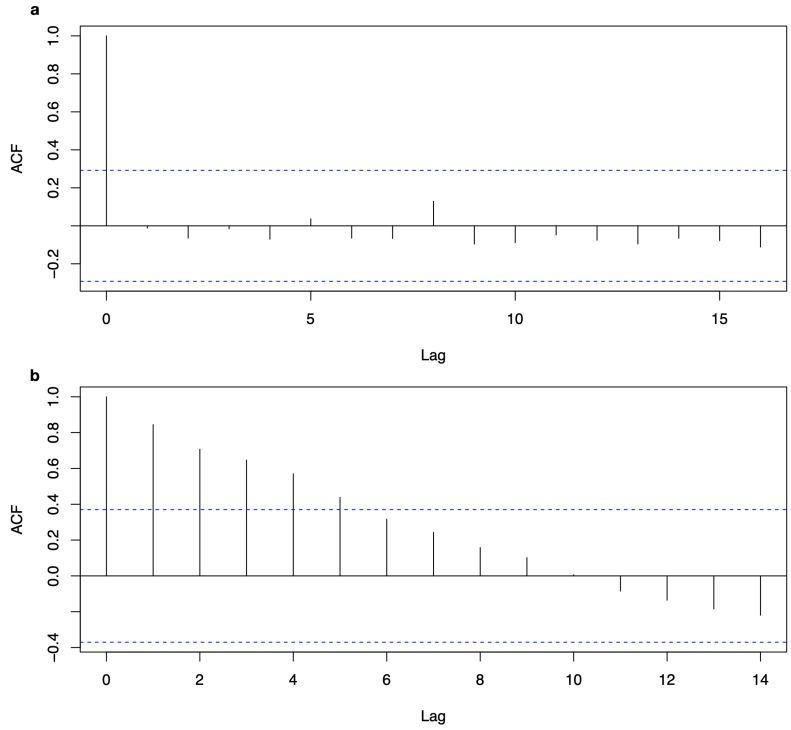
[70] T. Yates et al. “Efficacy and effectiveness of water, sanitation, and hygiene interventions in emergencies in low-and middle-income countries: a systematic review”. In: *Waterlines* (2018), pp. 31–65.

Supplementary Material

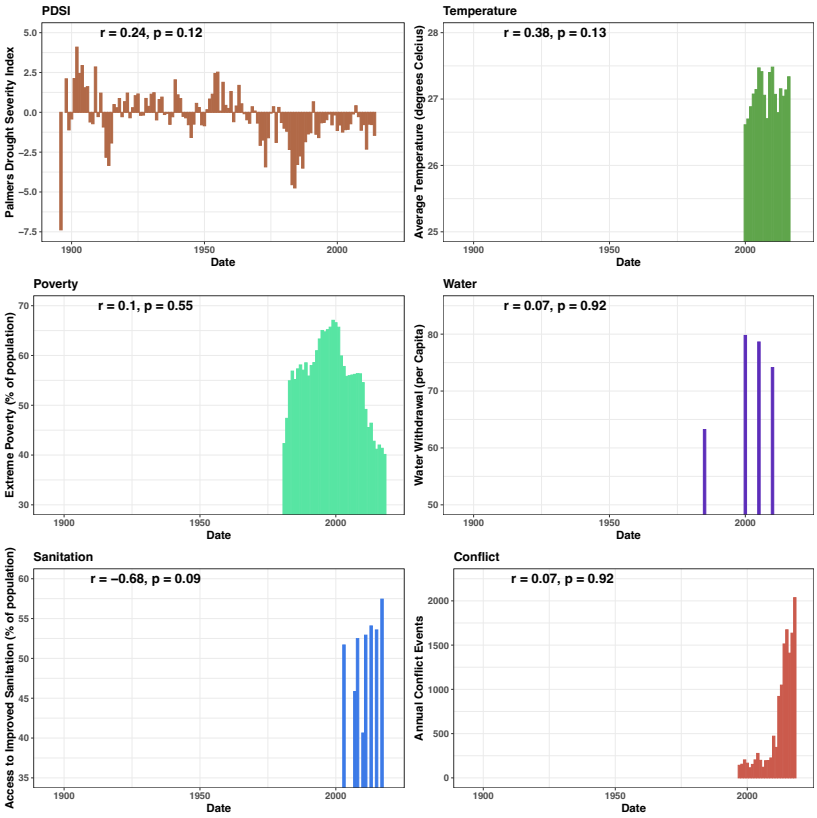
Supplementary Figures



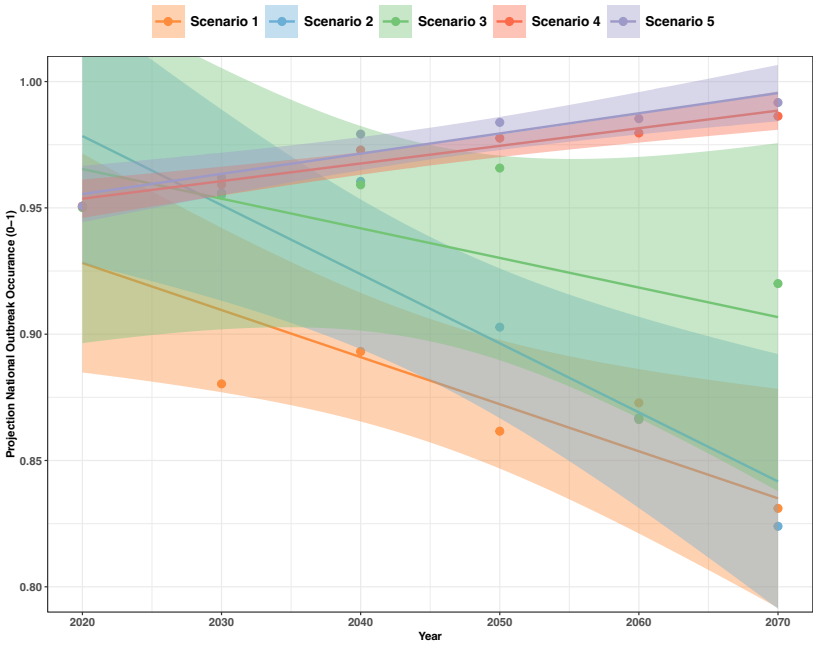
Supplementary Figure 6.1: Time series of cholera deaths for **a**, WHO data for 1970 to 2016, with a linear trend line [16], **b**, with a loess curve and **c**, the GHDx data for 1990-2016 [17], with a linear trend line and **d**, with a loess curve.



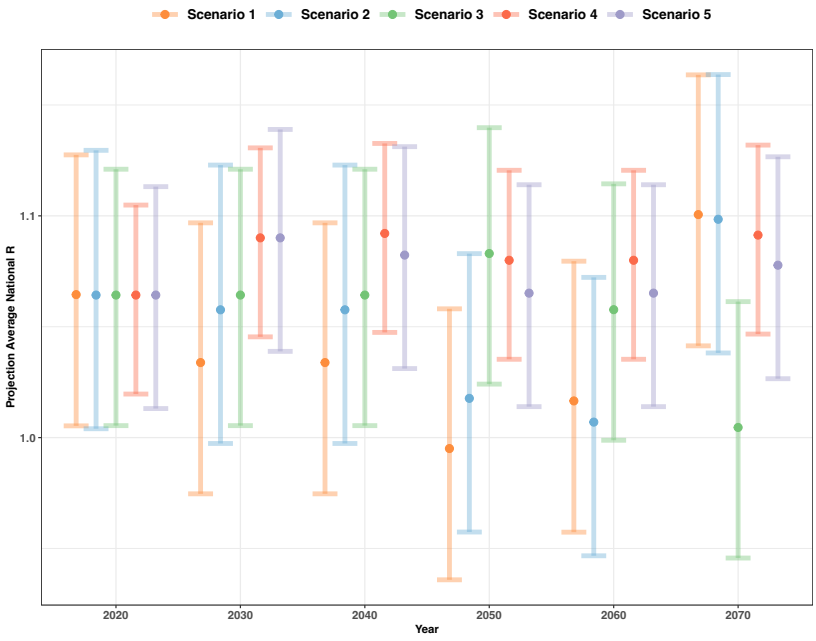
Supplementary Figure 6.2: ACF plots for **a**, the WHO data and **b**, the GHDx data. The dashed blue line represents the confidence interval (blue-dashed line at 95%), with ACF measured as the correlation coefficient of the residuals (between the time series and it lagged values). The lag is set to $10log_{10}(N/m)$, where N is the number of observations and m the number of series.



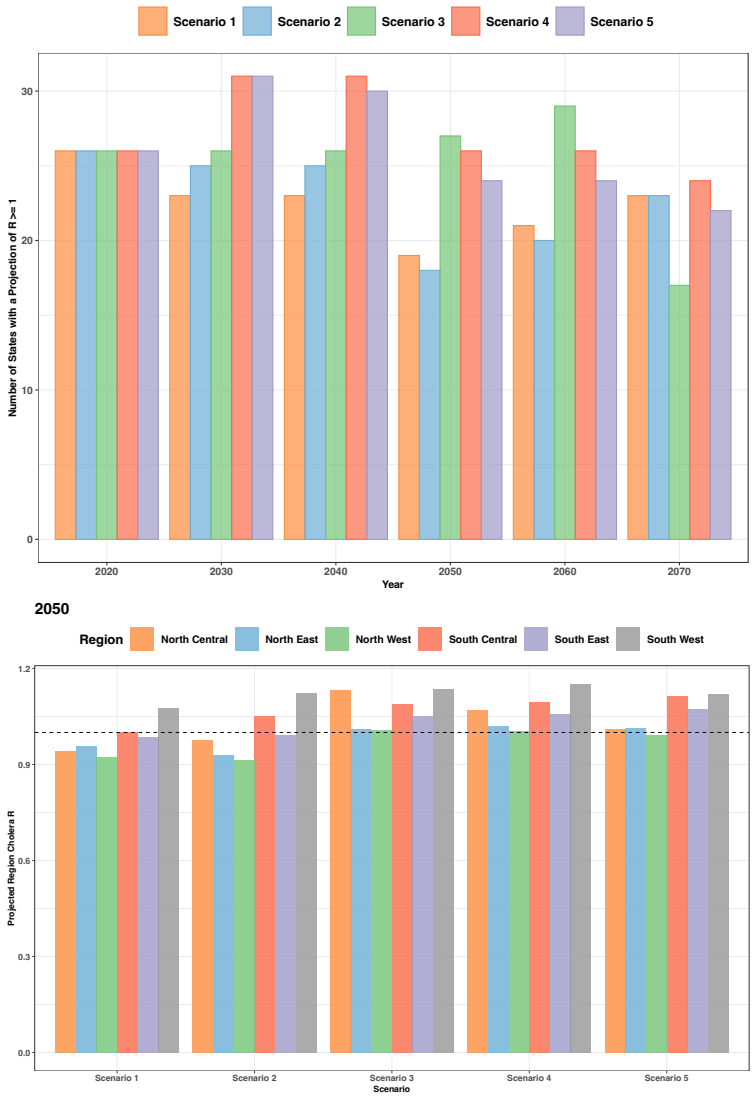
Supplementary Figure 6.3: Data for the covariates most commonly found as significant in the models fitted. Correlation coefficient (r) represents the correlation between the covariate and the WHO cholera deaths data [16] and the p-values. Most are not found to be significant, potentially due to a lack of complete data.



Supplementary Figure 6.4: National cholera projections for Nigeria, in cholera outbreak occurrence (0-1) for the five future scenarios, with a linear trend line and standard error.



Supplementary Figure 6.5: Averaged sub-national cholera projections for Nigeria, in cholera R_t, for the five scenarios to 2070 with 95% confidence intervals. For comparison with Figure 6.4.



Supplementary Figure 6.6: Sub-national projected changes in cholera transmission (R_t) for Nigeria. **Top panel**, number of states with projected R_t values over 1 for each year and scenario and **bottom panel**, average regional R_t value for each scenario at 2050.

The regions are based on the six Nigerian geopolitical zones. *North Central*: Benue, Kogi, Kwara, Nasarawa, Niger, Plateau, Federal Capital Territory. *North East*: Adamawa, Bauchi, Borno, Gombe, Taraba, Yobe. *North West*: Jigawa, Kaduna, Kano, Katsina, Kebbi, Sokoto, Zamfara. *South East*: Abia, Anambra, Ebonyi, Enugu, Imo. *South Central*: Akwa Ibom, Bayelsa, Cross River, Delta, Edo, Rivers. *South West*: Ekiti, Lagos, Ogun, Ondo, Osun, Oyo.

Chapter 7

Conclusion

7.1 Summary of the Motivations and Objectives

The primary motivation for the thesis was to understand why disaster-related disease outbreaks occur, even though disasters are not a new phenomenon. The thesis has helped answer this question by highlighting the complexities and multi-factorial nature of disaster-related disease outbreaks, further quantifying multiple risk factors and cascades. These outbreaks and their risk factors involve human behaviours and are somewhat dependent on the capacity of the population to adapt. Disaster-related disease outbreaks occur due to a breakdown in societal response and preparation, with several risk factors commonly reported as important in both the review and modelling studies. Quantifying these outbreak risks has been fundamental to understanding the health impacts of climate and global change here. Common risk factors included WASH, poorly managed displacement and poverty. Furthermore, certain regions were commonly impacted in the same disasters and resultant outbreak. For example, the review found that in conflict settings, poor access to healthcare and immunisations increase viral disease outbreaks.

Through the objectives of this thesis, key areas of disasters, infectious disease and climate change research have been brought together into one project. The thesis has provided more detail on the associated risk factors and links than previous studies, and investigated many areas deemed under-researched e.g., the links between drought and cholera in Africa, cholera projections and investigating the impacts of multi-hazard risks. The motivations (Introduction 1.5.1-3) of the thesis will be re-iterated and integrated throughout the Conclusion, to highlight how they have helped form the research questions and subsequent analysis presented here. How each objective (Introduction 1.5.4) has been achieved is stated below, followed by a summary of the thesis chapters (7.2):

1. *Create a comprehensive review* - The review identified disaster-related disease outbreaks as a

global issue, quantifying important characteristics and key areas of further research.

2. *Use novel methodological approaches and datasets to gain a greater understanding of cholera outbreak risk factors* - Four cholera datasets and three modelling approaches were used and analysed here to help understanding key cholera risk factors, particularly sanitation and poverty.

3. *Evaluate a range of social and development indicators* - Thirty five covariates, ranging across environmental and social indicators were investigated. Twenty three of these were related to social conditions including poverty, WASH, population, conflict, economics, health, education and nutrition.

4. *Apply the models to make quantitative predictions* - Scenario projections for cholera were made, accounting for climate and social changes, at both a national and sub-national scale to 2070, along with scenarios for current ideal conditions for outbreak prevention.

5. *Evaluate the achievability of global cholera targets* - The achievability of the GTFCC 2030 targets was assessed and further policy recommendations made to help achieve these targets by 2050 in Nigeria, while striving to meet the goals as soon as possible.

7.2 Summary of the Research Chapters

After introducing the areas of study in Chapter 1, Chapter 2 starts by investigating the scale of disaster-related infectious disease outbreaks through a systematic review. A greater understanding of the characteristics of outbreaks in a disaster setting, including common regions, disasters and aetiologies was gained. In the 132 studies reviewed, there were 137 different disasters and 140 disaster-related disease outbreaks. The outbreaks were grouped into several categories by disaster, region and disease. Many of the categorises were over-represented such as African conflicts leading to viral and vector-borne disease outbreaks and South and South East Asian hydrological-related outbreaks caused by water-borne bacterial pathogens. The analysis showed that human displacement, poor levels of WASH and insufficient housing were common risk factors relating to natural hazards. Whereas in conflict settings, poorly managed displacement and access to healthcare were more commonly reported. It was suggested in Chapter 2 though that disaster-related disease outbreaks are more complex than the relatively simple way they were categorised in the results and noted the commonality of multiple risk

reporting and the possibility of cascades in most of the reviewed outbreaks. Therefore, the remaining chapters focused on understanding and quantitatively analysing risk factors which lead to disease outbreaks in post-disaster settings.

For Chapter 3, drought-related cholera outbreaks in Africa were investigated using a covariate selection process and the data fit to generalised linear models, to further understand the environmental and social risk factors involved. Droughts are complex natural hazards and difficult to define, with comparatively little research compared to other hazards e.g., flooding (33% of the 137 reviewed disasters) and earthquakes (20%). The research gap was identified due to several reports of outbreaks during droughts, while still being relatively underrepresented, limiting the chances of a reporting bias. Cholera was identified as the most common aetiology in a drought setting in Chapter 2 and a pathogen which frequently had multiple risk factors. The disease has a high global burden and has proved difficult to eradicate. Meteorological drought was found to be a significant risk factor for African cholera outbreaks, along with a positive effect of population, temperature and poverty and a negative effect of freshwater withdrawal. The results helped to confirm the hypothesis that during droughts water is limited or mismanaged, while the population continues to shed cholera into the environment, increasing pathogen concentration and risky drinking water behaviours due to a lack of alternatives. National scenario projections across Africa, accounting for climate (temperature and drought) and global (water withdrawal and poverty) change, helped to shed light on the potential for increased sustainable development to offset future national cholera risk in Africa.

Chapter 4 used the Self Controlled Case Series methodology and conditional logistic regression models to investigate the association between conflict and cholera in two high burden countries, Nigeria and the DRC. Chapter 4 aimed to investigate why in high cholera burden countries (such as Nigeria and the DRC), the drought projections in Chapter 3 saw only negligible change in cholera risk, even with emission reductions and increased development. Other risk factors may have been affecting cholera in these countries, that were not considered in Chapter 3. An example of a cholera risk factor not considered, yet present in Nigeria and the DRC were conflicts. The models found that conflict was significantly associated with cholera outbreaks in both Nigeria and the DRC. Conflict increased the risk of cholera outbreak onset by 3.6 times in Nigeria and 2.6 times in the DRC, with some subnational areas showing a greater risk. In Nigeria, 19.7% of cholera outbreaks could be attributed to a conflict and in the DRC 12.3% of cholera outbreaks were attributed to conflict. Varying the exposure

periods and outbreak definitions proved model robustness and that cholera was particularly likely to arise within the first week following the conflict. The SCCS models not only proved to be a robust modelling approach but also showed the importance of rapid assistance in conflict settings to prevent outbreaks.

Chapter 5 used several measures of environmental and social extremes and pre-existing vulnerabilities, to investigate the impacts of multiple hazards using random forest models and cholera R_t. The chapter aimed to address several previous methodological limitations, including the use of binary outcome variables, reducing the understanding of cholera outbreak severity and data accuracy issues. Using confirmed cholera case data for 2018 and 2019, obtained from NCDC, incidence was calculated and used to model R_t, which was then used as the outcome variable for the model. A machine learning approached was taken, using random forest variable importance and several performance metrics for covariate selection and the best fit model identified, in terms of predictive power. The best fit model included PDSI, monthly conflict events, MPI and access to sanitation. A traffic-light system of cholera outbreak risk identified specific thresholds and triggers for cholera outbreaks in Nigeria. The system used two hypothetical scenarios needed to predict R_t below 1 (Green) and over 1 (Red). Additional spatial analysis helped understand the heterogeneities in the social and environmental extremes identified as important (conflict and PDSI). The scenarios found that sanitation access below 54% and MPI values above 0.38 put states at high risk of cholera in Nigeria. The traffic-light system and spatial analysis also highlighted the potential for sustainable development to reduce the impacts of disasters, as increasing sanitation and decreasing MPI reduced the effect of PDSI and conflict on R_t.

A motivation of the thesis was to understand where current policy falls short of preventing cholera outbreaks in a disaster setting and Chapter 6 aimed to bring together the evidence and understanding gained from Chapters 2-5 to evaluate the achievability of the GTFCC 2030 cholera goals in Nigeria. Evaluating whether global health targets can be met is vital to prevent them being forgotten or motivation being lost and the understanding gained from modelling work can inform this. Chapter 6 used historical data and trends and took the best fit models from Chapter 3 and 5 to project cholera to 2070 with 5 alternative scenarios. The scenarios built on the work of Chapter 3 but covered a wider range of future conditions, which included both improvements and regression. The historical data, projections and the evidence presented throughout the thesis, suggested that 2030

may not be a realistic target for a 90% reduction in cholera deaths in Nigeria. Research and policy on cholera elimination largely relies on improving outbreak response but long-term development is needed. Based on these results, a new 2050 target was suggested for Nigeria, looking at enhancing WASH services and alleviating poverty, especially in the northern states. Conflict was repeatably found to be detrimental to health in Nigeria and international commitment and coordination is needed in response to Boko Haram as a matter of global security. Continuing to degrade the groups territorial control and improving local-level stabilisation efforts will be vital in reducing attacks. Equitable development across the country and effective urban planning will be important in the future to achieve Nigeria's health targets and to empower the population to make informed health decisions. Pandemics should not be seen as one country or region's issue and an international approach (in terms of resources and coordination) is essential for prevention and control.

7.3 Applications

7.3.1 Applications to Climate Change

How societies will and can respond and adapt to climate change in the future is difficult to determine, but understanding future risks is vitally important. Disasters provide a unique opportunity to attribute a health outcome to a climate event and therefore climate change. This research has the potential to inform several areas of policy including climate change mitigation and adaptation policy, and disaster risk and reduction policy. Similar to Figure 3.1, which hypothesised the links between climate change, drought and cholera outbreaks, the links more generally found here, between climate change and cholera outbreaks and some of the pathways and cascades discussed in the thesis are summarised below in Figure 7.1.

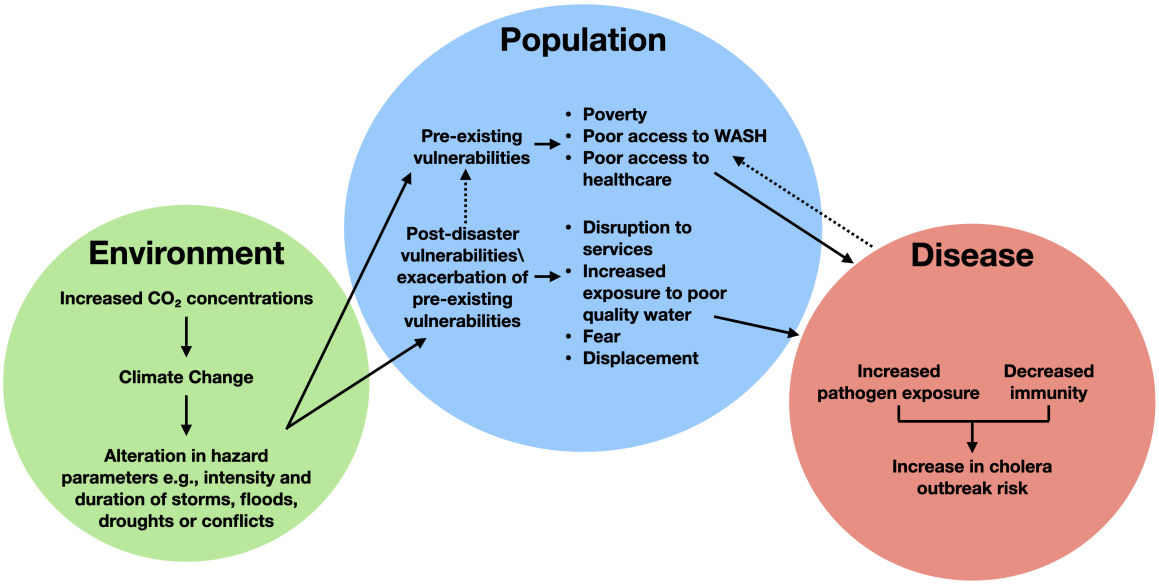


Figure 7.1: Pathways from climate change to cholera outbreaks, with a focus on natural hazards and sustainable development. The pathways link together how environmental changes in terms of natural hazards (Environment), interact with a vulnerable population (Population) and lead to cholera outbreaks (Disease) and how these can worsen pre-existing vulnerability (dashed arrows). The mechanisms are not considered a complete list of all potential pathways climate change may lead to increased cholera outbreaks, but instead bringing together some specific evidence from this thesis.

Chapter 2 identified disaster-related disease outbreaks as a global issue, but also stated regional similarities in terms of certain disasters and diseases, creating the opportunity for region/disaster-specific policy. Examples of these regional similarities, and therefore potential policy recommendations, included flooding leading to leptospirosis, particularly in Asia. Public health education and messaging is therefore important in a post-flood setting, to make sure that people avoid contact with floodwater and know the risks. Additionally, conflict in Africa and the Middle East disrupting access to healthcare and poor vaccination coverage was found to cause increases in viral diseases in children. Efforts should be taken globally to avoid the impact of conflict on childhood vaccination campaigns.

Risky drinking water behaviours are an issue during drought (Figure 3.1), and this thesis found evidence that they can lead to cholera outbreaks. The model results from Chapter 3 found that sufficient safe drinking water is an essential intervention in drought settings, while continually advocating for the sustainable expansion of freshwater availability. Risky water practices are not a choice but because alternatives are lacking and providing safe water, while also educating the population on the risks of

water sources during droughts could help reduce cholera risk. The work presented here largely focuses on low-income countries, where the effects of climate change will be felt the hardest, and more research is needed to prevent widening the health equity gap in the face of climate change.

The risks of climate change for health are far-reaching but gains in accuracy and precision of climate models and attribution studies has helped improve the global understanding of this issue. Climate models have continually demonstrated how climate change may alter hazard parameters, such as changes in frequency and intensity. The links and risk factors found here (such as the need for safely managed sanitation and provision of services) should be integrated into both climate change and disaster-risk reduction policy, in preparation for changing hazard parameters.

7.3.2 Applications to Cholera Policy and Control

The thesis highlighted how the results from modelling studies could directly feed into cholera policy. The quantification of cholera risk factors identified were used to develop national and sub-national recommendations. Understanding cholera risk factors was the secondary motivation of the thesis, along with using this to inform policy. The results from the modelling chapters shows poverty and sanitation as policy priorities for cholera control. As the seventh cholera pandemic continues, understanding cholera transmission factors are key to implementing outbreak mitigation strategies. Although Nigeria was mainly the focus for policy suggestions, many of the recommendations made would likely be relevant to other countries e.g., prioritising regions with sanitation access below 50%. At a global level, increasing access to testing, enhancing data collection of measurable risk factors (e.g., percentage access to sanitation and water, poverty headcount and number of healthcare facilities), environmental testing and effective urban planning, would undoubtedly help to control cholera in endemic countries.

7.3.3 Applications to Sustainable Development

The tertiary motivation of the thesis was to evaluate how sustainable development could benefit disease control. Several links and co-benefits between health and development have been described and should be used as further evidence for the need for greater long-term investment and coordination. A conclusion of several of the modelling chapters was the potential for the expansion of sustainable

development (particularly poverty alleviation) to offset the risks of both the disaster and therefore disease outbreak. As stated in the Introduction, marked development inequities between regions remain, an issue also found in Nigeria (Chapters 5 & 6), which significantly impacts the national and global disease trends. Furthermore, the results from Chapter 6 show how bringing all regions and countries to the same levels of development is important for disease control and finding key areas of prioritisation could help to focus fund allocation and motivation.

7.3.4 Methodological Applications

Several of the methods adapted and applied here (GLMs, SCCS and RFs) could be used to expand the body of work on disaster-related disease outbreaks. Additionally, the flexibility of conditional logistic regression models used in the SCCS method, could allow for more variables to be included in the analysis, such as social or environmental factors. Therefore, the flexibility of the modelling framework lends itself to several other interesting research questions. The projections created here could also be applied to other contexts and expanded. As stated, cholera forecasting and projection under global change is relatively understudied and the scenarios created should be considered as a starting point for more areas of global change and diseases. Furthermore, the work in Chapter 5 helped shed light on specific risk factors and ideal conditions to avoid cholera outbreaks (e.g., <50% sanitation), the concept of creating scenarios to inform the conditions needed to avoid outbreaks could be applied to a range of fields and diseases.

7.4 Future Work

7.4.1 Developing Data Metrics

As stated above, many of the methods used here could be applied to other countries, analysing different disasters and disease aetiologies, as this thesis could not comprehensively analyse all possible links and relationships. The benefits of doing so would be to help further inform policy and understanding and therefore lead to reductions of mortality and morbidity in disaster settings. To facilitate research and aid those working in government and humanitarian organisations, the best available data are

needed and both research institutions and governing bodies should continually stride to increase data quality. Several limitations of the data were pointed out here and a way to potentially tackle these limitations globally, to improve transparency and to make research more comparable, would be to identify a universal metric of reporting effort. Metrics have been used throughout this thesis including MPI and the Corruptions Perception Index. The Global Health Security Index [0], which judges how well prepared countries are for outbreaks, already has a detection and data parameter and could be built upon to create a more specialised metric, accounted for (perhaps by penalisation of a total figure) under-reporting, over-reporting, reporting lag and national barriers to report and test.

In terms of conflict, the development of additional metrics to better understand the impacts of the conflict would be helpful. Conflict severity is difficult to measure, due to its subjective nature and would perhaps lead to some conflicts being termed "less severe". However, in reality all conflicts are very disruptive to those who live in the conflict-affected or post-conflict areas. In terms of research though, it could be helpful in clarifying the extensive impacts of conflict and further understand the relationship between conflict and health. Possible variables to consider in a conflict severity metric could be the number of events, fatalities, financial costs, damage to schools and hospitals and loss of employment. To fully understand these complexities, those living in conflict-affected regions would need to be consulted and their insight would help gain a greater understanding of this issue.

7.4.2 Expanding Data Collection

As mentioned in Chapter 6, further data collection on risk factors is very important to fully understand the changes in cholera data over time. Without monitoring both the disease and its associated risks, the disease data are taken somewhat out of context. It makes it difficult to understand how and why changes in disease burden have occurred and is also a barrier to achieving health targets. Data are needed on both environmental concentrations of the pathogen, particularly in water sources commonly used for drinking and washing, and for socio-economic conditions such as WASH, poverty and healthcare. By collecting data on risk factors, it is easier to identify the cause of the disease case/death increase and act accordingly, implementing targeted interventions and identifying at-risk areas. It is also important for epidemiological research to help further untangle the mechanisms and characteristics of cholera outbreaks. For example, collecting and testing water samples would help to further understand the role of waterbodies in cholera transmission, which is still relatively contentious.

7.5 Final Thoughts

A result found across all spatial scales analysed here (global to sub-national) is that pre-existing vulnerabilities need to be addressed before disasters occur and by doing so, this will reduce the need for disaster assistance and humanitarian aid. Hazards will inevitably strike, as they have throughout history, but it is how we as a society deal with these hazards, that result in the disaster. Giving people agency to adapt to hazardous situations is important for their empowerment and to protect their health. Ultimately the way to reduce disaster-related disease outbreaks and improve global health is to invest heavily in sustainable development including poverty alleviation, expanding access to housing and WASH and increasing access to education and healthcare. Doing so will give people the opportunity to adapt and the knowledge they need to make informed health decisions.