

Drowning in Disease: Understanding Health Risks in a Changing Climate

An Exploratory Data Analysis into the Relationship Between
Flood Disasters, Healthcare Systems and Communicable Diseases



By C.P.M. van Dorst

Drowning in Disease: Understanding Health Risks in a Changing Climate

An Exploratory Data Analysis into the Relationship Between
Flood Disasters, Healthcare systems and Communicable Diseases

by

C.P.M. van Dorst

Master thesis submitted to Delft University of Technology in
partial fulfillment of the requirements for the degree of:

Master of Science
in
Complex System Engineering and Management

To be defended on 24th of June, 2024

Graduation Committee

First supervisor:	Dr. Tina Comes
Second supervisor:	Dr. Saba Hinrichs-Krapels
Advisor:	Julien Magana



Dear reader,

Welcome to the master's thesis "Drowning in Disease: Understanding Health Risks in a Changing Climate," the result of five months of dedication and hard work. In the search of a topic for my master thesis, I aimed to find a space where I could study the effects of climate change with the use of data analysis. By choosing a topic of societal relevance, I hope to inspire others to focus their research on global issues.

Water's power has always inspired me, whether watching rain pour outside my window or the rhythmic waves at the beach. However, this thesis has shown me the devastating effects water can have on people, especially with sudden and extreme rainfall. Understanding how climate change intensifies these events makes me nervous, but it also reminds me of the importance of respecting nature.

This project would not have been possible without the guidance of my supervisors, and therefore, I want to express my gratitude to Julien, Tina, and Saba. From the beginning, you provided me with the tools to tackle this challenging topic while giving me the freedom to be creative and explore research on my own.

I hope you enjoy reading this thesis and feel inspired to contribute to making the world a better place.

*C. van Dorst
Delft, June 2024*

Executive summary

Amidst the challenges posed by ongoing climate change, communities worldwide are increasingly facing the escalating frequency and severity of extreme weather events. Among these natural disasters, floods are particularly widespread, causing extensive destruction and posing significant threats to both the environment and human well-being. Health effects reported after floods include drowning, injuries, communicable and non-communicable diseases, and mental health disorders. While the majority of these risks are well documented across various settings, the risk of communicable diseases following flooding is highly context-specific. This research aims to address a knowledge gap by exploring the interplay between flood disasters, healthcare systems, and communicable diseases. It seeks to understand how various characteristics, including social, economic, institutional, and geographical interrelate with the health impacts of communicable diseases after flood events.

This research highlights the vulnerability of countries such as Peru, Sri Lanka, and Pakistan, where past floods have shown significant impacts on the incidence of communicable diseases. The resilience of healthcare systems is evaluated by analyzing both vulnerability and resilience factors, which together determine a country's susceptibility to health impacts from floods and its ability to manage these impacts. Vulnerability is influenced by socio-economic, public health, and behavioral factors, while resilience is shaped by the quality of healthcare system, institutional effectiveness, and individual capacity. The analysis illustrates that even with comparable healthcare system performance, significant differences in resilience exist between countries, highlighting the important role of individual factors. The findings suggest that strengthening local empowerment through strategies that support access to education and technology, individual safety and autonomy and civic participation can significantly strengthen community resilience. Furthermore, the positive correlation between a country's exposure to floods and vector-borne diseases indicates that countries already vulnerable to floods are at greater risk of communicable diseases. Therefore, it becomes even more critical to adopt community-driven solutions that use the strengths of individuals within these at-risk populations to effectively build strong resilience against the combined risks of floods and diseases.

The research used a quantitative approach with exploratory data analysis to understand the health impacts of flood disasters. Using a causal-comparative design, the study investigates cause-and-effect relationships by analyzing natural flood events. The methodology is divided into three phases. Initially, a conceptual framework is developed through a desk study and concept mapping to establish preliminary hypotheses about the health impacts of flood disasters. The second phase involves systematic data collection and analysis using secondary data sources. This phase utilizes exploratory data analysis, including both visualization and statistical techniques, to explore relationships between flood events, healthcare system, and communicable diseases. Additionally, principal component analysis is performed to identify indicators describing the relationship between health characteristics and increased incidences of communicable diseases after floods. In the final phase, insights from the data analysis are synthesized to develop an empirically validated framework.

Understanding the health impacts of floods and the factors that enhance community resilience is important for developing effective disaster response strategies. This knowledge supports the development of targeted mitigation measures aimed at addressing the challenges posed by communicable diseases following floods. By identifying key domains of healthcare vulnerabilities and resilience factors, and emphasizing individual empowerment, this research provides valuable insights that help in developing integrated, holistic approaches to disaster preparedness and response. This research represents an initial step in exploring this complex system and emphasizes the need for further exploration to guide the way toward resilient communities. While the future remains uncertain, promoting more resilient and healthier communities is important in preparation for potentially severe future impacts.

Contents

List of Tables	7
List of Figures	9
List of abbreviations	11
1 Introduction	13
1.1 Research Objective	14
2 Methodology	16
2.1 Research Approach	16
2.1.1 Causal-comparative research design	16
2.1.2 Advantages and limitations	16
2.1.3 Connection to CoSEM program	17
2.2 Research Sub-Questions	17
2.3 Methods	17
2.3.1 Conceptual framework	18
2.3.2 Database development	18
2.3.3 Data Analysis	19
2.3.4 Framework and Future Research	19
2.4 Research Structure	20
3 Conceptual Model	22
3.1 Theoretical Framework	22
3.1.1 Flood disasters	22
3.1.2 Health impact of floods	23
3.1.3 Scope of Health Effects	24
3.1.4 Principles of infectious disease after hydrologic disasters	25
3.1.5 Lag period communicable diseases	25
3.1.6 Mechanisms Cholera Transmission	26
3.1.7 Mechanisms Vector-borne Disease Transmission	27
3.2 Conceptual Framework	29
3.2.1 Selection of frameworks and studies	29
3.3 Review	30
3.3.1 Defining key concepts	30
3.3.2 Review of Objectives and drivers	31
3.3.3 Conceptual Framework Development: Defining Framework Domains	33
3.3.4 Conceptual Model Development: Interrelationships	35
3.3.5 Initial list of indicators	35
4 Database Development	38
4.1 Data sources	38
4.2 Data collection process	40
4.2.1 Selection of countries	40
4.2.2 Flood data	40

4.2.3	Epidemiological data	41
4.2.4	Health care system indicators	41
5	Analysis of Flood Impact on Communicable Diseases	44
5.1	Methodology	44
5.1.1	Exploratory Data Analysis	44
5.1.2	Visual Exploratory Data Analysis	45
5.1.3	Statistical Analysis	45
5.2	Results Visual Analysis	46
5.2.1	Results Visual Analysis - yearly data	46
5.2.2	Example plots weekly and monthly data	49
5.3	Results of Statistical Analysis	52
5.3.1	Correlation Analysis	52
5.3.2	Paired T-Test	52
5.3.3	Regression Analysis and scatter plots	54
5.4	Discussion	55
5.4.1	Visual Analysis	55
5.4.2	Statistical Analysis	56
5.5	Conclusion	58
6	Analysis of Healthcare System Resilience	61
6.1	Methodology Principal Component analysis	61
6.2	Results PCA	63
6.2.1	Selection and collection of indicators	63
6.2.2	Descriptive statistics	68
6.2.3	Pre-processing the data	69
6.2.4	Clustering countries based on flood-vectorborne disease exposure	69
6.2.5	Removing redundant data and verifying use of PCA	71
6.2.6	PCA and selection of number of components	73
6.2.7	Resilience Score	75
6.2.8	Varimax Rotation	75
6.3	Results PCA: categories	77
6.3.1	Low exposure	77
6.3.2	High exposure	79
7	Final Framework	81
7.1	Evaluation of PCA results	81
7.1.1	General considerations	81
7.1.2	Full sample size	81
7.1.3	Low/High exposure	82
7.1.4	Summary of key findings	83
7.2	Final framework	84
7.3	Interpretation of Final Framework	85

8 Discussion	88
8.1 Introduction	88
8.2 Methodological reflection	88
8.2.1 Exploratory data analysis	88
8.2.2 Data sources	89
8.2.3 Selection of appropriate techniques	89
8.3 Research limitations	90
8.3.1 Limitations in medical knowledge	90
8.3.2 Data	90
8.3.3 Selection of indicators	91
8.3.4 PCA	92
8.3.5 Validation	92
8.4 Theoretical implication: Contribution to knowledge gap	93
8.5 Practical implications: Remaining challenges and further research	94
8.6 Conclusion	96
Bibliography	97
Appendices	104
A Epidemiological data	104
B Overview of Reviewed Frameworks	106
A Social Determinants of Health	106
B Health impact pathway for flooding	106
C Flood vulnerability assessment framework	107
D Public Health Emergency Preparedness Framework	107
C Healthcare system indicators	108
D Formulas Statistical tests	111
A Paired t-test	111
B Pearson correlation test	111
C Linear Regression test	111
E Visualisation results	112
A Yearly data - Cholera	112
B Yearly data - Malaria	113
C Monthly and weekly data - Dengue	114
D Monthly and weekly data - Leptospirosis and ADD	116
F Scatter plots	118
G Descriptive statistics - histograms	120
H Correlation matrix of exposure indicators	123
I Shapiro Wilk	123

List of Tables

1	Overview of methods and data requirements for each research phase	18
2	Outline of data requirements for flood data, country characteristics and epidemiological data	19
3	Overview of reviewed frameworks addressing health and flood risk management . . .	32
4	Sources for the Conceptual Framework	33
5	Initial list of indicators based on the reviewed frameworks	36
6	Initial set of flood-prone countries included for analysis	40
7	Availability of epidemiological data divided into weekly, monthly and yearly available data including the time period of reported data	42
8	Summary of results visual analysis yearly data - Malaria	48
9	Summary of results visual analysis yearly data - Cholera	48
10	Summary of results visual analysis monthly and weekly data - Dengue	51
11	Summary of results visual analysis yearly data - waterborne	51
12	Results of Pearson Correlation Analysis of yearly reported malaria incidences versus yearly number of people affected by floods	52
13	Results of paired T-test for reported cases of Cholera, ADD, Leptospirosis and Dengue comparing flooded (+ 1 month) and non-flood periods in Bangladesh, Philippines, Sri Lanka, Australia and Peru.	53
14	Results of paired T-test for reported cases of Cholera, ADD, Leptospirosis and Dengue comparing flooded (+ 2 month) and non-flood periods in Bangladesh, Philippines, Sri Lanka, Australia and Peru.	54
15	Results of regression analysis, disease incidence vs. people affected by floods per year	55
16	Results of the visual and statistical analyses indicating the resilience of countries to communicable diseases following flood disasters.	59
17	Overview of indicators in the exposure category including unit and source	64
18	Overview of indicators in the vulnerability category including unit and source	66
19	Overview of indicators in the coping capacity category including unit and source . .	68
20	Dataset after pre-processing including tranformation, standardization and correcting theoretical orientation.	69
21	Countries categorized in two levels of exposure based on k-means clustering results .	70
22	Combined indicators describing the institutional and individual resilience	71
23	Indicators that have been removed in the PCA process including the reason for exclusion	71
24	Indicators included in PCA	72
25	Summary of results unrotated PCA	74
26	Resilience scores for each country based on the loadings of the original variables on each principal component, * indicating countries that were found to have a positive correlation between floods and vector borne disease outbreak(s).	75
27	Summary of rotated results PCA	76
28	Resilience scores for each country based on the loadings of the original variables on each rotated principal component, * indicating countries that were found to have a positive correlation between floods and vector borne disease outbreak(s)	76
29	Indicators included in PCA for Low Exposure countries	77
30	Results (Low expsoure, unrotated)	78
31	PCA results (low exposure, rotated)	78
32	Results PCA (high exposure, unrotated)	79

33	Results PCA (high exposure, rotated)	79
34	Summary of Findings	83
A.1	Significant outbreak of water-borne communicable diseases after flood disasters . . .	104
A.2	Significant reported outbreak of vector-borne communicable diseases after flood disasters + other reported communicable diseases	105
C.3	Healthcare System indicators	108
C.4	Table C.3: Continued	109
C.5	Table C.3: Continued	110
G.6	Summary of descriptive statistics for key indicators	122
I.7	Results of Shapiro-Wilk test for all indicators included in the PCA, indicating rejecting of H0	123

List of Figures

1	Research flow diagram describing the research activity, input/output and method used to construct each chapter of the thesis	20
2	Three photos illustrating examples of three different type of common floods	22
3	Effect of global warming on water cycle [Climate Council, 2022]	23
4	Overview of immediate, midterm and long term health impacts of floods	24
5	Epidemiological triad describing the relationship between the agent, host and environment	25
6	Transmission mechanisms of cholera transmission, representing waterborne diseases and vector-borne diseases	28
7	Conceptual framework of the relationship between floods, healthcare systems and communicable diseases	33
8	Data collection process for database development, including primary data sources and process goals.	38
9	Study protocol of the exploratory data analysis	44
10	People affected by floods vs. Malaria cases in Pakistan (suspected significance) . . .	47
11	People affected by floods vs. Malaria cases per year in Tanzania (no suspected significance)	47
12	Dengue incidence and flood events in 2010, 2013, 2017 and 2018 in Sri Lanka (suspected significance)	50
13	Dengue incidence and flood events in 2010, 2011, 2013, 2022, 2023 in Australia (without suspected significance)	50
14	Indicators in the Exposure category - Exposure to floods, exposure to waterborne disease and exposure to vector borne	64
15	Indicators in the Vulnerability category - Socio-economic, Ecological, Behavioural .	65
16	Indicators in the Coping Capacity category - Health Care System Resilience	67
17	Indicators in the Coping Capacity category - Institutional and individual resilience . .	67
18	Results of k-means clustering, based on the level of exposure	70
19	Correlation matrix of PCA for vector-borne diseases	72
20	Cumulative explained variance for the five retained principal components after performing principal component analysis	73
21	Visualisation of Exploratory Data Analysis results	74
22	Cumulative explained variance for the five retained PCs of the PCA for low exposure countries	77
23	Final framework	84
24	Future research agenda	95
B.1	Commission on Social Determinants of Health (CSDH) framework [Kumar, 2010] .	106
B.2	Health Impact Pathway for flooding [Few, 2007]	106
B.3	Flood Vulnerability Assessment Framework [Salami et al., 2017]	107
B.4	Public Health Emergency Preparedness	107
E.5	Caption for all subfigures	112
E.6	Reported number of malaria cases /1000 population at risk versus number of affected people by floods, 2000 - 2022 (various countries)	113
E.7	Monthly and weekly data - Dengue (Peru and Australia)	114
E.7	Monthly and weekly data - Dengue (Sri Lanka and Philippines	115
E.8	Monthly and weekly data - Leptospirosis (Australia and Sri Lanka)	116

E.8	Monthly and weekly data - Leptospirosis and ADD (Philippines, Bangladesh)	117
F.10	Scatter plots of Waterborne disease - Cholera, for Nigeria, Tanzania and Mozambique	118
F.9	Example of scatter plots with suspected linear relationship (a) and without suspected linear relationship (b)	118
F.11	Scatter plots of Vector-borne disease - Malaria, for Nigeria, Tanzania and Mozam- bique, Pakistan, China, South Sudan and Dominican Republic	119
G.12	Histograms of indicator data	120
G.13	Figure H.14: Continued	121
H.14	Correlation matrix of indicators describing level of exposure to floods and disease . .	123

List of abbreviations

Abbreviation	Definition
CoSEM	Complex System Engineering and Management
CSDH	Commission on Social Determinants of Health
EDA	Exploratory Data Analysis
EM-DAT	Emergency Event Database
GHDx	Global Health Data Exchange
GII	Gender-Inequality Index
GDP	Gross Domestic Product
HIV	Human Immunodeficiency Virus
IHME	Institute for Health Metrics and Evaluation
KMO	Kaiser-Meyer Olkin
LMIC	Low- and Middle-Income Country
MDPI	Multidimensional Poverty Index
NGOs	Non-Governmental Organizations
OPHI	Oxford Poverty & Human Development Initiative
PCA	Principal Component Analysis
PC	Principal Component
SDH	Social Determinants of Health
UN	United Nations
UNDP	United Nations Development Programme
V-EDA	Visual Exploratory Data Analysis
WHO	World Health Organization

Chapter 1

Introduction

1 Introduction

Amidst the challenges posed by ongoing climate change, communities worldwide are contending with the escalating frequency and severity of extreme weather events, which carry significant consequences for both the environment and human well-being. Among these natural disasters, floods stand out as one of the most widespread, causing extensive destruction and posing a significant threat to global populations [Doocy et al., 2013, Du et al., 2010]. The World Health Organization (WHO) explains floods as the condition that occurs when an overflow of water submerges land caused by heavy rainfall, rapid snow melt or a storm surge from a tropical cyclone or tsunami in coastal area [World Health Organization, 2024a]. Over the past decade, floods alone have accounted for an alarming 53,000 fatalities globally, often caused by the speed of onset of the flood [Alderman et al., 2012, World Health Organization, 2024a, Du et al., 2010].

Understanding the impact and scale of flood events worldwide is challenging due to a diverse array of factors including geographical variability, climate change influences, urbanization effects, infrastructure considerations, socioeconomic factors, interconnected systems, human behavior, and the necessity for international cooperation in flood management. Moreover, recent examples highlight the severity of flood impacts, such as the 2023 flood in Libya which resulted in 4,352 deaths and the displacement of 44,800 individuals. In total, 1.5 million people were impacted by the disaster, accounting for 22% of the total population. Furthermore, the disaster resulted in damages estimated at approximately US\$ 1.7 billion, equivalent to 3.6% of Libya's GDP, necessitating urgent recovery efforts. [World Bank, 2023, World Health Organization, 2023] Similarly, in 2022, one-third of Pakistan was submerged, affecting 33 million people and resulting in 1700 fatalities and 12,867 injured people [United Nations Children's Fund, 2022]. The flood was caused by record high monsoon rains, with over 190% of its normal rainfall. [Red Cross, 2023] However, floods are not confined to low- and middle-income countries (LMICs); they pose a significant threat to high income countries like the Netherlands as well, as highlighted in the 2021 floods in Limburg [Endendijk et al., 2023]. Heavy rainfall over several days in this southeastern province of the Netherlands and neighbouring countries caused rivers to overflow their banks and submerged towns and villages, resulting in over 130 fatalities in both Germany and Belgium. [Asselman et al., 2022] This flood resulted in extensive damage to infrastructures, homes, businesses, and agricultural land, leaving the flood as the second most expensive natural hazard of 2021 with an estimated total damage worth of 38 billion euro [Nederlandse Omroep Stichting, 2021]. Roads were washed out, bridges were damaged, and residents had to be evacuated from flooded areas. Furthermore, the flooding disrupted transportation networks and essential services, leading to significant economic losses and disruption to daily life.

Floods pose significant immediate dangers to human health, illustrated by the high number of injuries and deaths, but also long-term effects resulting from water contamination, displacement, damaged health infrastructure, and worsened living conditions have been marked. Health effects reported after floods include drowning, injuries, noncommunicable diseases, communicable diseases, and mental health disorders [Paterson et al., 2018]. While the majority of these risks are well documented and found across all settings, the risk of communicable disease following flooding is highly context-specific. Although several past studies show increases in infectious diseases weeks to months after flooding, caused by factors such as population displacement, changes in population density, and concerns about waste management and access to clean water, there remains uncertainty about the strength of this relationship and the factors potentially explaining it [Brown and Murray, 2013, Alderman et al., 2012, Walika et al., 2023]. A notable example of this is the massive 2022 floods in

Pakistan, previously mentioned, which significantly worsened the country's malaria crisis, leading to a fourfold increase in cases and emphasizing the severe health impacts [World Health Organization, 2024d]. In contrast to the well-documented tangible impacts of floods on homes, buildings, and infrastructure, there remains a considerable gap in our understanding of the interplay between flood disasters, health-care systems, and communicable diseases [Lee et al., 2020, Alcayna et al., 2022]. In order to manage the impacts of communicable diseases resulting from floods, extensive knowledge of health risks and the capacity of the health system to mitigate or manage the health consequences is required. What makes this especially difficult is that all floods are unique, dependent on different social, economic, demographic, and healthcare characteristics [Du et al., 2010]. Therefore, further research is necessary to understand these interrelations and to provide insights for adequate risk assessments.

Past experiences with flood disasters have revealed the vulnerability of countries to the devastating impacts of flooding, highlighting the need for effective flood management and adaptation strategies to mitigate risks and minimize the socio-economic consequences of future floods. Anticipation of increased flood occurrences due to rising sea levels and intensified precipitation events, both attributed to climate change, adds another layer of complexity to the challenge. As sea levels rise and precipitation patterns become more extreme, the frequency and severity of flooding events are expected to escalate, increasing this need for research and understanding to enhance resilience against flood risks. [Parry et al., 2007, Ramin and McMichael, 2009]. The progression of urbanization further complicates these challenges by exposing an increasing number of individuals and infrastructure to the risks associated with flooding events. With urban areas expanding and population density increasing, more people are living in flood-prone areas, heightening the potential for loss of life, property damage, and health impacts during flood events [Du et al., 2010]. Addressing these complex and interrelated challenges necessitates a comprehensive approach, combining effective flood management, climate change adaptation, and urban planning strategies to reduce vulnerability and enhance the resilience of communities to future flood risks.

1.1 Research Objective

The research aims to address a knowledge gap by studying the interplay between flood disasters, healthcare systems and communicable diseases. Past research primarily provides post-disaster evaluations describing observed health impacts and offering country-specific recommendations for mitigating health effects after disasters. However, these recommendations often remain tied to unique health effects observed in specific geographical locations and contexts. The challenge lies in gaining understanding of which characteristics, including the social, economic, institutional, and geographical factors, may interrelate with the health impacts of communicable diseases after flood events. The main research question is:

"Which healthcare system vulnerabilities and resilience factors affect health outcomes for communicable diseases after flood disasters?"

The research objective is to broaden the understanding of the relationship between flood disasters and the impacts of communicable diseases across diverse settings, investigating characteristics in 14 flood-prone countries worldwide. By examining the identification of communicable diseases resulting from floods and comparing the health care systems, the objective is to find correlations between countries' characteristics and factors increasing the risk of potential disease outbreaks. This understanding could contribute to future flood preparedness plans by informing risk assessment and providing insights into the challenges faced by different countries.

Chapter 2

Methodology

2 Methodology

2.1 Research Approach

2.1.1 Causal-comparative research design

The chosen research approach, grounded in quantitative methods and structured data analysis, is selected to address the central question of understanding the specific health impacts of flood disasters in specific environments. This research design involves multiple methods including data collection, analysis, interpretation of results, and framework development [Creswell and Creswell, 2023]. The study adopts a causal-comparative research design, investigating potential cause-and-effect relationships between existing variables and examining phenomena after the fact (Ex Post Facto), given the naturally occurring flood disasters analyzed. [Campbell and Stanley, 1966, Schenker and Rumrill, 2004].

The causal-comparative research design typically follows a sequence of five steps. Initially, the unexplained phenomena is observed, and existing theories related to the issue are explored. Subsequently, a hypothesis or conceptual model is formulated to clarify the observed phenomena. This leads to the development of predictions regarding the outcomes, along with a plan for hypothesis testing. The data is systematically collected and processed. The findings are then subjected to analysis to evaluate the initial predictions. The verified results are presented in a suitable format, concluding the research process. [Campbell and Stanley, 1966]

2.1.2 Advantages and limitations

The selected causal-comparative research design plays a significant role in guiding this thesis for several reasons. This approach allows for a comprehensive analysis of the impacts of communicable diseases following flood disasters in diverse environments, offering insights into how various factors contribute to the occurrence and spread of diseases in affected countries. By analyzing data collected in the aftermath of naturally occurring flood disasters, this design enables a thorough exploration of complex relationships and interactions among different components within the healthcare system and the broader socio-economic context. Moreover, the causal-comparative design adopts a holistic approach, enabling a nuanced understanding of the interconnectedness among multiple variables and their influence on health outcomes. This comprehensive perspective is essential for identifying patterns and trends that may not be apparent when examining individual variables in isolation. However, it is important to acknowledge the limitations inherent in this research design. For instance, the inability to manipulate research groups necessitates careful interpretation of findings, considering the potential influence of both known and unknown variables. Furthermore, while this research design allows for in-depth analysis of specific study environments, the findings may not always be generalizable to other populations or settings due to the unique characteristics of the study environment and the availability of data. Despite these limitations, the causal-comparative quantitative research design is considered suitable for guiding this data-focused research, as it provides a framework for investigating the complex relationships between flood disasters, healthcare systems, and communicable diseases.

2.1.3 Connection to CoSEM program

This Master Thesis is the final part of the programme of Complex System Engineering and Management (CoSEM), a master programme of the faculty of Technology, Policy and Management at the Delft University of Technology. This research requires an interdisciplinary approach, aligning with the overarching criteria of the CoSEM thesis. It aims to design a framework to contribute to an intervention in the decision-making process of disaster preparedness. This thesis is guided by a set of methods, focusing on data analysis approach. The thesis objective aligns with the fundamental aim of the CoSEM program, contributing to defining complexities in a globally connected world while aiming to develop interventions that consider both technical and socio-cultural dimensions.

2.2 Research Sub-Questions

Following the main research question "Which healthcare system vulnerabilities and resilience factors affect health outcomes for communicable diseases after flood disasters?" and the adoption of the causal-comparative research design, four research sub-questions are proposed. Initially, health risks in flood-prone areas and existing theories are explored to develop a conceptual model. Moving forward, the hypothesized interrelations of the conceptual model are evaluated by the results of data collection and analysis, reflecting on the approach's structured methodology, providing guidance in systematically collecting and strategically using data for analyzing the hypothesis. The results from this analysis are then evaluated to provide insights into the relationships of the data. Finally, these insights are integrated into the final framework, and gaps in knowledge are identified to provide recommendations for policy and decision-making processes.

1. What conceptual framework can be designed to understand the interplay between health system vulnerability, flood exposure, and communicable disease health outcomes?
2. Which countries exhibit the highest vulnerability to communicable disease outbreaks following flood disasters?
3. To what extent do resilience indicators explain the varying levels of vulnerability to communicable disease outbreaks after floods?
4. What empirically validated framework could be used to analyse the vulnerability and resilience of healthcare systems to communicable disease outbreaks following floods?

2.3 Methods

To answer the proposed research questions, three research phases can be distinguished. Initially, a conceptual model will be developed to establish a preliminary hypothesis. Then, through systematic data collection and analysis, the relationships between elements of this conceptual model will undergo testing, with the results serving as the foundation for validating, rejecting or expanding elements of the conceptual model. Finally, the insights from the data analysis will be evaluated and used to formulate the final framework and provide recommendations for future research directions. This chapter briefly discusses the various research methods used in the different phases of this research, outlined in Table 1, along with their corresponding data requirements. Each chapter of the thesis will be introduced by a section providing the detailed methodology of each method used to answer the corresponding research question.

Table 1: Overview of methods and data requirements for each research phase

Research phase	Method/approach	Data requirements
1. Designing conceptual framework	Desk Study Concept Mapping	Epidemiological data, Health system performance Vulnerability assessments Policy and preparedness evaluations
2. Data collection and analysis	Secondary Data Collection, Data Mining, Exploratory Data Analysis: - Visual analysis - Pearson correlation test - Paired T-test - Regression Analysis - Principal Component Analysis	Flood data Epidemiological data Country characteristics Health system data
3. Framework and future research	Synthesis - System Thinking approach	Results from data analysis

2.3.1 Conceptual framework

The first sub-question falls within the exploration phase, where an assessment of existing literature, evaluation reports, and theories will be conducted to examine health impacts associated with flood disasters. The data collection method used for this exploration is a desk study, focusing on key concepts such as variations in observed health effects across different countries, key drivers of health impacts, governance, and past studies that indicate causation between health system variables and health impacts of flood disasters. Scientific literature, government reports, publications from international health organisations, and news articles may be used to perform the desk study. The data requirements include epidemiological data linked to flood disasters, socioeconomic data to comprehend community vulnerabilities, and data on health system performance, policy and preparedness evaluations. Following the desk study, the conceptual framework will be designed using the concept mapping method, which results in a diagram visually representing relationship between concepts.

A conceptual framework is a comprehensive system of interlinked concepts, assumptions, beliefs, and theories that guides research efforts. It serves as a foundational theory to structure the study, providing a fundamental understanding of the phenomenon under investigation and shows how the research problem will be explored. Conceptual frameworks are particularly important in exploratory research design, where data is used to develop holistic insights into the research topic. They help illuminate the relationships among various elements and contribute to the formation of theories about the system. In contrast, a theoretical framework is more oriented toward empirical theories, directing deductive, theory-testing studies by providing explanations rooted in existing theories or concepts. Theoretical frameworks inform research questions, methods, and data analysis to establish causal or correlational patterns. To summarize, conceptual frameworks are exploratory and holistic, while theoretical frameworks are focused on testing and validating existing theories or concepts. [Tamene, 2016, Upadhyay, 2015]

2.3.2 Database development

Currently, there is no database encompassing all data records relevant to the research. Therefore, in order to address the main research question, several sets of data need to be collected to construct this database. The database should consist of various types of data describing past flood events, epidemi-

ological data, and characteristics of geographical, economic, social, and institutional performance of a country. Using secondary data collection, it involves utilizing various sources including online databases, government and institutional records, and publicly available data. The database is constructed in Excel. An outline of the required data, presented in Table 2, is established to guide the data collection during the project, and includes the following:

Table 2: Outline of data requirements for flood data, country characteristics and epidemiological data

Flood data	Characteristics of countries	Epidemiological data
Total number of floods	Health care system	Morbidity rates vector-borne diseases
Number of affected people	Socio-economic factors	Mortality rates vector-borne diseases
Location of floods	Infrastructural data	Morbidity rates water-borne diseases
Time span	Coping capacity indicators	Mortality rates water-borne diseases
Type of floods	Geographical data	
Magnitude		

2.3.3 Data Analysis

With the use of data analysis tools, the collected data will be analyzed to identify patterns and correlations between country-specific characteristics and health outcomes. Data mining methods will be utilized to filter through the large datasets. An Exploratory Data Analysis (EDA) will then be conducted to better understand the data, summarize its main characteristics, and identify primary patterns through visualization methods. This analysis is divided into two distinct phases. The first phase involves examining the impact of floods on reported cases of communicable diseases to determine countries' resilience. To accomplish this, a visual analysis will be conducted using time series plots and histograms. To further investigate initial findings, three statistical tests are suggested to assess statistical significance: the Pearson correlation test, the paired t-test, and regression analysis. To further analyze the data and identify indicators describing the relationship between country-specific characteristics and increased incidences of communicable diseases after floods, a principal component analysis (PCA) will be performed. This method aims to gain insights into the nature of relationships within the data. Python will be used to conduct this analysis. A synthesis of the results of the data analysis will be conducted to combine and integrate the findings, generating an understanding of the system and reflecting on the initial hypothesis.

2.3.4 Framework and Future Research

The insights gathered from earlier sub-questions lay the groundwork for addressing the fourth and final research sub-question. Here, patterns, relationships, and other key findings from the EDA are evaluated based on the conceptual model. The findings are contextualized using a holistic approach to understand possible underlying drivers and factors. A system perspective is employed to incorporate the interrelations suggested by the data analysis among system elements as described in the conceptual framework. Systems thinking involves skills to understand systems, predict behaviors, and make changes for desired outcomes, such as recognizing interconnections and understanding dynamic behaviors. [Arnold and Wade, 2015] Using the principles of systems thinking, the identified patterns from the second and third sub-questions will be integrated into the framework. The primary data requirement centers on the comparative analysis of health outcomes in flood-prone regions. Finally, remaining gaps in understanding will be identified to formulate a future research agenda, guiding potential next steps in understanding the complex system relationships.

2.4 Research Structure

Figure 1 illustrates the research structure of the study including the main elements and primary methods used to write each chapter. The first sub-question will be addressed in chapter 3 and is then followed by chapter 4 that describes the development of the database. In chapter 5 and 6, the results of the data analysis will be presented and discussed. The insights of the data analysis will be evaluated based on the conceptual model and these will be used to develop the final framework, presented in chapter 7. The report will be finalised with an 8th chapter that discusses the results, gives some final recommendations and concludes the thesis.

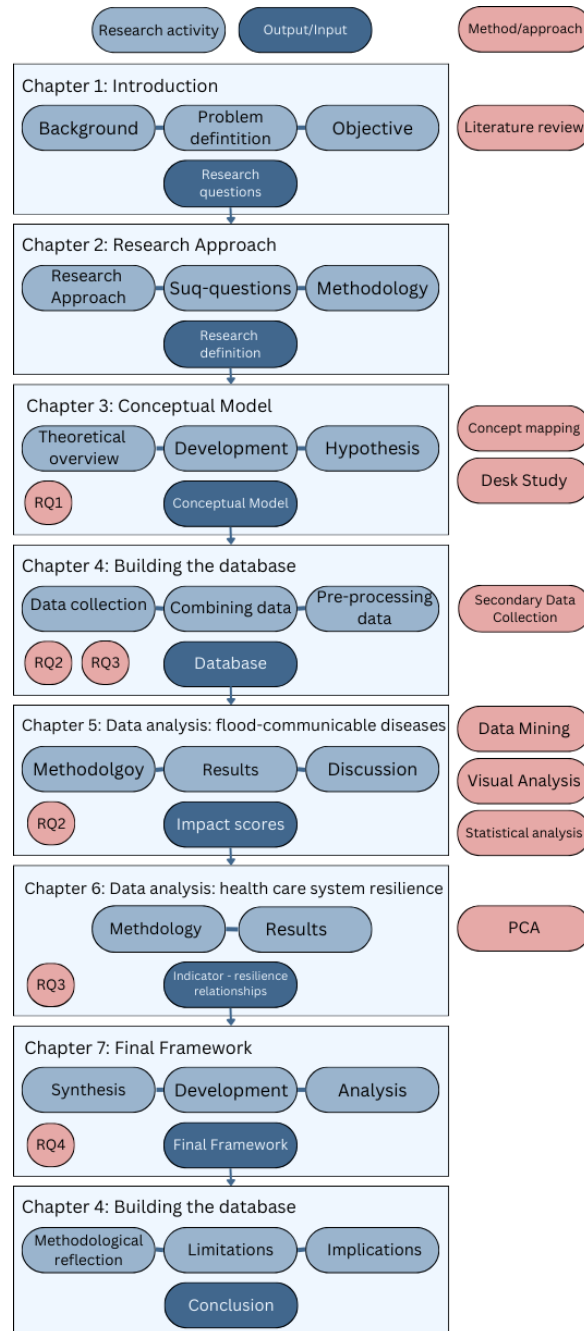


Figure 1: Research flow diagram describing the research activity, input/output and method used to construct each chapter of the thesis

Chapter 3

Conceptual Framework

3 Conceptual Model

In this chapter, the theoretical base of this research is being explained by utilizing past research studies and existing theories. To identify the main model domains, relationships, and determinants, four frameworks are being reviewed based on their objectives, drivers, and perspectives on key concepts. Furthermore, these frameworks are the foundation for the development of the conceptual framework that will be presented and explained as final part of this chapter.

3.1 Theoretical Framework

3.1.1 Flood disasters

Natural disasters can be explained as a situation or events where the ecosystem is distressed to an extent that overwhelms local capacity and a community's ability to adapt, often requiring external assistance [Liang and Messenger, 2018, Phalkey and Louis, 2016]. From all natural disasters, floods are the most frequent type and happen when an overflow of water submerges land that is usually dry [World Health Organization, 2024b, Abaya et al., 2009, Ochani et al., 2022]. Causes of floods vary and distinguish the different types of floods. Three common types of floods include riverine, flash and coastal floods, as illustrated in Figure 2 and briefly explained below [National Severe Storms Laboratory, 2023]:

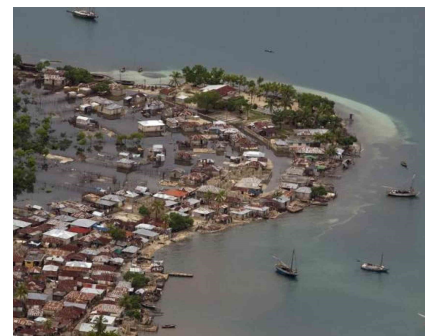
- **Riverine floods:** water levels rise over the top of river banks due to e.g. heavy rain, persistent thunderstorms, combined rainfall and snow melt, excessive tropical rain or an ice jam.
- **Flash floods:** excessive rainfall in a short period of time (<6 hours) cause torrents through river beds, urban streets or mountain canyon.
- **Coastal floods:** higher than average high tide cause inundation of land areas along the coast. These floods are often worsened by heavy rainfall and onshore winds.



(a) Riverine flood, Europe, 2022
[European Commission, 2023]



(b) Flash flood, India, 2013
[Sarkar, 2020]



(c) Coastal flood, Haiti, 2017
[PreventionWeb, 2017]

Figure 2: Three photos illustrating examples of three different type of common floods

Impact of climate change on floods

Floods are increasing in frequency and intensity due to climate change, demonstrating some of the most devastating consequences of extreme weather events [Bolan et al., 2024]. This increase is caused by several mechanisms that lead to alterations in the hydrological cycle. The primary cause, identified in the literature, is a rise in atmospheric moisture, as a warmer atmosphere can hold approximately

7% more moisture for every degree of warming [Bolan et al., 2024]. This extra moisture in the atmosphere leads to increased rainfall and a higher frequency of short, intense downpours, thereby increasing the risk of flash flooding, tropical cyclones, and hurricanes [Mallakpour and Villarini, 2015]. Furthermore, the additional heat in the atmosphere means there is more energy available for weather systems, altering the planetary water cycle and intensifying current climate patterns. As a result, precipitation becomes more concentrated with intense rainfall, heightening the risk of flooding in flood-prone areas. Lastly, climate change also elevates the risk of coastal flooding due to higher sea levels, as illustrated in Figure 3. [Liang and Messenger, 2018, Climate Council, 2022]

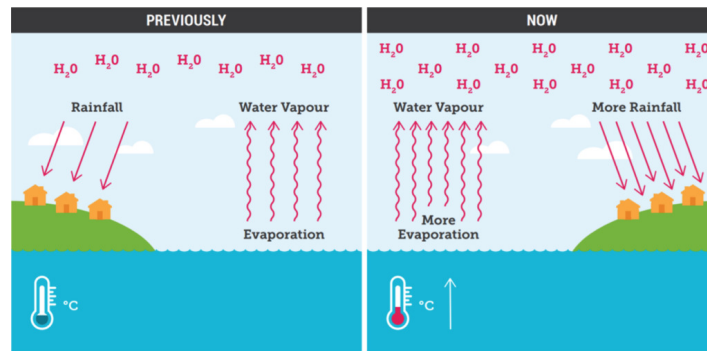


Figure 3: Effect of global warming on water cycle [Climate Council, 2022]

3.1.2 Health impact of floods

With the increasing frequency and intensity of flood disasters, it is likely that an increasing global health burden from the impacts will arise [Lee et al., 2020]. The health risks associated with floods encompass various sectors of health, and can be categorized into three groups based on their impact. The first group encompasses immediate health effects, occurring when the flood is present. The second group involves mid-term effects, spanning from days to weeks, and the third group consists of long-term effects, extending from months to years [Erickson et al., 2019]. An overview of the health impacts of floods is summarized in Figure 4.

1. Immediate health effects: These effects occur at the moment the flood occurs and includes trauma, injuries, skin infections, drowning, respiratory infections and gastroenteritis [Erickson et al., 2019, Mohajervatan et al., 2023]. A challenge identified in mitigating these effects is that affected people often require urgent medical care. However, the access to health care facilities and infrastructure is often disrupted due to significant damages [Okaka and Odhiambo, 2018, Englande, 2008]. Furthermore, health systems are considered to be insufficiently resilient, leading to inadequate responses during emergencies [Mohajervatan et al., 2023].
2. Midterm health effects: In the immediate aftermath of a flood disaster, the literature highlights various health effects, predominantly characterized by communicable diseases. Firstly, water contamination leads to water-borne diseases such as Diarrhoeal diseases, Hepatitis A, Leptospirosis, Typhoid fever, Vibrio Vulnificus, and Meningitis [Englande, 2008]. These communicable diseases are driven by factors like mass relocation, poor hygiene, and an increased presence of vector-borne diseases due to stagnant water providing breeding grounds. These include infections like coronavirus, influenza, measles, malaria, and dengue fever [Okaka and Odhiambo, 2018, Pal et al., 2016, Ochani et al., 2022]. Besides communicable diseases, there is a recognized health risk from toxicological diseases arising from chemical leaks and spills [Erickson et al., 2019].

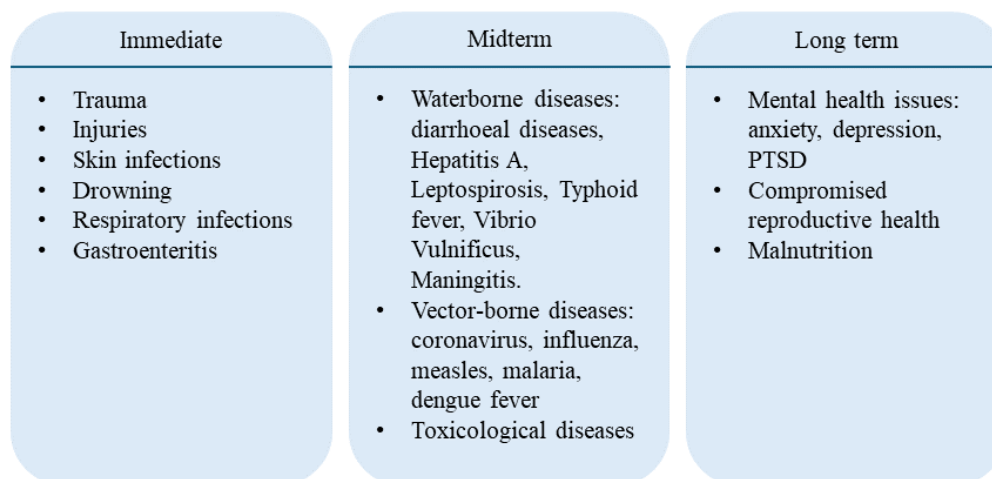


Figure 4: Overview of immediate, midterm and long term health impacts of floods

The prevalence of communicable diseases is substantial, with the specific disease varying and seeming to be influenced by multiple variables. [Mohajervatan et al., 2023, Abaya et al., 2009, Olanrewaju et al., 2019, Kouadio et al., 2012].

3. Long term health effects: Some health effects extend into long-term impacts, significantly affecting individuals well beyond the occurrence of the flood disaster. Particularly, reproductive health and mental health issues, including anxiety, depression, and post-traumatic stress disorder, emerge as primary concerns. The compromise of mental health seems to become apparent as a potential worldwide health effect of flood disasters. Additionally, the aftermath of a flood disaster may contribute to long-term malnutrition, with serious consequences. [Ray-Bennett et al., 2019, Mohajervatan et al., 2023, Abaya et al., 2009, Ochani et al., 2022].

3.1.3 Scope of Health Effects

The potential health impacts of major climatic hazards are compounded by the impacts they may have on health care systems: the institutions and services within society that are intended to protect and support human health [Few, 2007]. This thesis aims to analyse vulnerability and response not only in the sense of deriving indicators of risk, but in terms of understanding how and why the health impacts of hazard vary between groups in society and what determinants of the health system shapes the ability of people and institutions to cope [Few, 2007]. This thesis focuses on communicable diseases as they are considered to be mostly context-dependent, including the effectiveness of authorities and the emergency response within the health care system. In contrast, immediate health effects are often to distract from the intensity and geographical location of floods where long-term effects, especially mental health issues, are found worldwide and at every level of society, independent from geographical, institutional and socioeconomic factors. Since this research employs quantitative methods, relying on the availability of data, an initial search for epidemiological data of reported cases of communicable diseases was conducted to identify reported diseases in the context of flood hazards. During this search, literature was searched that linked an increase in incidence rates of a communicable disease to a flood disaster. As a result of this search, a Table was constructed indicating which diseases showed a positive correlation with floods. Following this result, the scope of diseases was determined to focus on: Cholera, Leptospirosis, Diarrhoeal disease, Malaria and Dengue. The Table summarising these results can be found in appendix A.

3.1.4 Principles of infectious disease after hydrologic disasters

Floods pose a significant threat to public health, particularly due to their potential to exacerbate the transmission of communicable diseases. The primary causes of such diseases in disaster scenarios can be categorized into four main areas: infections stemming from contaminated food and water, respiratory infections, vector-borne diseases, and infections resulting from wounds and injuries [Ligon, 2006]. In the context of floods, the rising waters create stagnant pools, which become breeding grounds for insects and other disease vectors. These conditions foster the proliferation of vector populations, increasing the risk of vector-borne disease transmission. Additionally, floods disrupt water supply and sanitation systems, contaminating these systems and potentially leading to waterborne diseases like Cholera, Diarrheal illnesses and Leptospirosis [Basaria et al., 2023].

Understanding the risk of infectious diseases following a flood involves considering the classic epidemiological triad: the external agent (microorganism), the susceptible host, and the environment that facilitates contact between the agent and host [Frost, 1976]. In flooded areas, the agents responsible for infections are typically those that existed in the region before the disaster, albeit with varying impacts on human health. Hosts, including survivors and responders, become vulnerable to infection through traumatic injury and exposure to contaminated environments. Factors such as poor hygiene, inadequate sanitation, and limited access to clean water and food further increase susceptibility to communicable diseases. Floods disrupt environmental barriers, contaminating water sources with sewage, wastewater, and agricultural runoff, thereby heightening the risk of disease transmission [Liang and Messenger, 2018]. The epidemiological triad is illustrated in Figure 5.

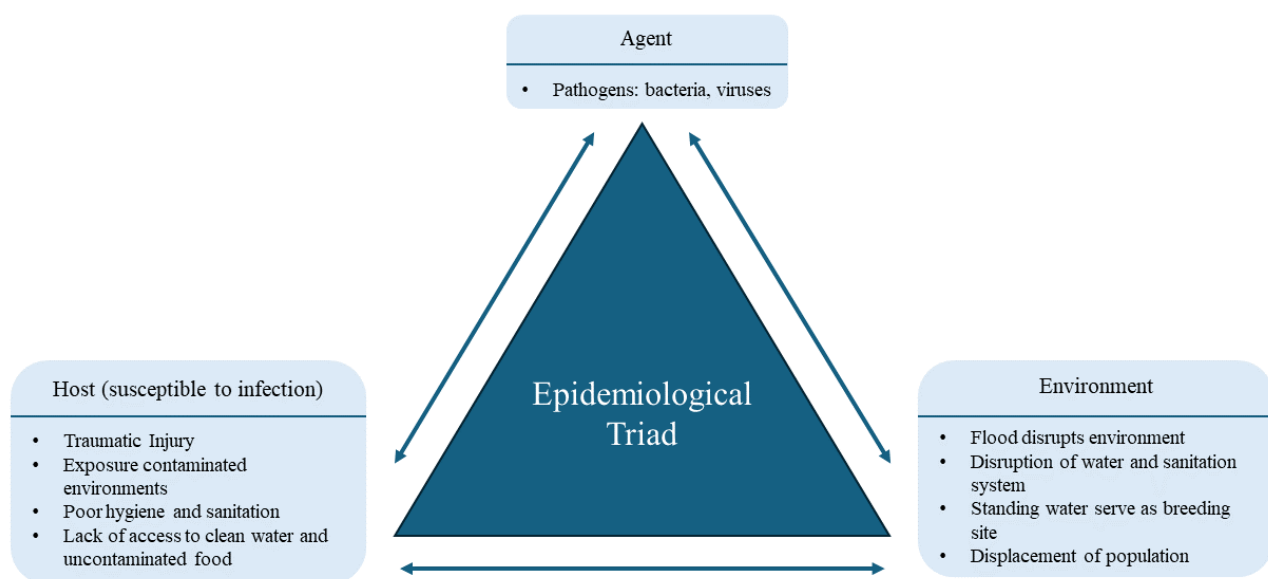


Figure 5: Epidemiological triad describing the relationship between the agent, host and environment

3.1.5 Lag period communicable diseases

When analyzing potential outbreaks of vector-borne and waterborne diseases following floods, it is crucial to consider the lag period. This time frame accounts for various ecological and biological processes that influence disease transmission. For example, heavy rains and flooding can lead to increases in vector populations, potentially causing a rise in disease transmission within 3-4 weeks

[Shortus et al., 2016]. Similarly, the risk of diseases such as malaria can increase due to the development cycles of vectors and pathogens, extending up to two months after increased precipitation and temperature [Gonzalez-Daza et al., 2023]. Waterborne diseases like Leptospirosis, which are influenced by the survival of pathogens in water and soil, show associations with rainfall over a period of 1-3 months, necessitating a 12-week lag period for accurate modelling [Chadsuthi et al., 2021]. Additionally, studies have shown consistent results with time lags of 1-2 months for disease incidence following flooding [Suwanpakdee et al., 2015]. Although the exact lag time for communicable diseases is difficult to determine precisely, based on previous research, a lag period of one to two months will be considered in this research.

3.1.6 Mechanisms Cholera Transmission

Extensive research has been conducted to explain the mechanisms driving cholera transmission. Initially, this research focused on waterborne diseases in general; however, cholera emerged as the most relevant due to its significant association with flooding and availability of data. In this section, the primary mechanisms that explain the relationship between flooding and cholera incidence, as highlighted in the literature, are discussed. Additionally, other factors that may influence the dynamics of cholera transmission are explored. These mechanisms are illustrated in Figure 6a.

Cholera transmission is intricately linked to environmental factors, among which those associated with flooding. Flooding events causes several processes that might lead to a change in transmission patterns. First of all, flooding lead to the mixing of sewers, exposed drains, reservoirs, and rivers, resulting in significant water contamination with Cholera [Akanda et al., 2009, Shackleton et al., 2023]. Floods not only adversely affect water sources but also sewerage systems, escalating the exposure of populations to cholera-causing bacteria [Hashizume et al., 2013]. This is intensified in case of urban floodings, when the risk of contact between individuals and these contaminated floodwaters is further increases [Mark et al., 2015]. Also, surface runoff during rainfall carries organic sediment, including fecal waste, into water sources, further contributing to contamination [Shackleton et al., 2023]. On a more biological level, it is suggested that the proliferation of Cholera is also facilitated by the increase in insoluble iron levels during flooding, improving the survival rate of Cholera [Hashizume et al., 2013].

Beyond flooding, other mechanisms such as low rainfall levels leading to higher bacterial concentrations and warmer ambient temperatures favoring bacterial proliferation are found to play a role in Cholera transmission. Furthermore, ocean temperatures also indirectly influence cholera dynamics on land through changes in monsoon rainfall and increased ambient temperatures, which Cholera prefers. Moreover, an increase in phytoplankton concentration directly correlates with higher Cholera abundance, as certain phytoplankton act as reservoirs for the bacteria, providing them with essential nutrients, and alter water pH to favor bacterial survival and reproduction. [Shackleton et al., 2023, Lutz et al., 2013]

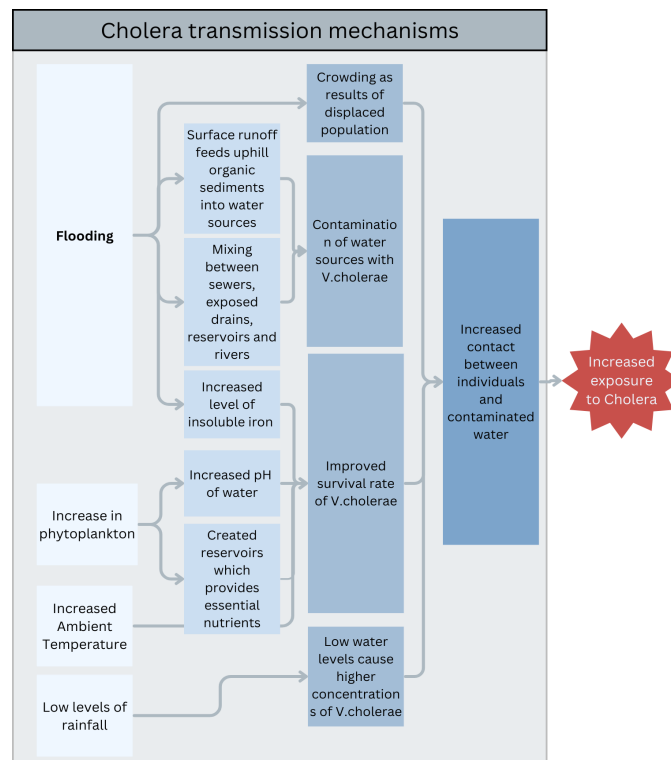
Lastly, socio-economic factors, such as population density and degraded sanitation conditions, are recognized as significant contributors to Cholera transmission, fostering increased human contact and providing favorable conditions for bacterial persistence [Shackleton et al., 2023]. In the context of this thesis, it is assumed that the flood-Cholera relationship has been established by prior literature. Through data analysis, the aim is to evaluate healthcare system determinants, to evaluate which determinants potentially affect the exposure and resilience to waterborne diseases in the context of floods.

3.1.7 Mechanisms Vector-borne Disease Transmission

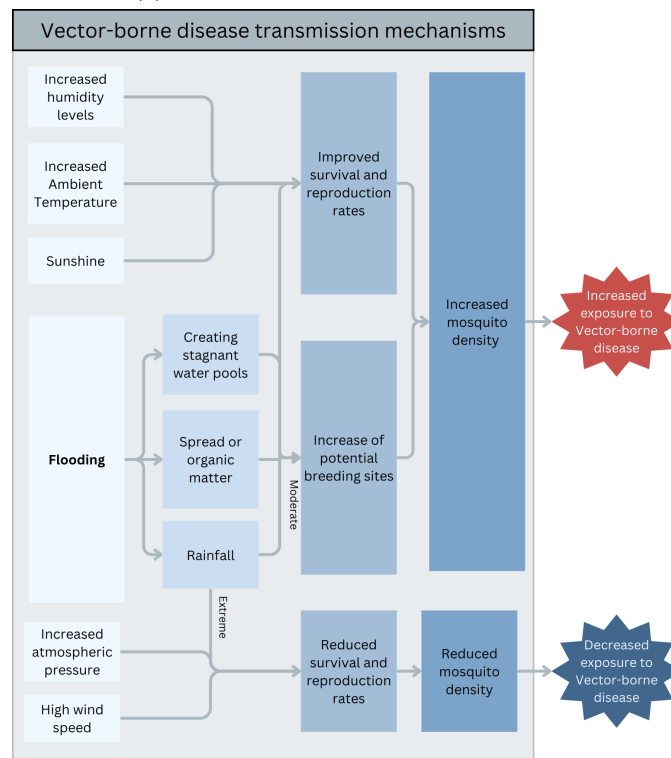
The complex relationship between flooding and the incidence of vector-borne diseases, such as Malaria and Dengue, is shaped by diverse meteorological factors. A primary mechanism driving this relationship is the role of flooding in increasing mosquito density through the creation of stagnant water pools and the spread of organic matter, providing additional breeding habitats. Factors influencing the impact of flooding on mosquito populations include climate patterns (like rainfall, temperature, and humidity) and water salinity [Coalson et al., 2021]. Flooding-induced precipitation can create favorable environments for mosquito breeding, facilitating their survival and reproduction. While moderate precipitation levels have been observed to enhance vector populations and disease transmission, extreme rainfall associated with flooding can produce contrasting outcomes as it may initially flush out breeding sites, thereby reducing vector density and disease transmission. [Wu and Huang, 2022, Morin et al., 2013, Viana and Ignotti, 2013] As heavy precipitation can also leave stagnant water behind, offering potential breeding grounds for adult mosquitoes, past studies indicate short-term (up to 1 month) decreases and subsequent (1–4 month) increases in incidence [Coalson et al., 2021]. This dual effect highlights the nuanced dynamics of flooding on vector-borne diseases [Wu and Huang, 2022].

Beyond flooding, several other meteorological and ecological factors play crucial roles in driving the transmission of vector-borne diseases. Rising temperatures, for instance, can accelerate the incubation and reproduction rates of disease vectors, amplifying their numbers and spread rate [Wu and Huang, 2022]. Similarly, increasing humidity levels facilitate disease transmission by creating more favorable environments for vector survival. Sunshine has also been positively correlated with disease incidence, while atmospheric pressure and wind speed exhibit negative associations. [Viana and Ignotti, 2013] Moreover, socio-ecological factors, changes in human host and vector behavior, and parasite genetic changes contribute significantly to disease transmission dynamics [Savi, 2022]. Another relevant aspect is the genetic complexity of the vector and the circulation of different serotypes, which likely also influence Dengue distribution, both in dry and rainy periods, and its ability to adapt to the human environment through breeding sites. Therefore, the vector does not depend exclusively on abiotic factors [Viana and Ignotti, 2013]. Understanding these multifaceted mechanisms is essential for developing effective strategies for disease control and mitigation efforts. The key factors explaining the flooding-vectorborne diseases are illustrated in Figure 6b.

In contrast to waterborne diseases, vector-borne diseases are found to be less dependent on socio-economic factors. However, monitoring, control and awareness strategies are deemed to be of increased importance [Savi, 2022], for example by increasing vaccination levels, perform early diagnostics tests and increase vector control [Khan et al., 2023].



(a) Transmission mechanism cholera



(b) Transmission mechanisms vector-borne diseases

Figure 6: Transmission mechanisms of cholera transmission, representing waterborne diseases and vector-borne diseases

3.2 Conceptual Framework

3.2.1 Selection of frameworks and studies

Extensive desk research identified relevant frameworks, emphasizing a systemic perspective beyond the medical field. Selected frameworks were chosen based on specific criteria for their comprehensive understanding of the healthcare system. Diverse domains and corresponding drivers were identified to construct a conceptual framework. This approach aims to encourage an holistic understanding of the complex healthcare environment by integrating insights from various disciplines. Before reviewing the frameworks and developing the conceptual framework, it is crucial to acknowledge their diverse objectives and perspectives, each contributing to a comprehensive understanding of the healthcare system's dynamics and its relationship with flood impacts on health.

Initially, the assessment explored the social determinants of health, laying the foundation for understanding the healthcare system's definition. Subsequently, Few's impact pathway for climatic hazards was evaluated, placing the healthcare system within the context of climatic risks. Following this, the framework for assessing urban settlements' vulnerability to flood risks was examined, bridging the conceptual model with flood risks specifically. Then, the Resilience Framework for Public Health Emergency Preparedness was reviewed to assess the system's resilience. Finally, various frameworks and supplementary studies were consulted to identify indicators for evaluating the system. [Roland et al., 2021] Below, brief descriptions of the frameworks are given and Table 3 shows the objectives and identified drivers for each framework. In appendix B, the frameworks are presented.

1. Social Determinants of Health (SDOH) are the non-medical factors that significantly influence health outcomes, including individuals' living conditions, socioeconomic status, education, employment, social support networks and access to healthcare. Addressing these determinants is essential for effectively managing the healthcare landscape from a systemic perspective. [Ansari et al., 2003, Kumar, 2010]
2. Few's health impact pathway for climatic hazards aims to analyze vulnerability and adaptation to climate hazards through the lens of environmental and social sciences, specifically focusing on health impacts. This framework seeks to explore how vulnerability to health impacts varies within societies and proposes a structured approach to understanding the interconnections between climatic hazards and health concerns. [Few, 2007]
3. The framework for assessing urban settlements' vulnerability to flood risks integrates cultural and behavioral factors to comprehend the socially constructed vulnerability among at-risk populations. Additionally, it incorporates urban political ecology to analyze socio-spatial-political profiles and institutional drivers of flood vulnerability. The overarching goal of this framework is to identify indicators, understand root causes, and develop flood risk management tools to reduce vulnerability and mitigate impacts. [Salami et al., 2017]
4. The Resilience Framework for Public Health Emergency Preparedness outlines the essential components of a resilient system, aimed at enhancing readiness for disasters and emergencies. This conceptual framework surpasses simple description by integrating principles of complexity, illustrating the dynamic nature of the system. Its primary objective is to identify proactive measures to strengthen preparedness and responsiveness in the face of crises. [Khan et al., 2018]

3.3 Review

The review of these frameworks serves several purposes. Firstly, it aims to define key concepts such as the healthcare system, factors influencing health outcomes during flood disasters, vulnerability, and indicators for measuring system performance. Another objective is to compare the objectives and identified drivers of the frameworks. This subsection presents the findings of the review.

3.3.1 Defining key concepts

Health care system and the determinants of health

To assess a health system's performance, it must first be clearly defined, including its boundaries and components. These boundaries define responsibilities and significantly influence the analysis's ability to identify factors impacting health system outcomes. In this research, it is essential to recognise the broader context of the health system and its interactions with the economic, political, and social surroundings. Aligning with the Health System Performance Framework [Frenk and Murray, 2000], the definition is set to:

“The resources, actors and institutions related to the financing, regulation and provision of health actions. Where health actions are any set of activities whose primary intent is to improve or maintain health.”

Improving or maintaining health can be done directly by improving service and quality of health services, however, health can also be indirectly improved by improving the environment in which people live. The health care system is thus more than just the performance and quality of health care facilities but rather is a combination of factors affecting the health of individuals and communities. The context of people's lives determine their health. All these factors is referred to as the determinants of health and include, among many other factors [World Health Organization, 2024]:

- Economic environment
- Individual behavior
- Social status
- Physical environment
- Education
- Social support networks
- Access to health services

Vulnerability

This thesis seeks to evaluate countries' vulnerability to communicable diseases following flood events. It focuses primarily on identifying indicators for risk assessment rather than strategies for mitigating these impacts. Each framework offers its unique definition of vulnerability.

Vulnerability is ...

“the susceptibility or exposure of disadvantaged individuals or groups to health-damaging conditions due to their unequal socioeconomic positions or circumstances.” [Kumar, 2010]

Vulnerability is ...

“a combination both of physical vulnerability (the likelihood of physical exposure to the hazard) and social vulnerability (susceptibility to its impacts).” [Few, 2007]

Vulnerability is ...

”circumstances triggered by various phenomena in the form of physical, social, economic, cultural and environmental factors which make a society, system or asset susceptible to natural and human-made hazards.” [Salami et al., 2017]

Vulnerability is therefore a central concept and is a complex term that can be explained as a combination of both physical and social vulnerability [Few, 2007]. It encompasses exposure (the risk of flood occurrence, infrastructure, land use), vulnerability (socioeconomic factors, demographics), and resilience or coping capacity (health systems, governance) [Salami et al., 2017]. These perspectives will form the foundation of the conceptual framework of this study, aiming to comprehend the complex relationship between floods, health systems, and the impacts of communicable diseases, drawing from existing theories, frameworks, and studies.

Resilience

Resilience, often explained as coping capacity, plays an important role in understanding the complex dynamics of healthcare systems. Within the Social Determinants of Health, as described by the WHO, coping capacity finds its place among the intermediary determinants, encompassing the psychosocial circumstances individuals face, including stressors, living conditions, and support systems [Few, 2007]. Here, the focus predominantly lies on individual resilience. However, as Few elaborates, resilience extends beyond the individual level to embrace collective coping capacities, shaped by a range of resources, behaviors, and broader societal processes. While Few emphasizes resilience in the context of health risks, particularly highlighting the ability to avoid infection and sustain functional health systems during hazardous events, Salami delves deeper, shedding light on adaptive coping mechanisms influenced by various factors such as perceptions, awareness, and strategies. In this exploration, attention is directed not only towards individual resilience but also institutional resilience. [Salami et al., 2017] Khan explores the crucial role of collaboration, trust, and community engagement in fostering resilience within public health systems. [Khan et al., 2018] These elements, along with others outlined in frameworks for resilient public health systems, emphasize the human and social dimensions essential for strengthening resilience and improving adaptive capacity. In essence, resilience emerges as a multifaceted concept encompassing individual, community, and institutional capacities to endure and rebound from adversity.

3.3.2 Review of Objectives and drivers

Table 3 provides an overview of the objectives and drivers of the selected frameworks. The objective of all frameworks is to address health-related issues and to provide an understanding on how to mitigate health risks or improve health outcomes within their respective area of focus. However, each framework takes a distinctive perspective on the context of health outcomes of flood disasters and these will be explained by identifying the relevant drivers of these frameworks.

Within the healthcare system, various categories of drivers contribute to explaining health system performance. To structurally assess the impact of flood hazards on healthcare system performance and health outcomes, several groups of indicators were identified through the assessment of multiple frameworks. All frameworks adopt a broad perspective on the drivers of the healthcare system, considering the political and institutional context. Additionally, they emphasize the significance of physical infrastructure and material circumstances. The CSDH identifies the socio-economic position of the individual as a primary determinant, along with behavioral factors, biological factors, and psychosocial factors, placing particular emphasis on individual positions. The Health Impact Pathway for

flooding focuses more on physical exposure to risk, including flood hazards and physical proximity. Salami approaches vulnerability from both an individual and systemic standpoint. The Public Health Emergency Preparedness framework takes a holistic approach, considering institutional, political, and cultural factors to evaluate the resilience of public health systems. These groups encompass physical exposure to flood hazards and disease outbreaks, socioeconomic and demographic factors, health system resilience, and institutional resilience.

Table 3: Overview of reviewed frameworks addressing health and flood risk management

Author / Institution	Name framework	Objective	Drivers
Solar O, Irwin A. World Health Organization	Commission on Social Determinants of Health (CSDH) framework	To provide an understanding of the SDH by delineating how social, economic, and political mechanisms contribute to socioeconomic positions, which subsequently influence individuals' exposure and vulnerability to health-compromising conditions.	Structural determinants: Socio-economic position, political context. Intermediary determinants of health: Material circumstances, behavioral factors, biological factors, psychosocial factors
Few R.	Health Impact Pathway for flooding	To integrate health considerations into the existing discourse on hazards, risk, and vulnerability, particularly within the realms of political ecology, disaster studies, and climate change adaptation.	Flood hazard, physical proximity, health risk effect, health outcome in social, cultural, economic, political, environmental context
Salami R, von Meding J, Giggins H	Flood vulnerability assessment framework	To address human settlements' vulnerability to flood disaster risk in cities by providing deep understanding of the flood risks.	Exposure, susceptibility and adaptive coping capacity in the context of households' or communities' social, economic, cultural, institutional and physical vulnerabilities.
Khan Y, O'Sullivan T, Brown A et al.	Public Health Emergency Preparedness framework	To describe the essential elements of a resilient public health system and how the elements interact as a complex adaptive system.	Institutional, political and cultural factors, material circumstances

3.3.3 Conceptual Framework Development: Defining Framework Domains

Based on the frameworks reviewed previously, the conceptual framework has been developed and is presented in Figure 7. In this model, the sources are indicated by numbers that correspond to those listed in Table 4. In this framework, the flood-disease risk can be explained as the risk of a potential communicable disease outbreak after a flood event. This flood-disease risk is influenced by four separate domains. Each domain either drives or reduces this risk and is affected by multiple factors. The conceptual framework provides valuable insights into identifying the four main domains that influence the flood-disease risk of a country: exposure to communicable diseases, exposure to floods, vulnerability of a flood-prone community, and resilience to the impact of floods on the community. In this subsection, these domains will be further elaborated on.

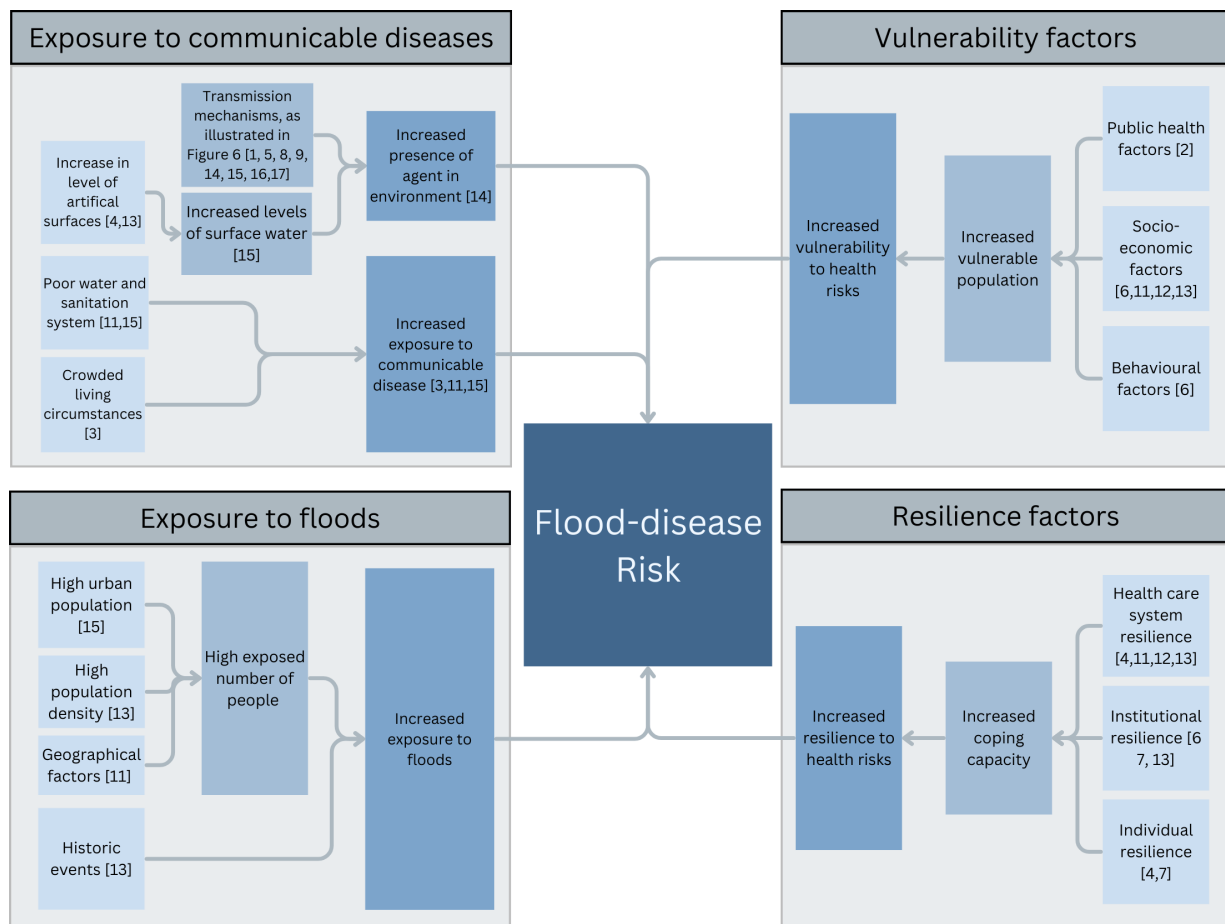


Figure 7: Conceptual framework of the relationship between floods, healthcare systems and communicable diseases

Table 4: Sources for the Conceptual Framework

Number	Source	Number	Source	Number	Source
1	[Akanda et al., 2009]	7	[Khan et al., 2018]	13	[Salami et al., 2017]
2	[Ansari et al., 2003]	8	[Lutz et al., 2013]	14	[Savi, 2022]
3	[Elimian et al., 2020]	9	[Morin et al., 2013]	15	[Shackleton et al., 2023]
4	[Few, 2007]	10	[Ndicunguye, 2012]	16	[Viana and Ignotti, 2013]
5	[Hashizume et al., 2013]	11	[Phalkey and Louis, 2016]	17	[Wu and Huang, 2022]
6	[Kumar, 2010]	12	[Roland et al., 2021]		

Exposure to Communicable Diseases

This domain focuses on understanding the physical exposure of a country or region to communicable diseases and its implications for the transmission of these diseases. Key factors include the epidemic risk index, which measures the likelihood of disease outbreaks in a given area. Material circumstances such as the quality of water and sanitation systems, the quality of housing, and land use also impact the magnitude of potential exposure to disease. Environmental mechanisms, explaining the transmission of communicable diseases, as illustrated in figure 6, are also included in the model.

Exposure to Floods

The second domain is determined by geographical and demographic factors. Geographical factors that could increase the exposure to floods include proximity to water bodies, such as rivers and coastlines, especially in regions with frequent and intense weather patterns like hurricanes and heavy monsoon rains. Additionally, low-lying topography exacerbates this vulnerability by allowing water to accumulate more easily, leading to severe and prolonged flooding. Indicators such as flood frequency, the annual number of people expected to be exposed to floods, population density, and the degree of urbanization help define a region's exposure to flood hazards. These factors contribute to understanding how many people are at risk during flood events and how often these events are likely to occur. High population density and urban areas tend to have higher exposure levels, which can amplify the risk of disease outbreaks post-flood.

Vulnerability

Individual vulnerability within communities can be explained as a set of conditions determined by various phenomena in the form physical, social, economic, cultural and environmental factors which make a community susceptible to natural and human-made hazards [Salami et al., 2017]. Factors determining the level of vulnerability include average household income, education levels, malnutrition, and age. Flood hazards themselves impact community vulnerability, particularly when recovery from previous events is incomplete, leading to disrupted income, damaged infrastructure, and food loss. Vulnerability extends beyond physical exposure to for example floods. Within this thesis, vulnerability is categorized into three main categories of factors: public health factors, describing the current health state of the population, socio-economic factors, affecting a persons' ability to engage in health activities, afford medical care and housing, and manage stress. and behavioral factors, describing skills to act properly in vulnerable situations.

Resilience

Resilience addresses a community's capacity to cope with flood risk in terms of their healthcare system, institutional resilience, and individual resilience. This domain reflects the efforts to enhance resilience in flood-prone areas through three distinctive categories: healthcare system, institutional and individual resilience. The ability to recover and adapt to flood events, maintaining functionality and reducing long-term impacts, is critical in mitigating flood-disease risks. Efforts to improve resilience include strengthening health care infrastructure, implementing early warning systems, and fostering community-based initiatives to support recovery and preparedness.

By examining these domains and their indicators, the conceptual model provides a comprehensive understanding of how various factors interact to influence the flood-disease risk of a community. This framework guides data analysis and helps in developing strategies to mitigate the risk of disease outbreaks following flood events.

3.3.4 Conceptual Model Development: Interrelationships

The conceptual framework presented in Figure 7 outlines the system of interlinked concepts, assumptions, and theories guiding this master thesis. These theories serve as a foundation to conceptualize the system for data analysis and to explore assumed relationships and interrelations within the system. The framework identifies various factors that might affect flood-disease risk, aiming to understand how these factors relate to each other and identify their interconnections. This conceptual framework forms the basis for the data analysis, guiding the thesis in two phases. In the first phase, past flood events and their impacts on the incidence of communicable diseases will be evaluated, focusing on exposure to both floods and diseases. This evaluation will inform the selection of countries, analyzing patterns between floods and disease incidences. Countries will then be categorized based on the historical impact of floods on disease incidence. In the second phase, the goal is to determine which factors from either the vulnerability or resilience categories explain these distinctions. Although literature suggests potential contributing or mitigating factors, this research aims to identify the most influential ones. The objective is to determine which category of factors has a stronger influence on flood-disease risk. Therefore, vulnerability and resilience factors will be further described in the following chapters, and indicators will be defined for analysis.

3.3.5 Initial list of indicators

Based on the frameworks and past studies reviewed during the desk study, an initial list of indicators is constructed, shown in Table 5. This list will be the starting point for collecting data for the database.

Table 5: Initial list of indicators based on the reviewed frameworks

Domain	Category	Indicator	Sources
Exposure to communicable diseases	-	Land use	[Salami et al., 2017, Few, 2007, Ndicunguye, 2012]
Exposure to communicable diseases	-	Quality of water and sanitation system	[Phalkey and Louis, 2016, Salami et al., 2017, Roland et al., 2021, Ndicunguye, 2012]
Exposure to floods	-	Annual expected people exposed to floods	[Salami et al., 2017]
Exposure to floods	-	Flood frequency	[Phalkey et al., 2017]
Exposure to floods	-	Population density	[Phalkey and Louis, 2016]
Exposure to floods	-	Proximity to river	[Salami et al., 2017]
Vulnerability	Behaviour	Education	[Salami et al., 2017, Roland et al., 2021, Kumar, 2010, Ansari et al., 2003]
Vulnerability	Behaviour	Literacy	[Roland et al., 2021]
Vulnerability	Socioeconomic	Age	[Phalkey and Louis, 2016, Salami et al., 2017, Ndicunguye, 2012, Ansari et al., 2003]
Vulnerability	Socioeconomic	Average household income	[Roland et al., 2021, Kumar, 2010, Ndicunguye, 2012, Ansari et al., 2003]
Vulnerability	Socioeconomic	Family size	[Salami et al., 2017, Roland et al., 2021]
Vulnerability	Socioeconomic	Gender	[Phalkey and Louis, 2016]
Vulnerability	Socioeconomic	House types	[Salami et al., 2017]
Vulnerability	Socioeconomic	Occupation	[Salami et al., 2017]
Vulnerability	Socioeconomic	Poverty	[Ansari et al., 2003]
Vulnerability	Socioeconomic	Quality of housing	[Phalkey and Louis, 2016, Ndicunguye, 2012, Ansari et al., 2003]
Vulnerability	Socioeconomic	Quality of neighborhood	[Salami et al., 2017]
Vulnerability	Socioeconomic	Race	[Salami et al., 2017, Kumar, 2010, Ansari et al., 2003]
Resilience	Health care system	# of health workers / population	[Roland et al., 2021, Khan et al., 2023]
Resilience	Health care system	# of hospital beds / population	[Roland et al., 2021]
Resilience	Health care system	# of hospitals / population	[Roland et al., 2021]
Resilience	Health care system	% of population with health insurance	[Roland et al., 2021]
Resilience	Health care system	Accessibility of health care system	[Phalkey and Louis, 2016, Few, 2007]
Resilience	Health care system	Access to insurance	[Salami et al., 2017, Few, 2007]
Resilience	Health care system	Availability of resources	[Phalkey and Louis, 2016, Roland et al., 2021, Khan et al., 2018]
Resilience	Health care system	Distance to hospitals	[Roland et al., 2021]
Resilience	Health care system	Health expenditure per capita	[Frenk and Murray, 2000]
Resilience	Health care system	Quality of health care system	[Phalkey and Louis, 2016]
Resilience	Individual	Civic participation	[Salami et al., 2017, Few, 2007]
Resilience	Individual	Flood insecurity	[Few, 2007]
Resilience	Individual	Social cohesion	[Salami et al., 2017, Ansari et al., 2003]
Resilience	Institutional	Communication	[Khan et al., 2018]
Resilience	Institutional	Corruption	[Phalkey et al., 2021]
Resilience	Institutional	Emergency response plans	[Khan et al., 2018]
Resilience	Institutional	Trust in institutions	[Salami et al., 2017]
Resilience	Institutional	Warning systems	[Salami et al., 2017, Khan et al., 2018]

Chapter 4

Database Development

4 Database Development

The development of the database is the critical phase of this research as it determines the scope for exploring the interrelations between health care systems, communicable diseases outbreaks and floods. This chapter aims to provide the required database necessary to answer the second and third research questions that explore the correlations between health system characteristics and the incidence of communicable diseases after floods. To systematically collect this data, a search strategy was formulated. First, an initial selection of countries took place to frame the search. Then, the process continued with several parallel processes where three types of data were retrieved. First, the flood data highlighting past flood disasters, affected people and frequency. Most importantly, the seasonal patterns and time span of the floods were mapped. Secondly, epidemiological data was gathered through numerous databases and website to retrieve as detailed possible data on the surveillance of communicable diseases during time of flood disasters. Lastly, vulnerability and resilience indicators were collected. These indicators were identified based on the conceptual model presented in the previous chapters and corresponding theories. The data collection process is illustrated in figure 8.

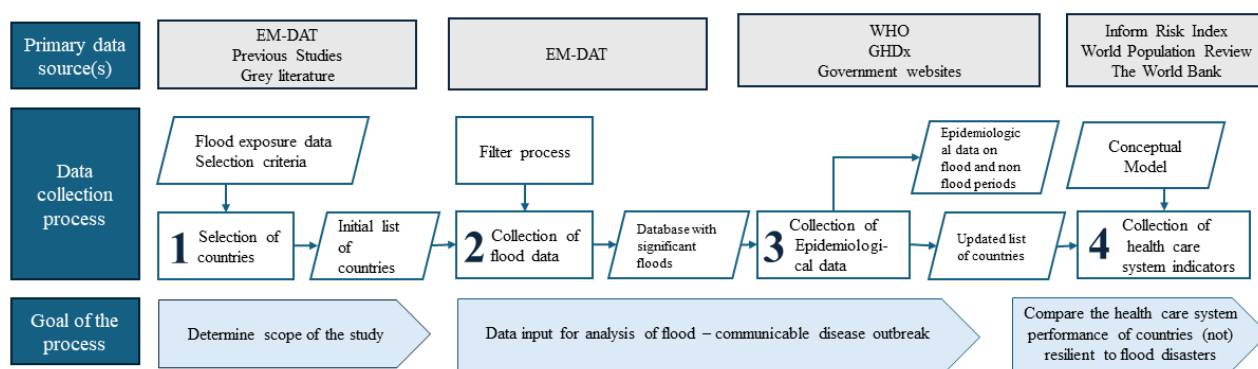


Figure 8: Data collection process for database development, including primary data sources and process goals.

4.1 Data sources

To develop the database, an extensive search for data was performed to acquire a set of data relevant to the context of this research. Therefore, a wide range of data sources was used to collect the required data. The primary data sources used for the development of the database are discussed in this subsection.

1. EM-DAT

EM-DAT is a valuable data repository that systematically documents human and economic losses worldwide in the aftermath of disasters meeting specific criteria. These include incidents resulting in at least 10 fatalities, affecting 100 or more people, triggering a declaration of a state of emergency, or necessitating international assistance. The database categorizes disasters as events that overwhelm local capacities, prompting requests for external aid. These unpredictable and often abrupt occurrences cause substantial damage, destruction, and human distress. EM-DAT offers users access to flood-related data from 2000 onwards for all countries globally. Despite its broad coverage, occasional data gaps may indicate either relatively minor floods or inadequacies in reporting, emphasizing the need for careful interpretation. [Centre for Research on the Epidemiology of Disasters, 2024]

2. World Health Organization

The World Health Organization (WHO) is an United Nations agency responsible for international public health with primary objective to ensure that all people attain the highest possible level of health. The WHO plays a crucial role in responding to health emergencies, promoting health equity, and addressing various health challenges worldwide, such as infectious diseases, noncommunicable diseases, and the impact of environmental factors on health. The global health observatory served as an important source for the collection of data for this research as the database allows for a structure search based on indicators, categories or countries. It was used to collect epidemiological data on Cholera and Malaria and also provided input for several health system performance indicators, specifically health system resilience and exposure indicators. [[World Health Organization, 2024b](#)]

3. Global Health Data Exchange

The Global Health Data Exchange (GHDx) is a platform developed and maintained by the Institute for Health Metrics and Evaluation (IHME) and is a catalog of health-related data from various sources. The GHDx provides access to a wide range of health data, including information on disease prevalence, mortality rates, risk factors and health system performance. The IHME's main goal is to provide data on important health metrics to policymakers, researchers and the public. For this research, the GHDx was primarily consulted to collect worldwide epidemiological data on Malaria, Diarrhoeal Disease, Cholera and Dengue.

4. Ministries of health

Although the WHO and GHDx provided a variety of datasets on epidemiological, it mainly consist of yearly numbers. To identify seasonal patters in incidence rates and to be able to analyse the relationship between floods and communicable diseases outbreaks, the government website, specifically their ministries of health, of all countries were searched for datasets, reports or dashboards on monthly or weekly incidence rates of communicable diseases.

5. INFORM Risk Index

The INFORM Risk Index is a global risk assessment for humanitarian crises and disasters developed by UN agencies, donors, NGOs and research institutions. It identifies countries at high risk of humanitarian crises including floods. The Index model has three dimensions of risk: Hazards and exposure, vulnerability and lack of coping capacity. For this research, it provided valuable data on the health system performance indicators, specifically on the exposure and socio-economic indicator domains of the conceptual framework.

6. World Population Review and the World Bank

Various demographic data and statistics related to global population trends are documented by the world population review website. The world bank has an open access data base that provides data on global development. These data sources were used to identify characteristics with regard to demographics and the socio-economic indicators of countries.

4.2 Data collection process

4.2.1 Selection of countries

For this thesis, the top 20 countries of three different measurements of flood risk have been selected to compile an initial list. These measurements include:

1. Countries most exposed to floods by risk index [Salas, 2023]
2. Origin of floods with the highest total affected people since 2000 worldwide, as well as per continent, to understand variation in the characteristics of the selected countries obtained through the public EM-DAT platform
3. Percentage of population at risk [Rentschler et al., 2022]

This process resulted in a list of 41 countries. Subsequently, an online search for epidemiological evidence from past flood disasters was conducted to evaluate the relevance of including these countries in the research. This list of countries was then used to collect data on communicable diseases outbreaks after past flooding disasters. During this initial phase of the data collection, another five countries were excluded due to insufficient availability of relevant data. The initial list of countries, presented in Table 6, will serve as the starting point for the project.

Table 6: Initial set of flood-prone countries included for analysis

Asia	Africa	America	Europe	Oceania
Bangladesh	Mozambique	Dominican Republic	Russia	Australia
Vietnam	Tanzania	Peru	Czech Republic	
Cambodia	Nigeria	Guyana	Serbia	
Iraq	Somalia	Guatemala	United Kingdom	
Sri Lanka	South Sudan			
Pakistan				
China				
India				
Philippines				

4.2.2 Flood data

The initial set of countries served as the starting point for collecting relevant data on past floods. The International Disaster database was utilized to retrieve this data, providing information on the type and origin of the flood, location, magnitude, time span, total affected people, and total damage. This data was retrieved to gather information regarding countries' past experiences and vulnerabilities concerning flood disasters. Firstly, the frequency of flood disasters occurring in a country was calculated, alongside the number of floods and days of flooding per year. Additionally, it is necessary to identify the total number of affected people per year and the seasonal distribution of floods.

However, the criteria for inclusion in the International Disaster Database permit the reporting of flood disasters that might not directly impact the incidence of communicable diseases. Referring back to the conceptual framework, health risk factors that increase exposure to flood-related health risks include the disruption of water and/or sanitation systems, contact with contaminated water and/or food, and the displacement of populations. Therefore, based on this, the assumption is made that the number of affected people (deaths, injuries, homelessness) provides the most indication for a possible connection between floods and communicable disease outbreaks. Furthermore, the total reported

damage provides an indication of possible disruptions to water and sanitation systems and infrastructure. Furthermore, damage might relate to loss of health infrastructure including essential drugs and supplies and complicating evacuation of patients potentially leading to an increased risk of diseases. [Du et al., 2010] Also, an increased level of damage might indicate a longer recovery process, complicating managing the increased need of health after a flood event and increasing secondary stressors such as economic stressors, for example loss of property. [Stephenson et al., 2014, Lock et al., 2012] Lastly, the time span of a flood might also correlate with reported floods, as a continuous flood might create favorable conditions for the proliferation of vector populations and could impact the commencement of the recovery process. [Morin et al., 2013] Furthermore, a long time span increases the exposure duration of individuals to contaminated water. [Akanda et al., 2009] Therefore, the retrieved data was filtered based on these different criteria, and a significance score was assigned to each flood. For several data rows, information on one or more criteria were missing. Therefore, this information was manually added, with the assumption that no data for total affected or total damage meant no affected individuals or damage and was thus set to zero. The data for total affected appeared to be the most complete and was therefore considered the principal variable. To assign this score, first, the data for total affected, total damage, and days of floods were standardized. For this, formula 1 is used. For each country, floods were selected that yielded a positive significant score, indicating that the summarized data is above the mean of the dataset, suggesting an above-average impact.

$$\text{Significance score} = (0.8 \times \text{total affected}) + (0.1 \times \text{total damage}) + (0.1 \times \text{days of floods}) \quad (1)$$

4.2.3 Epidemiological data

The initial scope of the data collection process for epidemiological was determined by the initial set of countries, the health impact scope determined in chapter 3, including: Cholera, Leptospirosis, Diarrhoeal disease, Malaria and Dengue and the time span of past flood events. The aim was to collect as detailed as possible data, consisting of weekly or monthly morbidity and mortality rates. The availability of reliable data was limited and therefore also yearly data was included in the database. Mortality rates were not sufficiently available in online open sources and were therefore excluded from the data collection. Epidemiological data was reported when at least one flood period could be compared to two non flooded periods, determined by the collected data on past floods. Based on the availability of data, countries were categorized in three categories; weekly data availability, monthly data availability and yearly data availability, presented in Table 7. Based on the availability of this data, the methodology for analysis is adjusted accordingly. As a result of this process, several countries were excluded from the research including: Vietnam, Iraq, Laos, Congo, Somalia, South Sudan, Guyana, United States of America, Guatemala, Russia, Czech Republic and Cambodia.

4.2.4 Health care system indicators

The initial set of indicators was obtained during the development of the conceptual model and laid foundation for the data collection. The data collection was systematically divided into the three main domains as defined in the conceptual model; vulnerability, people's susceptibility and resilience. During the data collection, proxy indicators were added to represent indirect indicators defined in the previous phase. A complete list of these indicators can be found in Table C.3 of the appendix. Data was collected for these countries that allowed for the collection of epidemiological data during flood and non flood periods, described in Table 7.

Table 7: Availability of epidemiological data divided into weekly, monthly and yearly available data including the time period of reported data

Weekly data	Disease	Time period
Pakistan	Cholera, ADD, Dengue, Malaria	2021 - present
Philippines	Leptospirosis, Dengue	2019, 2021, 2022
Sri Lanka	Leptospirosis, Dengue	2007 - present
Peru	Dengue	2020 - present

Monthly data	Disease	Time period
Bangladesh	Cholera, ADD	2018 - 2019
Australia	Leptospirosis, Dengue, Malaria	1991 - present

Yearly Data	Disease	Time period
China	Malaria	2000 - 2015
India	Cholera	2000 - 2022
Tanzania	Malaria, Cholera	2000 - 2022
Nigeria	Malaria, Cholera	2000 - 2022
Mozambique	Malaria, Cholera	2000 - 2022
United Kingdom	Cholera, Leptospirosis, Dengue	2007 - 2019

Chapter 5

Analysis of Flood Impact on Communicable Diseases

5 Analysis of Flood Impact on Communicable Diseases

The database developed in the previous chapter is used to perform the EDA to analyse the correlations between health system characteristics and incidence of communicable diseases after floods. To address this question, the analysis is distinguished into two different phases. The first phase involves examining the impact of floods on reported cases of communicable diseases to determine countries' resilience to the impacts of floods on the incidences of communicable diseases. Based on these findings, the performance of the healthcare systems of these countries will be analyzed in the second phase to identify vulnerability and resilience indicators that explain the varying outcomes in health effects of communicable diseases after floods.

5.1 Methodology

5.1.1 Exploratory Data Analysis

An EDA was selected as the preferred method because it allows for the examination of multiple variables and outcomes across data strata. It is a method that explores data to discover patterns, locate outliers, and identify relationships between outliers to test hypotheses. [Matthieu Komorowski et al., 2016] Typically, it is carried out as an initial and preliminary step before formal analysis or modeling to solve a research problem. For this reason, the term exploratory analysis is used. [Luijken et al., 2022] Most EDA techniques are graphical in nature, as their main role is to explore and gain insights into the structure of the data. A distinction can be made between univariate, which concerns only one variable or outcome, and multivariate, which involves multiple variables. A combination of both will be used in this research. [Matthieu Komorowski et al., 2016] EDA serves various goals, of which the following are relevant for this research:

1. Gain insight into the data and understand data structure
2. Visualize potential relationships between variables
3. Detect outliers

To present the process and results of this analysis, the reporting suggestions for EDA from [Luijken et al., 2022] are being considered. To illustrate the flow and objectives of the analysis, a study protocol is presented in Figure 9. The results of the analysis will be accompanied by a proposed research agenda. Other considerations include that conclusions should not be based on significance values only, interpretation of findings should be in line with the nature of analysis and all summarised results should be reported.

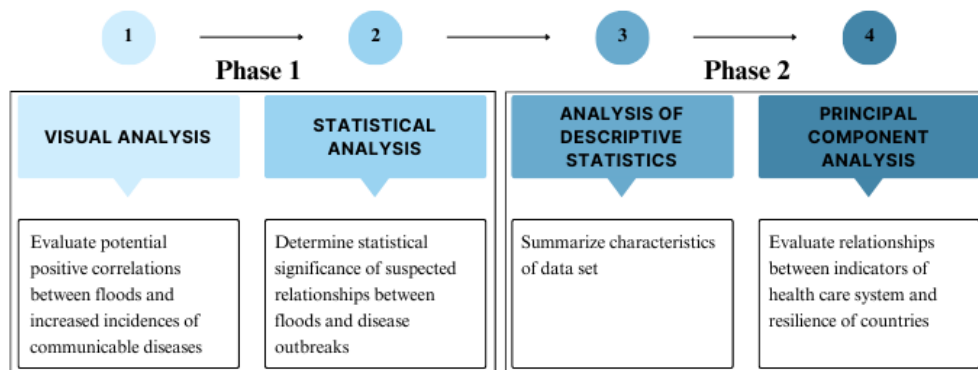


Figure 9: Study protocol of the exploratory data analysis

5.1.2 Visual Exploratory Data Analysis

Being able to visualise changes in incidence rates and seasonal patterns throughout the year is an effective approach to better understand data and take some first steps in exploring the data and identifying patterns and relationship. This section will discuss different time-series data by using some visual exploratory data analysis (V-EDA) methods. It is aimed to evaluate if the data can show whether the incidence of communicable diseases is suspected to be impacted by floods. This is done by first performing a visual inspection by plotting the data. By doing this, trends over the weeks, months of years can be visually inspected and provides a initial idea whether noticeable differences in epidemiological data during flood and non-flood periods can be suspected. The choice of type of plot is based on the availability of the data. Below, the different types used for this V-EDA are briefly elaborated on [Matthieu Komorowski et al., 2016]:

1. *Time plots* are used to visualise weekly and monthly data and illustrate data points at successive intervals of time. This allows for identifying recurring seasonal patterns. In these plots, the significant floods, as determined by the formula provided in subsection 4.2.2, are marked including their origin and start date.
2. *Time plot + histogram* are used to visualise yearly data against the total number of affected people in the corresponding year. Although, the yearly data is more difficult to interpret as the lack of details gives more room to other factors that might influence the number of reported cases, it still hints at a potential relationship between flooding and disease outbreaks.
3. *Scatter plots* of number of reported cases versus total number of affected people. These are used to analyse whether there might be a relationship between an increase in total number of affected people per year and the reported incidence rates.

5.1.3 Statistical Analysis

To further analyse the suspected significance in differences, statistical tests are performed to determine whether the data is significant. The nature of the data determines the use of a specific test. The formulas of these test can be found in appendix D.

Paired t-test

Monthly or weekly available data is tested through a two-tailed paired T-test where the mean difference between two paired groups is tested to determine the statistical significance. [Niven et al., 2012] A paired t-test is used when you have two sets of paired observations or measurements, for example measuring the same set of objective before and after an event or intervention. Two tailed refers to the directionality of the hypothesis being tested. It essentially checks if the two groups being compared are significantly different from each other, regardless of the direction of the difference. Here, the mean of incidence rates during flood periods are compared with the incidence rates in these same weeks/months of years without floods. This method allows for better control for individual differences between observations, such as seasonal variation or differences in the environment that could affect the number of reported incidences [Heo et al., 2008].

Pearson correlation

For data available on a yearly base, the Pearson correlation is performed. The reason for selecting this method is that it is able to test correlation despite non normality of the dataset. This test measures an association between two variables, primarily a linear relationship between two continuous variables. [Pearson, 1931, Schober and Schwarte, 2018] Here, the disease incidence rate was tested against the

total number of people affected by floods in a specific year. This is done to test the proposed null hypothesis which is; there is no correlation between the number of people affected by a floods in country X and the number of incidences of disease X. The significance of the reported data can again be determined by comparing either the Pearson Correlation Coefficient (r) to the critical value or by comparing the p-value with the selected significance level.

Regression analysis

Additionally, a regression analysis is performed to analyse the relationship between the number of affected people and number of reported incidences of a disease. A regression analysis is typically used to study how a response variable depends on one or more indicative variables. [Cook, 2015] The number of affected people is the independent, indicative variable because it is used to predict or explain variation in the number of reported cases, the dependent variable. With this analysis, it is possible to determine if the number of people affected has a statistically significant effect on the number of reported cases.

5.2 Results Visual Analysis

Each dataset of epidemiological data was plotted to visually inspect patterns and trends in the data. Although an increase in the number of incidences is complex and affected by multiple factors, it can provide a suspicion of significance that can be tested using additional statistical methods.

5.2.1 Results Visual Analysis - yearly data

For the datasets that provided yearly reported incidence numbers, the absolute number of cases was plotted against the people affected by floods in the corresponding years. To calculate the total number of affected people, all floods included in the EM-DAT were considered for the analysis. Pakistan was the only country that showed a possible relationship between the number of affected people by floods and the communicable diseases incidence rates. Figure 10 shows the number of people affected by floods versus the number of reported Malaria cases in Pakistan for the period 2000 to 2021. There are two indications to suspect a possible relationship between these two variables. There is a peak in reported cases after 2010, a year with a relatively large number of people affected by floods. In 2023, there is another peak in both incidences and number of affected people which can possibly be explained by the floods that occurred during these year.

In contrast, the plot of malaria incidence versus number of affected people in Tanzania, illustrated in Figures 10 and 11 shows a relatively constant number of Malaria cases since 2009 after which several peaks in number of people affected by floods can be identified. There is no suspected relationship between the two variables. The results of the visual analysis of the yearly plots are summarised in Table 8 and 9. The plots of all countries included in this study are presented in appendix A and B.

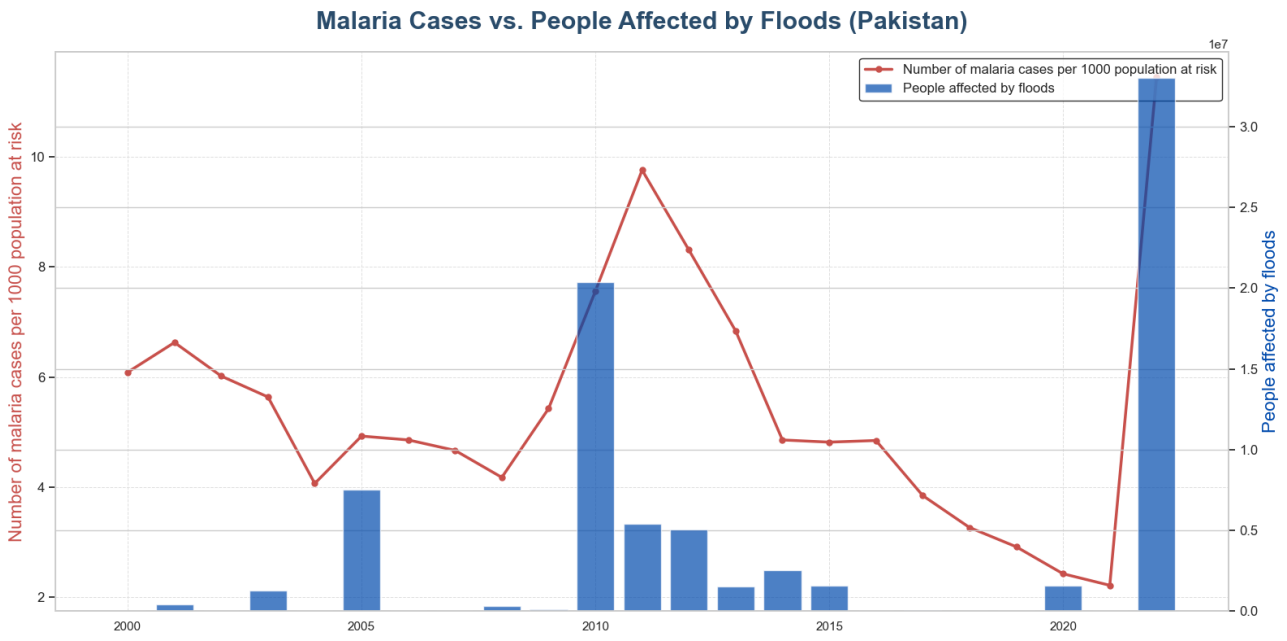


Figure 10: People affected by floods vs. Malaria cases in Pakistan (suspected significance)

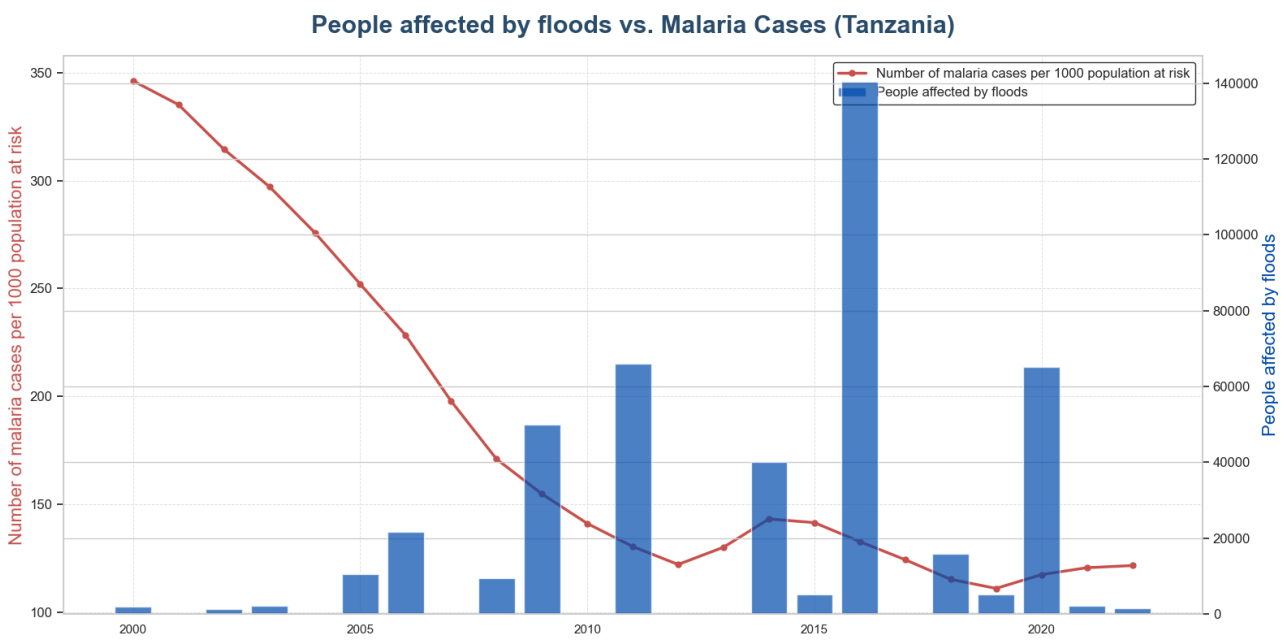


Figure 11: People affected by floods vs. Malaria cases per year in Tanzania (no suspected significance)

Table 8: Summary of results visual analysis yearly data - Malaria

Country	Findings	Suspected significance	Plot
China	<ul style="list-style-type: none"> - Malaria cases have decreased over time - Malaria cases are relatively low - No identified relationship between affected people and incidences 	No	E.6f
Dominican Republic	<ul style="list-style-type: none"> - Outlier in number of affected people in 2016 - Outlier does not correlate with malaria cases - No identified relationship between affected people and incidences 	No	E.6e
Mozambique	<ul style="list-style-type: none"> - Malaria cases are relatively stable over time - In the early 00's, higher number of people affected by floods - Decrease in malaria cases corresponds with lower numbers of affected people - Last years show little variance in reported malaria case, despite variance between affected people - Null hypothesis cannot be rejected 	+/-	E.6a
Nigeria	<ul style="list-style-type: none"> - Several outliers in number of affected people (2010, 2012, 2018 and 2022) - Malaria cases do not show correlation with these outliers 	No	E.6b
Pakistan	<ul style="list-style-type: none"> - Outliers in number of affected people (2005, 2010, 2022) - Outliers show possible correlation with malaria cases 	Yes	10
South Sudan	<ul style="list-style-type: none"> - Malaria cases are relatively stable over time - Number of affected people have increased over time and show high numbers from 2019 - Malaria cases do not increase from 2019 	No	E.6g
Tanzania	<ul style="list-style-type: none"> - Malaria cases have decreased over time - Number of affected people have increased over time - Plot does not show correlation between the two variables 	No	11

Table 9: Summary of results visual analysis yearly data - Cholera

Country	Findings	Suspected significance	Plot
Mozambique	<ul style="list-style-type: none"> - Number of affected people have decreased over time - Cholera cases are heavily fluctuating - Although cholera cases are also decreasing over time, the years showing peaks in incidences do not show a relationship with the peaks in number of people affected by floods 	+/-	E.5a
Nigeria	<ul style="list-style-type: none"> - Several outliers in number of affected people (2010, 2012, 2018 and 2022) - Cholera cases are heavily fluctuating - Outliers of cholera cases do not correspond with outliers in number of affected people 	No	E.5b
Tanzania	<ul style="list-style-type: none"> - Cholera cases seems to have decreased over time - Missing data for 2014 - Outliers of cholera cases do not show relationship with outliers in number of affected people 	No	E.5c

5.2.2 Example plots weekly and monthly data

Weekly incidence datasets were plotted over multiple years, enabling the observation of seasonal patterns, such as those related to rain seasons or temperature variations, as well as differences between years with and without flood disasters. Significant floods were highlighted in these plots to facilitate the identification of potential associations between flood events and increases in reported incidences. The dark blue X marks the start of the indicated flood. The yellow dashed lines indicate persistent flooding. The plot of two countries, Sri Lanka and Peru, raised suspicion for a possible relationship between a significant flood and a following outbreak. Figure 12 shows the dengue incidences for four different years (2010, 2013, 2017 and 2018) in Sri Lanka. In both 2010 and 2017, flooding occurred. Years 2010, 2013 and 2018 show a relatively stable number of incidences throughout the year. However in 2017, a significant outbreak of Dengue occurred in the 10 weeks following the flood in May 2017 raising suspicion for a possible relationship between the two variables. The Dengue incidences of five different years were also reported for Australia, as shown in plot E.7b. Floods occurred in three of those years (2010, 2011 and 2022). The corresponding reported number of Dengue cases in the month of these floods do not show significant increases that might raise a suspicion of possible connection between the floods and number of reported incidences. A summary of the main findings are provided in Tables 10 and 11. The visualisations are shown in appendix D and C.

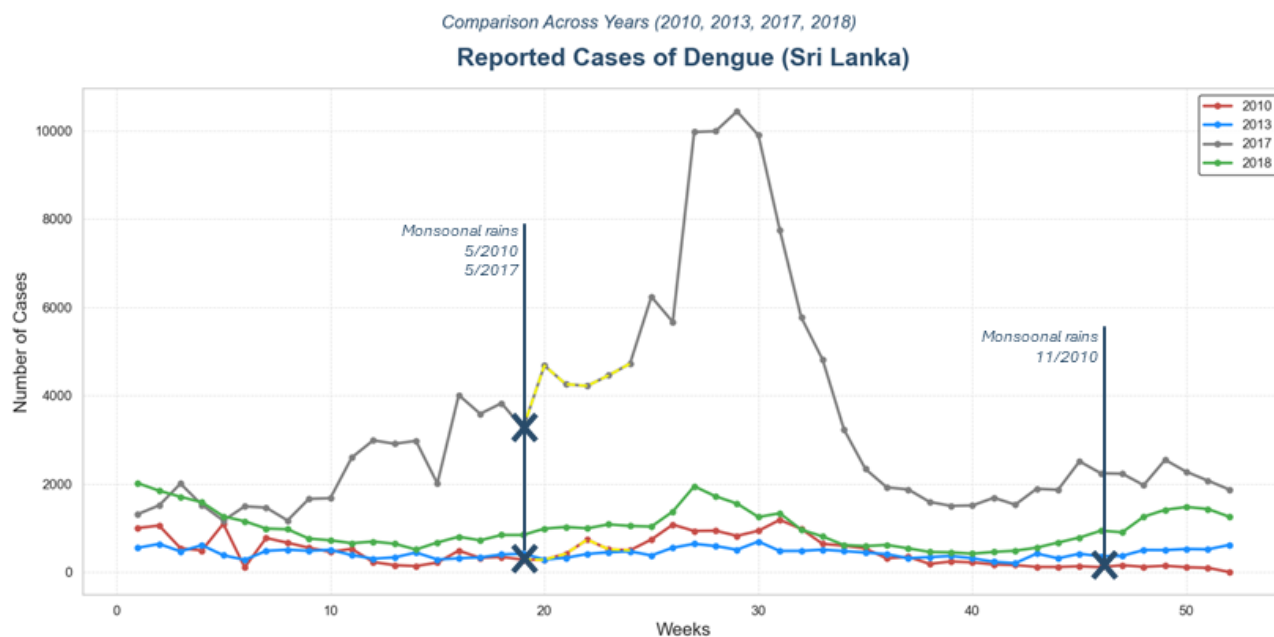


Figure 12: Dengue incidence and flood events in 2010, 2013, 2017 and 2018 in Sri Lanka (suspected significance)

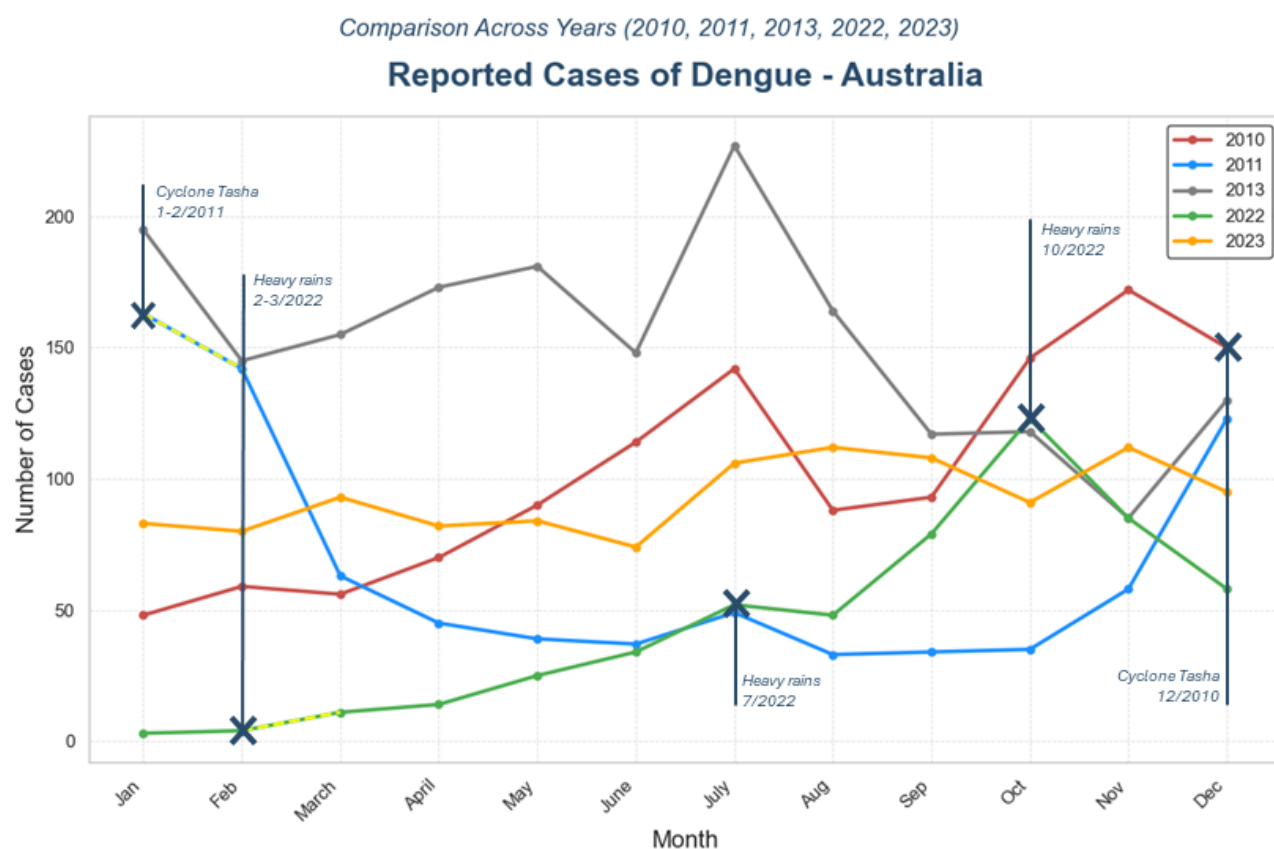


Figure 13: Dengue incidence and flood events in 2010, 2011, 2013, 2022, 2023 in Australia (without suspected significance)

Table 10: Summary of results visual analysis monthly and weekly data - Dengue

Country	Findings	Suspected significance	Plot
Australia	- Increased average incidence rates during winter with peaks in July - Peaks in reported dengue cases do not consistently follow floods	No	E.7b
Peru	- Significant peak in reported cases during weeks 15-24 following floods in beginning of 2023 - No clear seasonal variation - 2021 and 2022 show relatively stable number of reported cases	Yes	E.7a
Philippines	- Relatively stable number of reported cases in first 4 months of the year - 2/3 years of reported data show increase in reported cases after month 5 - No significant variance can be identified between flooded vs. non-flooded years	No	E.7d
Sri Lanka	- Significant peak in reported number of dengue cases in 10 weeks following monsoonal rains in 2017 - 2010, 2013, 2018 show relatively stable number of reported cases - Suspected seasonal variation during weeks 25-31 and week 34-52	Yes	E.7c

Table 11: Summary of results visual analysis yearly data - waterborne

Country	Disease	Findings	Suspected significance	Plot
Australia	Leptospirosis	- Reported cases seem to have a seasonal variation with increases reported cases from April to June - Two months after 2022, feb flood the reported cases show a peak - Reported cases significantly increase during cyclone Tasha but rapidly decrease after	+/-	E.8a
Bangladesh	ADD	- Reported cases of ADD are relatively stable throughout the year - A decline in reported cases can be identified after flood in 2014	No	E.8d
Philippines	Leptospirosis	- 2019 and 2022 show an increase in reported cases after month 7 - Lacking data for these months for 2021 - No relationship can be identified between flood in 2021 and reported cases in following months	No	E.8c
Sri Lanka	Leptospirosis	- Reported cases are fluctuating throughout the year - Increase in reported cases in month after monsoonal rains in 2010, however, other years also show an increase during this month - Inconclusive results	No	E.8b

5.3 Results of Statistical Analysis

5.3.1 Correlation Analysis

For the countries with data available per year, a correlation analysis was performed to evaluate the possible relationship between the reported incidence number per year and the corresponding affected people in that same year. The results of these analyses are shown in Table 12. The Pearson correlation coefficient indicates whether there is a linear relationship between the two variables. 1 indicates a perfect positive linear relationship, -1 indicates a perfect negative linear relationship and 0 indicates no linear relationship. This coefficient can be compared with the critical value to determine the significance of the data. This critical value is determined by the degree of freedom and the selected level of significance. For this study, the level of significance is set to 0.05. A Pearson Correlation coefficient higher than the critical value indicates statistical significance of the data. This relationship can also be evaluated by calculating the p-value. A p-value lower than the level of significance enables to reject the null hypothesis. For Pakistan, both the correlation coefficient and the p-value indicate significance. The results of the correlation analysis for Mozambique showed significance based on the p-value. The other countries did not show any significance within the data, which corresponds with the findings from the visual analysis.

Table 12: Results of Pearson Correlation Analysis of yearly reported malaria incidences versus yearly number of people affected by floods

Country	Pearson Correlation coefficient	Critical Value	Significant (coefficient)	P-value	Significant (p-value)
Pakistan	0.697	0.537	yes	0.000221	yes
Nigeria	-0.222	0.549	no	0.308	no
Mozambique	0.415	0.549	no	0.0489	yes
Tanzania	-0.299	0.549	no	0.165	no
China	-0.324	0.549	no	0.0677	no
South Sudan	-0.324	0.549	no	0.132	no
Dominican Republic	-0.170	0.549	no	0.450	no

5.3.2 Paired T-Test

Paired T-tests were performed to evaluate the data that was available on a monthly or weekly base since this data allowed for comparing on a more detailed level. In this test, months or weeks when floods occurred are compared to these same months or weeks in years without flooding. In this test, the t-statistic is calculated and compared to the corresponding critical value along with a comparison between the calculated p-value and the significance level of 0.05. Since the lag time of communicable diseases, described in subsection 3.1.4, is difficult to measure precisely, it was decided to perform the paired T-test both for a flooded period + 1 month after the end date of the flood as reported in EM-DAT (Table 13), and a flooded period + 2 months (Table 14). Following these results, the data corresponding to the incidences of Dengue in the Philippines, Sri Lanka, Australia and Peru show a significance when comparing reported number of cases of Dengue in flooded and non-flood periods. Furthermore, Sri Lanka and Australia show significance in reported number of cases of Leptospirosis. The test showing a positive significant correlation between flooded and non-flooded years are highlighted in blue.

Table 13: Results of paired T-test for reported cases of Cholera, ADD, Leptospirosis and Dengue comparing flooded (+ 1 month) and non-flood periods in Bangladesh, Philippines, Sri Lanka, Australia and Peru.

Country	Disease	Time period	T-statistic	P-value	Significant?
Bangladesh	Cholera	7:8 2016 vs 2018	0.435	0.739	no
Bangladesh	Cholera	8:9 2017 vs 8:9 2018	2.500	0.240	no
Bangladesh	ADD	8:10 2014 vs 8:10 2013	1.076	0.394	no
Philippines	Leptospirosis	1:8 2019vs 2021	1.126	0.297	no
Philippines	Leptospirosis	1:8 2021vs 2022	0.097	0.925	no
Philippines	Dengue	1:8 2019 vs 2021	-5.693	0.001	yes
Philippines	Dengue	1:8 2021 vs 2022	-2.369	0.050	yes
Sri Lanka	Dengue	flooded vs non-flooded	7.104	0.000	yes
Sri Lanka	Dengue	45:51 2010 vs 2013	-5.616	0.001	yes
Sri Lanka	Dengue	45:51 2010 vs 2018	-6.652	0.001	yes
Sri Lanka	Dengue	19:28 2010 vs 2013	2.878	0.018	yes
Sri Lanka	Dengue	19:28 2010 vs 2018	-5.633	0.000	yes
Sri Lanka	Dengue	19:28 2017 vs 2013	7.313	0.000	yes
Sri Lanka	Dengue	19:28 2017 vs 2018	0.000	0.000	yes
Sri Lanka	Leptospirosis	flooded vs non-flooded	-1.113	0.271	no
Sri Lanka	Leptospirosis	45:51 2010 vs 2013	2.042	0.087	no
Sri Lanka	Leptospirosis	45:51 2010 vs 2018	7.950	0.000	yes
Sri Lanka	Leptospirosis	19:28 2010 vs 2013	-0.460	0.656	no
Sri Lanka	Leptospirosis	19:28 2010 vs 2018	4.288	0.002	yes
Sri Lanka	Leptospirosis	19:28 2017 vs 2013	-1.543	0.157	no
Sri Lanka	Leptospirosis	19:28 2017 vs 2018	0.375	0.716	no
Australia	Leptospirosis	1:3 2011 vs 2013	4.610	0.044	yes
Australia	Leptospirosis	1:3 2011 vs 2023	3.263	0.082	no
Australia	Leptospirosis	2:4 2022 vs 2013	3.333	0.079	no
Australia	Leptospirosis	2:4 2022 vs 2023	-0.113	0.921	no
Australia	Dengue	1:3 2011 vs 2013	-1.615	0.248	no
Australia	Dengue	1:3 2011 vs 2023	1.096	0.387	no
Australia	Dengue	2:4 2022 vs 2013	-26.582	0.001	yes
Australia	Dengue	2:4 2022 vs 2023	-18.577	0.003	yes
Peru	Dengue	5:20 2021 vs 2022	-11.088	0.000	yes
Peru	Dengue	1:20 2023 vs 2022	0.471	0.643	no

Table 14: Results of paired T-test for reported cases of Cholera, ADD, Leptospirosis and Dengue comparing flooded (+ 2 month) and non-flood periods in Bangladesh, Philippines, Sri Lanka, Australia and Peru.

Country	Disease	Time period	T-statistic	P-value	Significant?
Bangladesh	Cholera	7:9 2016 vs 2018	0.623	0.597	no
Bangladesh	Cholera	8:10 2017 vs 2018	1.353	0.309	no
Bangladesh	ADD	8:11 2014 vs 2013	1.338	0.273	no
Philippines	Leptospirosis	1:12 2019vs 2021	0.959	0.358	no
Philippines	Leptospirosis	1:12 2021vs 2022	0.055	0.957	no
Philippines	Dengue	1:8 2019 vs 2021	-3.379	0.006	yes
Philippines	Dengue	1:8 2021 vs 2022	-2.882	0.015	yes
Sri Lanka	Dengue	19:32 2010 vs 2013	4.215	0.001	yes
Sri Lanka	Dengue	19:32 2010 vs 2018	-5.567	0.000	yes
Sri Lanka	Dengue	19:32 2017 vs 2013	9.121	0.000	yes
Sri Lanka	Dengue	19:32 2017 vs 2018	8.771	0.000	yes
Sri Lanka	Leptospirosis	19:32 2010 vs 2013	-0.333	0.745	no
Sri Lanka	Leptospirosis	19:32 2010 vs 2018	1.232	0.240	no
Sri Lanka	Leptospirosis	19:32 2017 vs 2013	-1.051	0.313	no
Sri Lanka	Leptospirosis	19:32 2017 vs 2018	-0.156	0.878	no
Australia	Leptospirosis	1:4 2011 vs 2013	3.346	0.044	yes
Australia	Leptospirosis	1:4 2011 vs 2023	2.937	0.061	no
Australia	Leptospirosis	2:5 2022 vs 2013	3.191	0.050	yes
Australia	Leptospirosis	2:5 2022 vs 2023	0.887	0.440	no
Australia	Dengue	1:4 2011 vs 2013	-2.251	0.110	no
Australia	Dengue	1:4 2011 vs 2023	0.616	0.581	no
Australia	Dengue	2:5 2022 vs 2013	-33.968	0.000	yes
Australia	Dengue	2:5 2022 vs 2023	-14.280	0.001	yes
Peru	Dengue	5:24 2021 vs 2022	-11.460	0.000	yes
Peru	Dengue	1:14 2023 vs 2022	1.891	0.071	no
Australia	Dengue	1:3 2011 vs 2013	-1.615	0.248	no
Australia	Dengue	1:3 2011 vs 2023	1.096	0.387	no
Australia	Dengue	2:4 2022 vs 2013	-26.582	0.001	yes
Australia	Dengue	2:4 2022 vs 2023	-18.577	0.003	yes
Peru	Dengue	5:20 2021 vs 2022	-11.088	0.000	yes
Peru	Dengue	1:20 2023 vs 2022	0.471	0.643	no

5.3.3 Regression Analysis and scatter plots

For several countries, a regression analysis was performed based on annual reported cases and the number of people affected by floods. The meaning of the variables can be explained as follows:

- R-squared is a statistical measure of how well the regression model fits the observed data and represents the proportion of the variance in the number of reported cases that is explained by the number of people affected by floods in the model. Typically, higher values might indicate that the model explains more of the variance in the number of reported cases.
- The coefficient associated with the affected people variable represents the estimated change in the number reported cases for an one-unit increase in the affected people variable.
- The p-value associated with each coefficient in the regression output indicates the probability of observing the coefficient's value (or one more extreme) under the null hypothesis that there is no relationship between the two variables. P-values close to zero provides reason to reject this null hypothesis.

Table 15: Results of regression analysis, disease incidence vs. people affected by floods per year

Country	Disease	R squared	Coefficient (affected people)	p-value (affected people)	Significant?
Nigeria	Malaria	0.049	-7.16E-06	0.308	no
Nigeria	Cholera	0	0.0002	0.957	no
Pakistan	Malaria	0.486	1.99E-07	0	yes
Mozambique	Malaria	0.172	2.70E-05	0.049	yes
Mozambique	Cholera	0.09	0.0023	0.164	no
Tanzania	Malaria	0.09	-0.0007	0.165	no
Tanzania	Cholera	0.06	0.0329	0.271	no
China	Malaria	0.15	1.84E-10	0.068	no
South Sudan	Malaria	0.105	-2.43E-05	0.132	no
Dominican Republic	Malaria	0.029	-9.00E-08	0.45	no

A summary of the results is provided in Table 15. These results indicate a statistical significance for both Mozambique and Pakistan. To further analyse these findings, the results of the regression analyses have been visualised with scatter plots to get a visual indication of how well the data fit the model. Two examples of these plots are provided in Figures F.9a and F.9b. Based on these plots, it can be concluded that no linear relationship can be identified. Therefore, the decision was made to continue the analysis by analyzing the scatter plots only. The scatter plots are shown in Figures F.10 and F.11 of appendix F.

5.4 Discussion

5.4.1 Visual Analysis

Yearly data

The visual analysis of yearly data in this study covers people affected by floods and all reported cases of diseases within a given country over the course of an entire year. The results provide several insights. First, in general, the incidence rates of Malaria have decreased over time, while Cholera does not seem to follow this trend. Notably, when comparing waterborne and vector-borne diseases, vector-borne diseases demonstrate greater stability over time, whereas instances of cholera show pronounced spikes. This difference calls for further examination of the underlying reasons for such variations. Another notable point is that the prevalence of Malaria is significantly higher in African countries than in other parts of the world. Lastly, while this approach allows for the identification of significant floods and potential disease outbreaks, its effectiveness is limited by the inability to distinguish minor outbreaks due to the scale of the data. Moreover, the reliability of reported cases is questionable, raising concerns about the quality of reporting. Conversely, the apparent significance in the data may be influenced by various factors beyond flood events, as explained in the conceptual model. Additionally, the absence of noticeable patterns in yearly data across most countries encourages investigation into other potential contributing factors, such as war, drought, extreme weather conditions, technological advancements in reporting, or other natural disasters.

Weekly and Monthly data

Enhancing the level of detail in the analysis provides a clearer insight into potential connections between flood events and disease outbreaks. Specific patterns linking floods to the incidence of Dengue have been observed in Sri Lanka and Peru, although these connections are less distinct in other countries. A central question arises regarding the timing of disease outbreaks, especially con-

cerning vector-borne diseases which often proliferate in stagnant water following floods. However, determining the duration of elevated water levels and its precise impact on mosquito reproduction is challenging. To address this uncertainty, the statistical analysis considers both one and two-month intervals, aiming to partially mitigate this uncertainty. Nonetheless, variations in the duration of flooding, influenced by factors such as the specific location of the flood, urban versus rural settings, and land usage patterns, further complicate the analysis.

Seasonal patterns play a crucial role in understanding fluctuations in incident rates, recurring annually. Identifying these patterns facilitates distinguishing between increases in incidences attributable to seasonal variations and those arising from other factors. In Sri Lanka, a notable seasonal pattern emerges, characterized by heightened reported cases during weeks 25 to 31 and 47 to 50, corresponding to mid-June to mid-July and the end of November to the beginning of December. This pattern aligns with the Southwest monsoon season, bringing rainfall to the southwest from May to September, and the northeastern monsoon, associated with higher rainfall from October to January, while temperatures remain relatively stable throughout the year. Similarly, in Australia, a seasonal trend is evident in reported cases of Leptospirosis, with peaks observed in late summer and autumn, from February to May.

5.4.2 Statistical Analysis

Correlation analysis

Pakistan exhibits convincing significance, demonstrating a positive correlation that aligns well with the results of visual analysis. In contrast, while Mozambique shows significance in terms of p-value, closer examination of the visual plot reveals discrepancies. In the early 2000s, the total number of people affected by floods was higher, which correlated with a somewhat increased number of Malaria cases per 1000 population at risk. However, upon closer examination of the data from 2010 onwards, during which reporting quality presumably improved, the incidence rate appears stable, with no evident patterns indicating a relationship between the total number of flood-affected individuals and the incidence rate of Malaria. Consequently, drawing any definitive conclusions becomes challenging. Notably, all other countries fail to exhibit any significant correlation. Moreover, the identified correlations tend to be negative, suggesting that as the number of people affected by floods increases, the incidence rate of malaria tends to decrease.

Paired T-test

The paired t-test specifically compares the period of flooding with the months following the flood. However, it is important to consider that the test does not account for the months prior to the flood or the relative change during these months. It examines the mean of the two sets of data being compared, assuming that the number of reported cases follows a seasonal pattern and thus a similar trend throughout the year. However, earlier discussions revealed that these seasonal patterns are not clear in the plotted data, making it difficult to explain the observed patterns solely through seasonal variability. Consequently, the number of reported cases in similar periods in different years might not exhibit a consistent trend, complicating comparisons. For example, Australia shows statistical significance when comparing 2011 and 2013. Although the mean of the flooded year is higher, the number of cases decreases during and after the flood, suggesting a negative effect of the flood on the incidence rate. Therefore, this cannot be considered a significant result.

The sign of the t-statistic is also crucial to consider. A positive t-statistic implies that the mean of the first dataset surpasses that of the second dataset, while a negative t-statistic suggests the opposite: the mean of the second dataset exceeds that of the first. In our results, the first dataset consistently represents the flooded period. Upon closer examination, it is apparent that significant tests often reveal a higher mean in the non-flooded period, indicating a significant trend. Essentially, besides Australia, the paired t-test primarily indicates statistical significance in increased reported cases during flooded periods for Sri Lanka (Dengue). When comparing the flooded period + 1 month and the flooded period + 2 months, only one notable difference arises: the statistical significance of Leptospirosis in Sri Lanka. However, upon closer analysis, this result, when compared with visual analysis, reveals fluctuating trends, suggesting potential data quality issues.

Regression analysis and scatter plots

Examining Table 15, it is clear that nearly half of the variance in the data points in Pakistan can be explained by the regression model. Furthermore, the p-value of 0 indicates that the regression model fits the data well. Similarly, the p-value for Mozambique also suggests statistical significance. However, with only 17% of the variance explained by the regression model in Mozambique, it is insufficient to draw a linear trend, especially given the small sample size of this study. This limitation is also apparent in Figure F.9a. Consequently, it was considered necessary to conduct further analysis by examining scatter plots to identify potential alternative trends.

One notable observation from the scatter plots is that a majority of the data points are concentrated on the left side, indicating that the number of people affected by floods each year is relatively equal, with only a few outliers. Regarding the scatter plots representing cholera cases, the outlier in Nigeria indicates a high number of affected people not corresponding to a high number of Cholera cases. Conversely, in Tanzania and Mozambique, the outlier with a high number of affected people also shows a high number of Cholera cases in the corresponding year. In the Malaria plots, Pakistan displays a somewhat positive linear relationship. However, Nigeria, Tanzania, and South Sudan contradict this relationship, as data points with a high number of people affected by floods do not correspond to data points with a high number of Malaria cases, suggesting no clear relationship. The scatter plot for China appears more random compared to others, possibly due to the country's large population size and relatively low number of Malaria cases, which are dispersed across the entire country. Across all scatter plots, there are instances of high disease incidence despite a low number of people affected by floods, indicating that increased incidences do not consistently correlate with floods.

5.5 Conclusion

In conclusion, the combined visual and statistical analyses offer insights into the complex patterns and relationships between flood disasters and subsequent disease reporting. While clear patterns were observed in yearly plots, particularly in Pakistan, other plots resulted in inconclusive findings due to challenges in identifying minor outbreaks, concerns about reporting reliability, and the influence of factors beyond floods on data significance. Vector-borne diseases showed greater stability over time compared to cholera spikes. Weekly and monthly plots allowed for detailed analyses, revealing detailed connections. However, determining precise flood durations remains a challenge, impacting result interpretation. While seasonal patterns were identified in Sri Lanka and Australia, statistical analysis highlighted a positive correlation in Pakistan and initial suggestions of significance for Mozambique, although recent years raise doubts. The paired t-test emphasized significance in countries identified visually, including Australia for *Leptospirosis* during floods. However, evidence against the null hypothesis is limited considering visual analysis and relative change. Scatter plots revealed no consistent linear relationship between affected people and reported cases, with instances of high correlations (e.g., Malaria in Pakistan, Cholera in Mozambique and Tanzania) and divergent cases (e.g., Malaria in the Dominican Republic, South Sudan, Tanzania, Nigeria, and Cholera in Nigeria).

The results of this initial phase of the EDA suggest that factors beyond floods may influence the relationship with elevated reported incidences. It is assumed that countries showing a positive correlation between flood-affected populations and increased incidence rates may be less resilient to communicable diseases in the context of flooding compared to those lacking this correlation. The findings regarding the patterns and relationship between floods and the incidence of communicable diseases and the distinction of countries based on their impact level are summarized in Table 16, where green indicates no evidence for the impact of floods on communicable diseases and red indicates evidence for this impact. It is proposed to introduce two categories: impacted and not-impacted. To be categorized as impacted, a country should have shown significance in both the visual and the statistical analysis. Following [Luijken et al., 2022], no conclusions will be drawn solely based on significance values. The thresholds for supporting significance indicating impact from floods are as follows:

1. For the visual analysis, this is the case when a correlation between a high number of affected people by floods or historic flood events showed a correlation with an increased number of communicable disease incidence.
2. For the statistical analysis, statistical significance should have been identified, and the context of this result, including relative change in incidences and consideration of lag time, should support this statistical value.

The impacted countries include Pakistan, Peru, and Sri Lanka. As the context of the statistical significance for Australia did not support the findings, Australia is not considered to be impacted by the floods. Mozambique only shows a statistical significance and thus no conclusions with regard to the impact level will be drawn. As the data availability for the impacted countries solely concerns vector-borne diseases, it has been decided to exclude waterborne diseases from further analysis. This decision is rooted in the understanding that no meaningful comparisons can be made among countries regarding waterborne diseases, as none are considered more impacted than others in this context.

Remarkably, the results of this analysis indicate that countries demonstrating characteristics that one could link to higher risks, such as high incidence rates for vector-borne diseases and low incomes,

like the African countries, are not necessarily more susceptible to the impacts of floods on incidence rates. This raises questions about the factors that influence susceptibility and how countries with high and low impacts might be distinguished.

Table 16: Results of the visual and statistical analyses indicating the resilience of countries to communicable diseases following flood disasters.

(a) Countries with yearly available data

Country	Visual Analysis	Statistical Analysis	Impact Level
China	No impact	No impact	Not impacted
Dominican Republic	No impact	No impact	Not impacted
Mozambique	No impact	Mixed	Inconclusive
Nigeria	No impact	No impact	Not impacted
Pakistan	Impact	Impact	Impacted
Tanzania	No impact	No impact	Not impacted
South Sudan	No impact	No impact	Not impacted
United Kingdom	No impact	No impact	Not impacted

(b) Countries with weekly or monthly available data

Country	Visual Analysis	Statistical Analysis	Impact Level
Bangladesh	No impact	No impact	Not impacted
Philippines	No impact	No impact	Not impacted
Australia	No impact	Mixed	Inconclusive
Peru	Impact	Impact	Impacted
Sri Lanka	Impact	Impact	Impacted

Legend:

- No impact - No significant impact of floods on communicable diseases.
- Impact - Significant impact of floods on communicable diseases.
- Mixed - Inconclusive or mixed results.

Chapter 6

Analysis of Healthcare System Resilience

6 Analysis of Healthcare System Resilience

In the previous chapter, the impact level of floods on the incidence of communicable diseases was explored. Building on these findings, this chapter aims to evaluate the relationships between healthcare system performance indicators and the post-flood increase in communicable diseases. The objective is to identify which factors contribute to variations in resilience to flood-related health impacts across different countries. To achieve this, Principal Component Analysis (PCA) will be employed as part of the EDA to reduce the dimensionality of the dataset and clarify the variances within the healthcare systems. This chapter will begin with a detailed outline of the methodology used, followed by a comprehensive presentation of the results.

6.1 Methodology Principal Component analysis

PCA is a method used for reducing the dimensionality of datasets and thus increasing interpretability, but at the same time aiming to minimize information loss. With this technique, new uncorrelated variables are created that successively maximize variance [Jolliffe and Cadima, 2016]. This thesis aims to assess vulnerabilities and resilience of health care system based on a large set of indicators and therefore, dimensionality reduction is deemed useful. This subchapter will explain the methodology of performing the PCA. The process suggested for this research consists of 11 consecutive steps, and are as follows:

1. Selection of indicators.
2. The collection of data
3. Pre-processing the data: transformation, normalization, and theoretical orientation.
4. Clustering countries based on exposure level
5. Verify the use of PCA
6. Create a correlation matrix and assess the collinearity
7. Remove redundant data
8. Perform PCA with standardized input values
9. Select no. of PCs to be used for further analysis
10. PCA matrix rotation
11. Interpretation of results

Selection and collection of indicator data

An initial set of indicators was selected based on conceptual, theoretical, and/or empirical justifications from previous research to represent each of the domains: exposure, vulnerability, and resilience. A comprehensive set of indicators should be collected to adequately describe the determinants of health in the context of flood-disease impact. In accordance with the proposed conceptual model and corresponding frameworks, indicators should be identified that illustrate the extent to which a country performs in each domain. The collection of data can utilize all available open data sources. It is important to consider both the timeline and completeness of the data.

Descriptive statistics can be used to summarize characteristics of datasets to get a better sense of the variables by examining the data. The selected descriptive statistics for this research includes the mean, standard deviations, min and max values. By evaluating the descriptive statistics, the dataset becomes more straightforward and initial trends and patterns might be identified. The data is suggested to be visualised using histogram plots.

Pre-processing the data

After collecting the raw variables, the data should undergo transformation, standardization, and correction based on theoretical orientation. Transformation involves converting absolute variables into percentages, averages, rates, or differences to account for variations in, for example, the total population of a country. Subsequently, the data should be standardized, which entails scaling all variables to ensure they have comparable reference points. The selected method for standardization is z-score normalization, which standardizes values based on the mean and standard deviation within the dataset. If a data point equals the mean of the variable, its standardized value becomes 0. Consequently, values below the mean become negative and those above become positive, with the magnitude determined by the standard deviation [Maharrani et al., 2024]. Lastly, the orientation of each variable should be adjusted so that larger values theoretically correspond to higher resilience in the context of communicable diseases after floods. [Cutter et al., 2014]

Clustering countries based on exposure level

Before performing the PCA, an additional step is proposed to facilitate comparison between countries that face similar flood-health risks. It is suggested to cluster the countries within the dataset based on their exposure level. The exposure level is defined as the sum of a country's exposure to vector-borne diseases and exposure to floods. For performing the clustering, k-means clustering is utilized. K-means clustering is an unsupervised machine learning algorithm used to partition a dataset into a predetermined number of clusters, grouping similar countries and minimizing within-cluster variation of exposure levels. The k-means clustering algorithm generates clusters using the mean value of objects within the cluster. [Ikotun et al., 2023]

Verify the use of PCA

Two tests are proposed to verify the suitability of data for PCA Analysis. First, Bartlett's Test is applied to compare the obtained correlation matrix to the identity matrix. This test determines whether the p-value is smaller than the significance level of 0.05, where the null hypothesis can be rejected. Basically, it tests multicollinearity and checks the redundancy between variables. However, not too much data should be removed as a perfectly uncorrelated dataset is unsuitable for PCA. Therefore, the Kaiser-Meyer Olkin (KMO) statistic statistically measures the data adequacy based on the level of correlation between indicators. This adequacy refers to the proportion of variance among variables. Values closer to 1 indicate higher correlations and thus better suitability. A value below 0.5 indicates the data is inappropriate for use in PCA. [Maharrani et al., 2024]

Removing redundant data

The number of indicators is being reduced based on the correlation matrix. Collinearity is assessed to determine whether variables might be redundant, for example, if multiple variables describe the same mechanisms. This phenomenon is called double counting and leads to outcomes that over represent some mechanisms, resulting in an uneven reflection of each mechanism's influence on the flood-disease resilience index. Therefore, it is decided to remove one of each pair of indicators with a Pearson's R larger than 0.85.

Perform PCA and select number of components

PCA values are calculated based on eigenvalues and eigenvectors, which represent the data distribution of a dataset. Through PCA, the original set of indicators is transformed into new variables known as principal components (PCs), where the number of PCs is fewer than the initial set of indicators. These PCs capture the essential indicators that explain the variances within the dataset.

Geometrically, PCs represent the directions of the data that explain the maximum variance. Eigenvectors are ranked in descending order based on their eigenvalue, indicating the amount of variance explained by the PCs. In this analysis, the Kaiser criterion was used to determine the number of components to retain, indicating that all eigenvectors with eigenvalues greater than 1 should be retained. [Jorgensen and Hansen, 2012]

PCA Matrix Rotation

To clarify the interpretation of indicators, the varimax rotation is being considered. Varimax rotation aims to rotate the factor axes in a way that maximizes the variance of the squared loadings within each factor, while also minimizing the number of variables that have high loadings on each factor. It is used to make the data orthogonal by increasing the interpretability [Török, 2018]. In this process, the sum of variances of the rotated components is equal to the unrotated components so no variance is lost. However, the successive maximization of the unrotated PCs, resulting in a more evenly distributed total variance between components after rotation. [Jolliffe and Cadima, 2016] Important to note is that this step is not required, so an unrotated solution could also be considered for interpretation.

Interpretation of results

Interpretation of results is primarily done by constructing a so-called resilience score based on the retained PCs. This score consists of the weighted sum of the component scores, calculated by multiplying the loading matrix with the normalized indicator values. As a result, each country is assigned a score for each retained PC, providing insight into the contributions of dominant variables to the variance between the countries in the dataset. By doing this, a high value is assigned to communities, that score high on an indicator, that also has high variance. This potentially leads to insights what indicators might need to improve to increase the resilience.

6.2 Results PCA

6.2.1 Selection and collection of indicators

This subsection presents an overview of all selected indicators. Based on the initial indicators (Table 5), derived from reviewed frameworks and theories, several sources were utilized to find data. When data were available for all countries, a criterion was directly included. However, more often, alternative indicators describing similar concepts were selected to cover all domain categories from the conceptual framework (Figure 7). These indicators are used to evaluate the health risks associated with communicable diseases after a flooding event. Therefore, each indicator aims to describe characteristics relevant to this context. To structure the selection of indicators, they are categorized into three main domains, encompassing exposure, vulnerability, and resilience, following the INFORM risk index model [European Commission and Joint Research Center, 2024]. Based on the review of existing frameworks in Chapter 3, more detailed categories are defined within each domain. This section provides an overview of these indicators, including a description of each indicator, its unit, and the data source. Furthermore, it describes the relevance and reason for the inclusion of the indicators.

Exposure

The exposure assessment is divided into three main categories: exposure to flooding, exposure to waterborne diseases, and exposure to vector-borne diseases and are used to assess the risk of potential flooding and a country's vulnerability to diseases. They provide insight into the likelihood of a country being affected by floods and its susceptibility to disease outbreaks.

Sections 3.1.6 and 3.1.7 outlined the mechanisms behind these exposures and distinguished between relevant indicators for waterborne diseases and vector-borne diseases. For instance, the transmission of Cholera is closely linked to the quality of water and sanitation systems, as well as the number of individuals sharing the same water source. In contrast, vector-borne diseases are predominantly influenced by factors such as the size of the vector population and environmental conditions. Urban population was added as an indicator since urbanization is recognized as a factor enhancing flood risk [Salami et al., 2017]. Furthermore, several authors mentioned age as a possible indicator. For this reason, 'Children under 5' was included as an indicator, as young children belong to the group most susceptible to vector-borne diseases [World Health Organization, 2014]. 'Elderly (65+)' was also included, as this group is more vulnerable to health risks. Figure 14 and Table 17 provide an overview of these indicators.

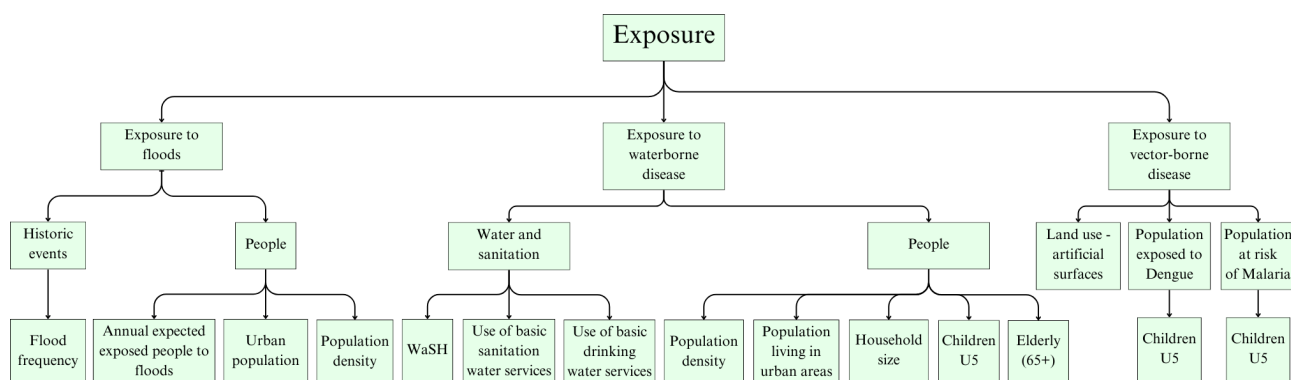


Figure 14: Indicators in the Exposure category - Exposure to floods, exposure to waterborne disease and exposure to vector borne

Table 17: Overview of indicators in the exposure category including unit and source

Category	Indicator	Unit	Source
Exposure to floods	Annual expected exposed people to floods	Index	INFORM Risk
Exposure to floods	Population density	People per sq. km of land	INFORM Risk
Exposure to floods	flood frequency	Average number of floods/year	EM-DAT
Exposure to floods/ Exposure to waterborne	Population living in urban areas	%	INFORM Risk
Exposure to waterborne	Quality of sanitation and drinking water	Score	WaSH Monitoring Programme
Exposure to waterborne	Housing overcrowding rate	%	Worldmapper
Exposure to waterborne	People using at least basic sanitation services	% of population	INFORM Risk
Exposure to waterborne	People using at least basic drinking water services	% of population	INFORM Risk
Exposure to waterborne	Household Size	Index	INFORM Risk
Exposure to vector-borne	Land use - artificial surfaces	% of total land	OECD
Exposure to vector-borne	Number of children <5	% of population	Our World in Data
Exposure to vector-borne	Population ages 65 and above	% of population	Our World in Data
Exposure to vector-borne	Population exposed to Dengue	Index	INFORM Risk
Exposure to vector-borne	Population at risk of Malaria	Index	INFORM Risk

Vulnerability

The second domain addressed in the evaluation of the health risks is vulnerability. Vulnerability is a complex term that can be explained as a combination of both physical and social vulnerability [Few, 2007]. To address these, a distinction was made between socio-economic vulnerability, describing social and economic factors of a population, health vulnerability, describing processes that affect the birth, growth and survival of individuals and behavioural vulnerability describing peoples susceptibility to change. Figure 15 and Table 18 provide an overview of these indicators.

To evaluate the economic position, the Multidimensional Poverty Index (MPI) was selected as an indicator, which measures acute multidimensional poverty and considers health, education, and living standards. The MPI is published annually by the Oxford Poverty & Human Development Initiative (OPHI) and the United Nations Development Programme (UNDP) [United Nations Development Programme, 2020]. To evaluate the financial position of a population, GDP per capita and mean income were selected. Additionally, the employment-to-population ratio was included as an alternative to the type of occupation ([Salami et al., 2017]), as it generally suggests greater economic stability and opportunities for individuals. The social position is described by demographic indicators and the Gender Inequality Index (GII). The GII, published by the UNDP [United Nations Development Programme, 2024], measures gender-based disadvantages based on three indicators: empowerment, reproductive health, and the labor market. A higher GII score indicates increased levels of inequality within a country, disadvantaging women. Health vulnerability is described by the prevalence of HIV, under-5 mortality, under-5 malnutrition, and the percentage of people affected by natural disasters in the past three years. Under-5 mortality reflects the environmental, economic, and social conditions in which children live, including the healthcare system. Because morbidity data is often unavailable, this indicator is frequently used to identify vulnerable populations [World Health Organization, 2024c]. These health vulnerability indicators were included to evaluate overall vulnerability, influenced by risk factors, environmental conditions, and health infrastructure, indirectly measuring critical health aspects. The percentage of people affected by natural disasters was included to evaluate whether communities might still be recovering from a recent disaster, potentially increasing their vulnerability to a new one. Lastly, some indicators were selected to describe behavioral vulnerability based on individuals' susceptibility to change. This is described by the literacy rate, educational attainment, internet usage and mobile cellular subscription.

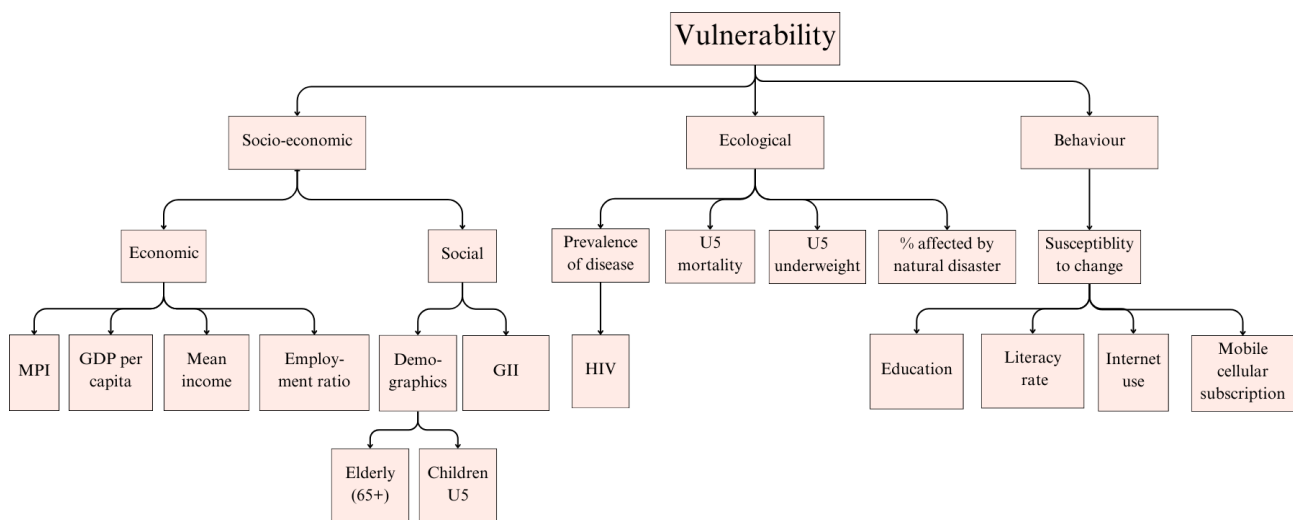


Figure 15: Indicators in the Vulnerability category - Socio-economic, Ecological, Behavioural

Table 18: Overview of indicators in the vulnerability category including unit and source

Category	Indicator	Country	Source
Socio-economic	Multidimensional Poverty Index	Index	UNDP
Socio-economic	GDP per capita	USD	INFORM Risk
Socio-economic	Mean Income	USD	CEIC
Socio-economic	Employment to population ratio, 15+	%	World Bank
Socio-economic	Gender-inequality index	Index	UNDP
Ecological	Prevalence of HIV	%	INFORM Risk
Ecological	Under 5 mortality	%	INFORM Risk
Ecological	Under 5 underweight	%	INFORM Risk
Ecological	% affected by natural disaster (last 3 years)	% of population	INFORM Risk
Behaviour	Individuals using the internet (% of population)	%	INFORM Risk
Behaviour	Educational attainment, at least completed primary, population 25+ years	%	World Bank
Behaviour	Literacy rate	%	World Population Review
Behaviour	Mobile cellular subscriptions	Per 100 people	World Bank

Coping capacity

The coping capacity of a country describes its ability to manage and respond to disasters, including those related to communicable diseases following flood events. Within this framework, coping capacity is evaluated across three main dimensions: healthcare system resilience, institutional resilience, and individual resilience. Healthcare system resilience assesses the capacity of the medical infrastructure to effectively respond to disasters and deliver essential care during crises. This aspect focuses on factors such as physical infrastructure, availability of medical personnel, and financial resources. The 'financial priority to health' and 'supply chain' were included as additional indicators to describe health care capacities. Additionally, the performance of the health care system is evaluated based on indicators including, 'Access to healthcare', 'Maternal mortality rates', 'life expectancy', and emergency response capabilities described by the indicators 'Detection and reporting', 'Rapid response' and 'Prevention', also indicating the availability of adequate warning systems [Salami et al., 2017, Khan et al., 2018].

Institutional resilience evaluates the effectiveness of government policies, planning, and overall governance in mitigating the impact of disasters. Key indicators in this domain include measures of government effectiveness, state legitimacy, order and security and constraints on government power. While indicators related to natural disaster preparedness would ideally be included, their availability was limited due to data constraints. 'State legitimacy' and 'Constraints on government power' were included to assess the level of trust in institutions and the prevalence of corruption, respectively, as these factors profoundly influence governance effectiveness and societal stability.

Lastly, the selection of indicators for assessing individual resilience includes factors that shape individuals' lives and their communities. 'Political rights' and the 'freedom of peaceful assembly' were chosen to reflect the degree of civic participation and the health of democratic processes, crucial for fostering citizen engagement and accountability within society [Salami et al., 2017, Few, 2007]. Additionally, indicators such as 'freedom over life choices', 'count on help', and the 'equal access index' were incorporated to capture dimensions of social cohesion [Ansari et al., 2003]. These indicators reflect individuals' autonomy, social support networks, and equitable access to opportunities and resources, all essential components of cohesive and resilient communities. Monitoring these indicators allows for a nuanced understanding of the inclusivity, fairness, and connectedness of communities, informing efforts to address inequality and promote positive social outcomes. Moreover, the role of positive psychosocial coping mechanisms in mitigating the adverse effects of exposure to

violence and crime emphasizes the importance of considering mental health and well-being in the selection of indicators. High levels of violence and crime not only pose direct physical risks but also contribute to mental health challenges, social disintegration, and barriers to accessing healthcare services. Therefore, indicators related to violence and crime were included as indicators of these psychosocial stressors, reducing one's individual resilience [Lock et al., 2012].

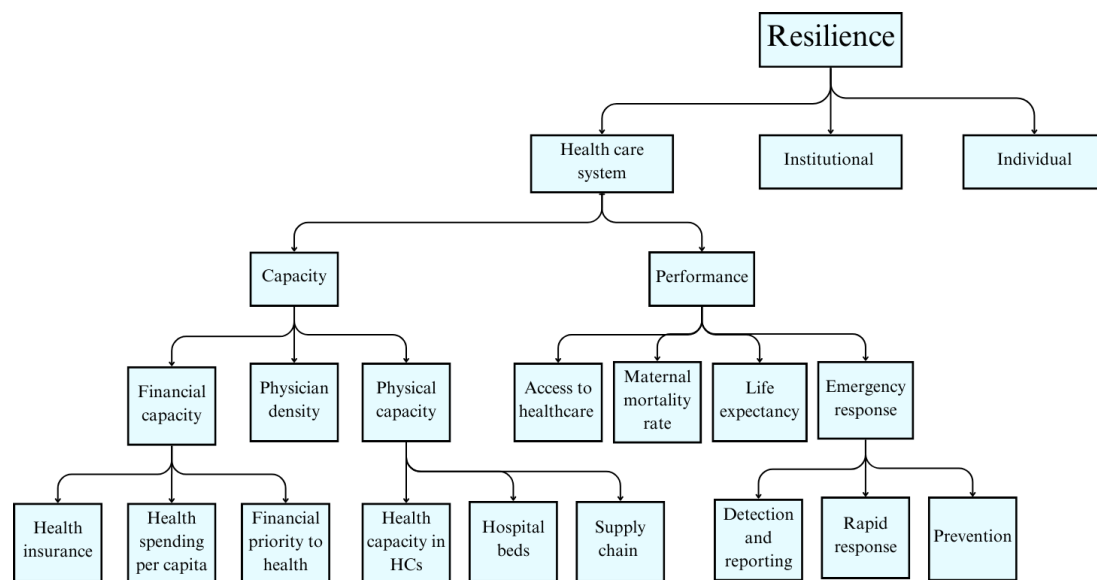


Figure 16: Indicators in the Coping Capacity category - Health Care System Resilience

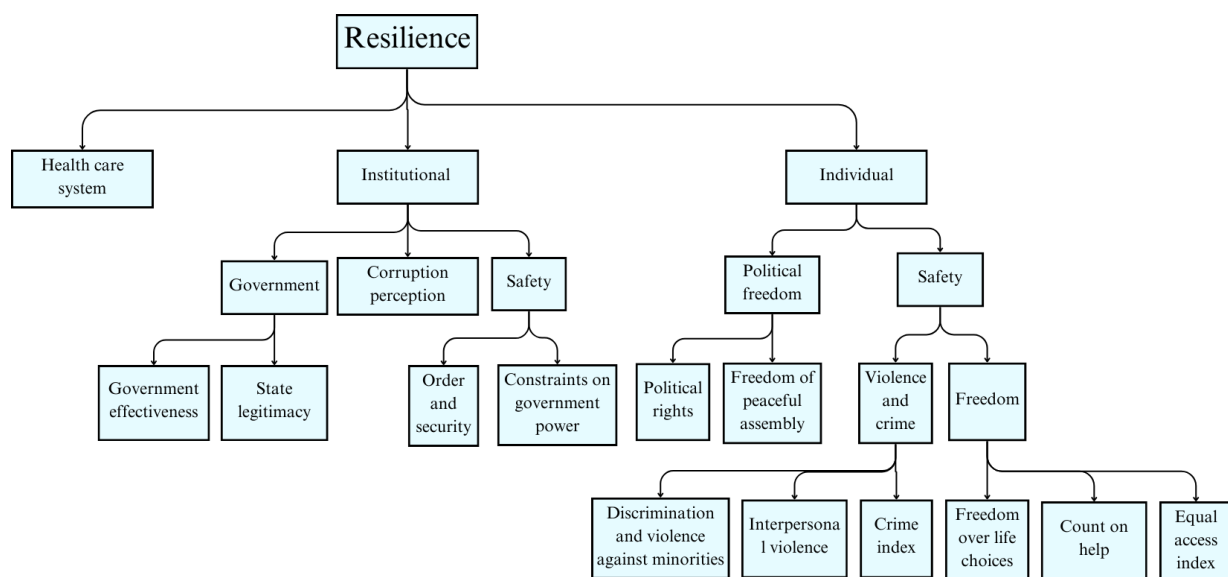


Figure 17: Indicators in the Coping Capacity category - Institutional and individual resilience

Table 19: Overview of indicators in the coping capacity category including unit and source

Category	Indicator	Country	Source
Health System Resilience	Current health expenditure per capita	USD per capita	INFORM Risk
Health System Resilience	Access to healthcare index	Index	INFORM Risk
Health System Resilience	Physician density	/ 10,000 population	WHO
Health System Resilience	Hospital beds	/ 10,000 population	WHO
Health System Resilience	Percentage of population covered by insurance	%	Our World in Data
Health System Resilience	Health capacity in clinics, hospitals and community care centres	Index	GHS Index
Health System Resilience	Supply chain for health system and health care workers	Index	GHS Index
Health System Resilience	Prevention	Index	GHS Index
Health System Resilience	Detection and reporting	Index	GHS Index
Health System Resilience	Rapid response	Index	GHS Index
Health System Resilience	Health spending US\$ per capita	\$	WHO
Health System Resilience	Maternal mortality rate	/ 100,000 births	INFORM Risk
Health System Resilience	Life expectancy	number	HDI
Health System Resilience	Priority to health (GGHE-D%CHE)	%	WHO
Individual Resilience	Political Rights	Index	Social Progress Index
Individual Resilience	Freedom of peaceful assembly	Index	Social Progress Index
Individual Resilience	Freedom over life choices	Index	Social Progress Index
Individual Resilience	Count on help	Index	Social Progress Index
Individual Resilience	Discrimination and violence against minorities	Index	Social Progress Index
Individual Resilience	Equal access index	Index	Social Progress Index
Individual Resilience	Interpersonal violence	Index	Social Progress Index
Institutional Resilience	Government Effectiveness	Index	INFORM Risk
Institutional Resilience	Constraints on government power	Index	INFORM Risk
Institutional Resilience	Corruption Perception Index	Index	INFORM Risk
Institutional Resilience	Order and security	Index	Rule of Law Index

6.2.2 Descriptive statistics

After the collection of the dataset, an initial analysis was performed to improve comprehension of the data. For each indicator included in the dataset, the descriptive statistics were calculated to analyse the structure of the data. These results are presented in Table G.6 and visualised using histograms, which can be found in appendix G. There are several points that stand out from these statistics:

- High variability in economic indicators including health expenditure and GDP per capita suggesting economic and healthcare investment disparities, which may possibly correlate with differences in health outcomes.
- Disparities in basic sanitation services use, highlighting that access to sanitation is still an issue in many of the countries included in the analysis, affecting exposure to waterborne disease.
- High levels of housing overcrowding potentially enhancing the spread of diseases.
- Wide range in internet usage and educational attainment, possibly impacting communication during emergencies and access to health information.
- High variability in health system capacity and performance highlighting disparities.
- Demographic variability with large variations in urban population and population density, possibly affecting countries' exposure to risks and resource allocation.

Table 20: Dataset after pre-processing including tranformation, standardization and correcting theoretical orientation.

	Education	HIV prevalence	U5 underweight	Internet Usage	Phone subscriptions	DTP3 coverage
Australia	2.42	0.52	1.35	1.73	0.28	0.96
Bangladesh	-0.24	0.52	-0.74	-0.46	0.39	1.14
China	-0.77	0.39	1.31	0.84	0.78	1.20
Dominican Republic	0.58	0.26	1.25	1.31	-0.26	0.30
India	-0.25	0.49	-1.64	-0.18	-0.43	0.36
Mozambique	-1.73	-3.19	0.05	-1.28	-1.63	-1.07
Nigeria	-0.07	0.13	-0.31	0.17	-0.14	-1.37
Pakistan	-1.19	0.49	-0.79	-1.14	-0.44	0.24
Peru	0.52	0.42	1.34	0.77	0.97	0.18
Philippines	1.16	0.49	-0.37	0.07	1.44	-1.31
South Sudan	-0.23	-0.13	-1.26	-1.70	-2.00	-1.79
Sri Lanka	-0.03	0.52	-0.53	0.60	1.38	1.02
Tanzania	-0.17	-0.91	0.33	-0.74	-0.34	0.12

6.2.3 Pre-processing the data

After the collection of the data, the dataset was reviewed to identify missing values or inconsistencies. Indicators that lacked data for more than 2 countries were excluded from the database. Some indicators related to poverty, for example the multidimensional poverty index lacked data for Australia and the United Kingdom as this index is aimed at developing countries. Therefore, values corresponding to a low level of poverty was assigned. For other variables, the mean of the dataset was calculated to fill in missing values. Python was then used to standardize the data using the z-score method, after which the theoretical orientation was assigned. [Cutter et al., 2014, Maharrani et al., 2024] The results of this step are presented in Table 20.

6.2.4 Clustering countries based on flood-vectorborne disease exposure

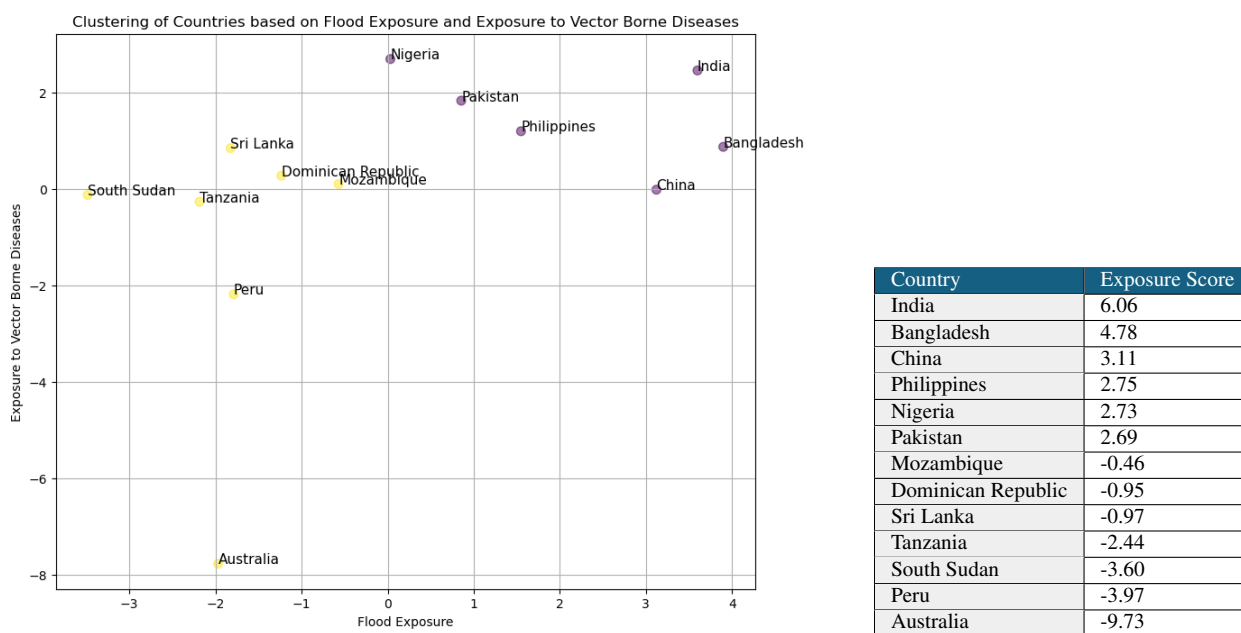
In exploring the interplay between flood disasters, outbreaks of communicable diseases, and the underlying social determinants of healthcare systems, it is essential to consider not only the determinants themselves but also the contextual factors of exposure to floods and diseases. As the level of exposure might have a potential relation to either the vulnerability of countries or their level of preparedness for events, clustering the countries reduces the uncertainty surrounding the effects of exposure level, simplifying the identification of potential relationships within healthcare systems. Therefore, the level of exposure should be considered when evaluating the potential for an outbreak following a flood event.

For this reason, it was suggested to cluster the countries, using k-means clustering, based on the indicators corresponding to exposure to floods and exposure to vector borne disease, as shown in Table 17. To verify the selection of indicators, the correlation matrix was assessed, shown in appendix H. Positive correlations are identified between the indicators describing the exposure to floods. Children U5 and artificial surfaces also show positive correlation with population exposed to dengue and population at risk of malaria, indicating that these factors are contributing to the exposure. Another important observation is the UK's lack of exposure to vector borne diseases, which is indexed to be 0. For this reason, the UK will also be excluded from further analysis when comparing the healthcare systems.

Table 21: Countries categorized in two levels of exposure based on k-means clustering results

Category 1: Low Exposure	Category 2: High exposure
South Sudan, Sri Lanka, Tanzania, Peru, Mozambique, Dominican Republic	Nigeria, Pakistan, Philippines, India, China, Bangladesh

Before conducting the k-means clustering, several preparatory steps were taken. Firstly, the dataset, comprising indicators describing exposure, was standardized using the z-score method. This process assigned a score to each country, indicating their level of flood exposure and exposure to vector-borne diseases, by summing up the standardized scores of corresponding exposure indicators. Subsequently, the number of clusters (k) was specified as two, and the k-means clustering model was initialized with a random seed for reproducibility. The algorithm then fit the k-means model to the standardized data to determine cluster centroids [Ikotun et al., 2023]. To visualize the outcomes, the exposure to floods was plotted against the exposure to vector-borne diseases. The results of the k-means clustering are depicted in Figure 18, revealing two distinct clusters: one with relatively low exposure to both floods and vector-borne diseases, and another with relatively high exposure, as summarized in Table 21. These clusters primarily differ in terms of flood exposure, rather than exposure to vector-borne diseases, which is faced to a more similar extent. Australia, identified as an outlier, will not be directly compared with other countries in further analysis. These clusters are crucial to consider when evaluating PCA results, providing insight into the context of healthcare systems and the level of risk faced by each country.



(a) K-means clustering to cluster countries based on their flood and vector-borne disease exposure level (b) Overall exposure score based on standardized indicator data

Figure 18: Results of k-means clustering, based on the level of exposure

6.2.5 Removing redundant data and verifying use of PCA

An initial step in removing redundant data was to observe for any overlapping indicators describing the same characteristic to prevent double-counting. Secondly, a distinction was made between the dataset to analyze vector-borne and waterborne diseases. For waterborne diseases, there was no evidence indicating which countries might be more or less resilient and therefore the dataset will be adjusted accordingly. Before calculating correlations, certain exposure indicators were removed based on the indicator tree diagram, illustrated in Figure 14. These included indicators related to the quality of water and sanitation systems, the percentage of people using at least basic sanitation, population density, the percentage of the population living in urban areas, and household size. Additionally, some indicators with high correlations describing the same type of resilience characteristic were combined by averaging them. The new indicators are presented in Table 22. Subsequently, a correlation matrix was calculated to identify correlations that were too high or too low. Correlations above 0.85 were evaluated to determine if they overlapped too much and were describing the same mechanisms. If this was indeed the case, one of these indicators was removed. In total, 25 indicators were removed from the analysis. Four were removed because they described the same characteristics, five were used for describing exposure to waterborne diseases, another five were used to cluster the countries based on their exposure to flood and vector borne diseases and an additional 15 were removed due to high correlation. An overview of the indicators removed during this process and the reason for exclusion is provided in Table 23. As a result of this process, a list of 22 indicators was constructed as illustrated in Table 24. Each category is represented in the final list of indicators and is often representing other variables highly correlated with. As a result of removing redundant data, the resulting correlation matrix is presented in Figure 19.

Table 22: Combined indicators describing the institutional and individual resilience

Original indicators	New indicator
Prevention, Detection and Reporting, Rapid response	Government Emergency Response
Political rights, Freedom of peaceful assembly	Individual Political Freedom
Freedom over life choices, Count on help, Discrimination and violence against minorities, Equal access, interpersonal violence	Individual autonomy and Safety

Table 23: Indicators that have been removed in the PCA process including the reason for exclusion

Removed indicators	Reason for exclusion
Mean Income, State legitimacy, Medical doctors, Healthcare access	Overlap with indicators describing the same characteristic
WaSH (Water, Sanitation and Hygiene, People using at least basic sanitation services, Household Size, People using at least basic drinking water service	Indicators describing exposure to waterborne diseases
Annual expected exposed people to floods, flood frequency, Land use - artificial surfaces, Population exposed to Dengue, Population at risk of Malaria	Indicators describing exposure to flood-vector borne diseases
Literacy rate, U5 mortality, Recently affected by natural disaster, Current health expenditure per capita, Government Emergency Response, Urban population, Life expectancy, Physician Density, Access to healthcare, Corruption Perception Index, Regulatory enforcement, Multidimensional Poverty Index, Employment to population ratio, Gender-inequality index	Correlation >0.85

Table 24: Indicators included in PCA

Indicator	Category	Number in Category
Number of children under 5	Vulnerability: Socio-economic	
Mean income	Vulnerability: Socio-economic	
Population ages 65 and above	Vulnerability: Socio-economic	
Population density	Vulnerability: Socio-economic	4
People living with HIV - Adult 15+	Vulnerability: Ecological	
Under 5 underweight	Vulnerability: Ecological	2
Educational attainment	Vulnerability: Behaviour	
Individuals using the internet	Vulnerability: Behaviour	
Mobile cellular subscriptions	Vulnerability: Behaviour	3
		Total Vulnerability: 9
Hospital beds	Resilience: Health care system	
Population covered by health insurance	Resilience: Health care system	
Health capacity in HC	Resilience: Health care system	
Supply chain for health care system	Resilience: Health care system	
Maternal mortality rate	Resilience: Health care system	
Priority to health (GGHE-D%CHE)	Resilience: Health care system	
Population with access to 3 doses DTP3	Resilience: Health care system	7
Government Effectiveness	Resilience: Institutional Resilience	
Order and security	Resilience: Institutional Resilience	
State legitimacy index	Resilience: Institutional Resilience	3
Individual Political Freedom	Resilience: Individual Resilience	
Individual autonomy and safety	Resilience: Individual Resilience	
Housing overcrowding rate	Resilience: Individual Resilience	3
		Total Resilience: 13
		Total: 22

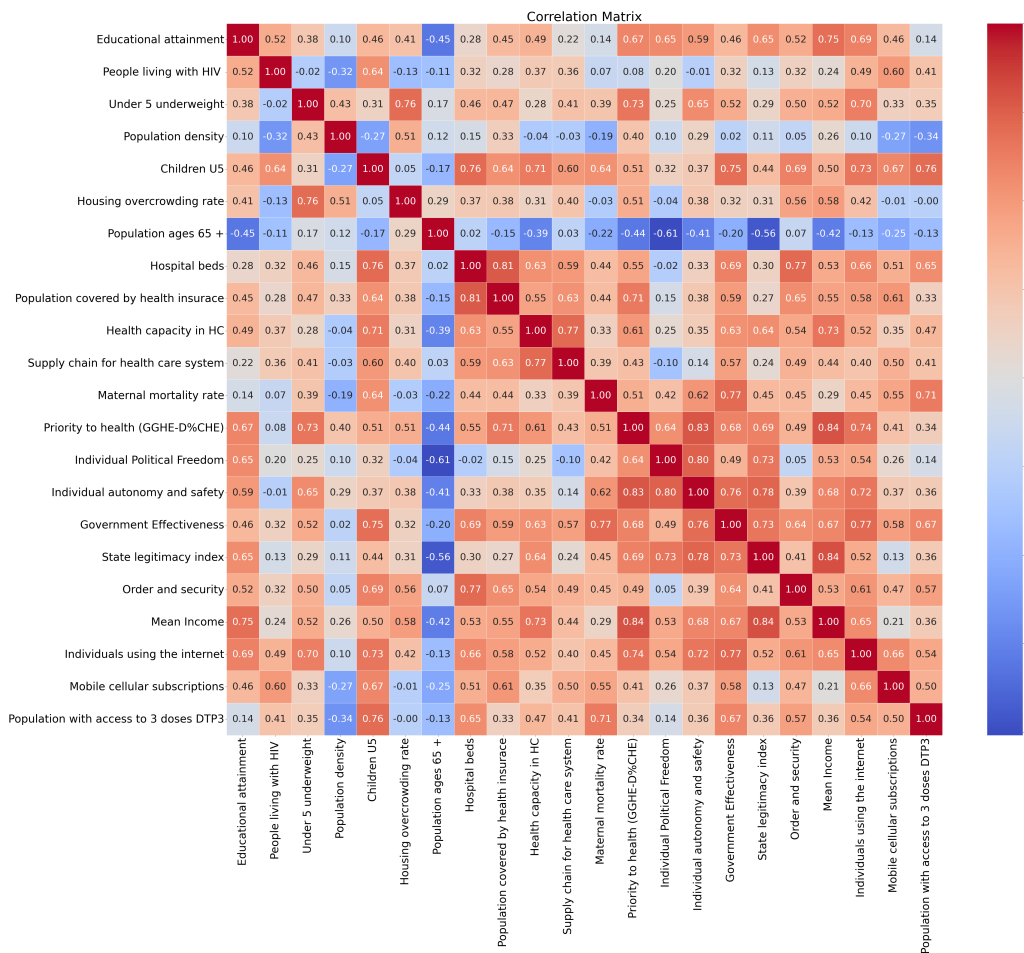


Figure 19: Correlation matrix of PCA for vector-borne diseases

Bartlett's test for sphericity was performed to assess the suitability of the data for PCA. It is important to note that Bartlett's test assumes a normal distribution, which may not hold true for all variables. A Shapiro-Wilk Test was conducted to test for normality, and it was found that not all variables followed a normal distribution. Results of this test can be found in Appendix I. Although, the p-value of three indicators showed that it is not possible to reject H_0 , they will be included for analysis as normal distribution is not mandatory. However, it should be considered during the interpretation of the results.

Subsequently, to test the suitability of the data for PCA, the Bartlett's test was conducted for vector-borne diseases. The resulting p-value was 0.00, indicating a statistically significant result. Additionally, the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was calculated to be 0.60, suggesting mediocre but sufficient adequacy for PCA.

6.2.6 PCA and selection of number of components

The indicators in Table 24 were used for PCA analysis to derive the dataset's PCs. Following the Kaiser criterion, which suggests retaining eigenvectors with eigenvalues greater than 1, six PCs were selected for analysis [Braeken and Van Assen, 2017]. To validate this, the cumulative explained variance was plotted against the number of PCs. Figure 20 demonstrates that six PCs explain nearly 90% of the variances between countries, a reasonable amount.

Several methods can be used to visualise the results of the unrotated PCA, including scatter plots and loading scores heatmaps, as shown in Figure 21. The scatter plot (Figure 21a) visualizes the reduced-dimensional representation of the data. The loading scores (Figure 21b) represent the correlation coefficients between the original indicator and the PCs. The contribution of each original indicator to the variance captured by each PC is indicated. Thus, the higher the absolute loading score, the more variance between countries is explained by that indicator. Based on the loadings scores, the most dominant variables describing each PC can be summarised, as shown in Table 25.

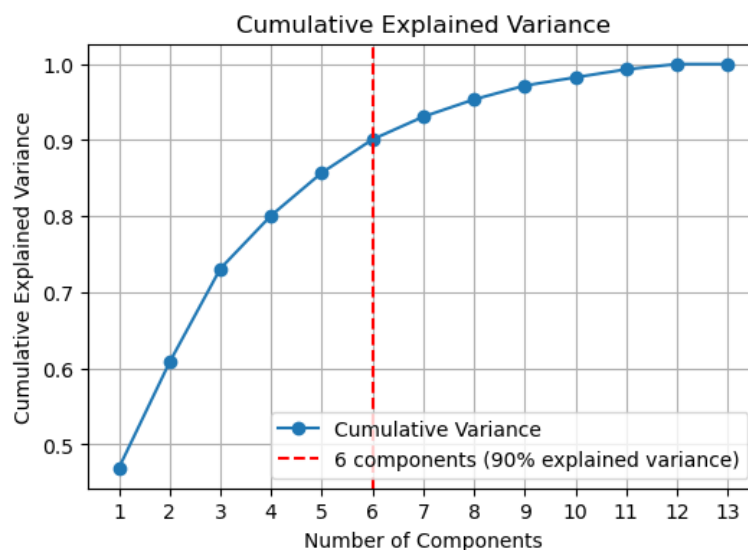


Figure 20: Cumulative explained variance for the five retained principal components after performing principal component analysis

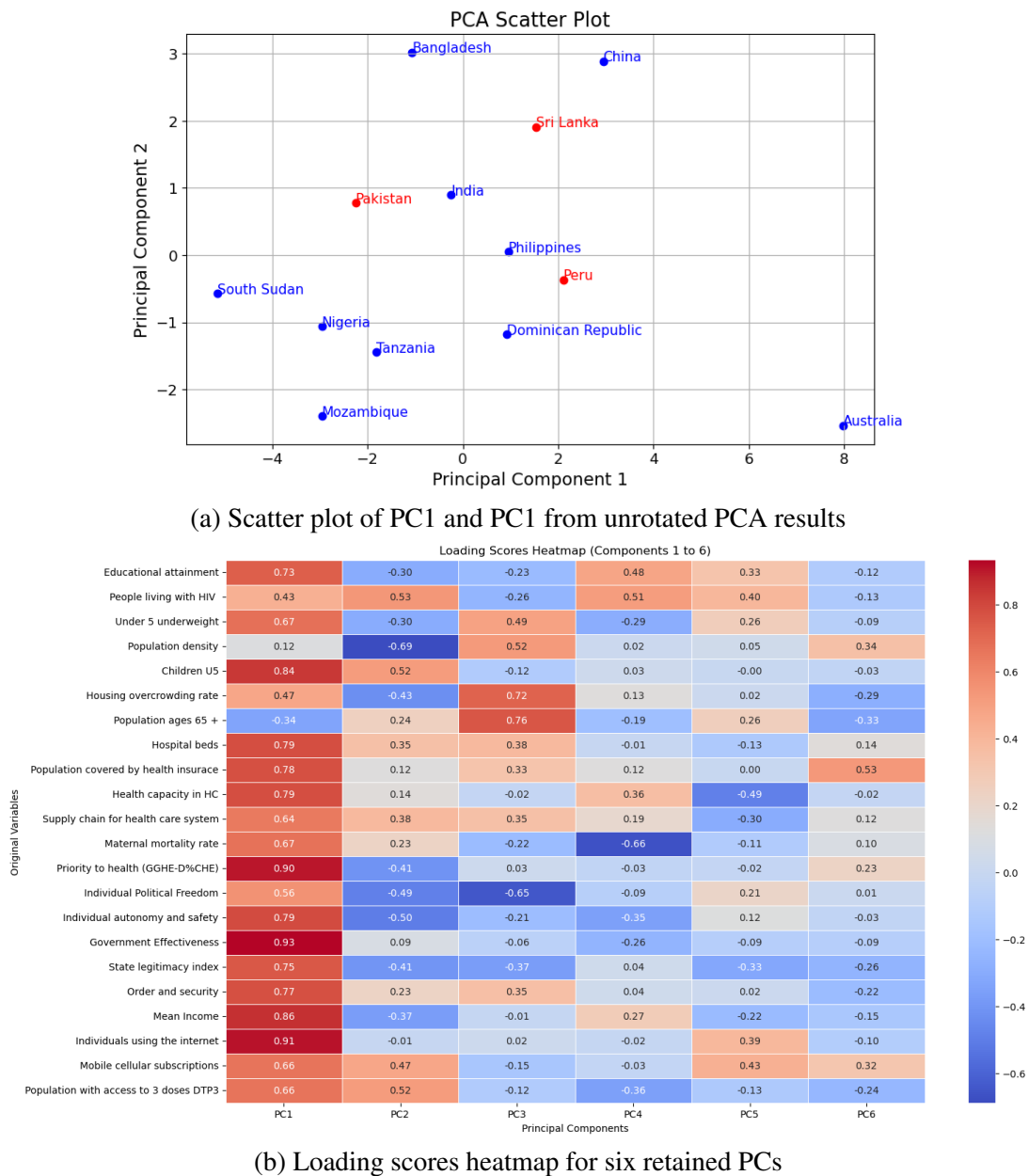


Figure 21: Visualisation of Exploratory Data Analysis results

Table 25: Summary of results unrotated PCA

Component	Variance	plus or min	Dominant variables	Loadings
PC1	46.90%	+	Government effectiveness Individuals using the internet Priority to health	0.93 0.91 0.90
PC2	13.91%	-	Population density	-0.69
PC3	12.29%	+	Housing overcrowding Population 65+	0.72 0.76
PC4	6.98%	-	Maternal mortality rate	-0.66
PC5	5.62%	-	Health capacity	-0.49
PC6	4.38%	+	Insurance	0.53

6.2.7 Resilience Score

A resilience score can be calculated based on the loadings of the original variables on each of the PCs. These scores are calculated as the dot product of the standardized data and the loadings of the selected PCs. The total score is the sum of the scores across all retained PCs for each country, providing an overall measure of the countries' resilience. The scores for each PC and the corresponding overall score are shown in Table 26, offering insight into which PCs and dominant variables might contribute to the variances explained between countries. Firstly, it is notable that, based on this method, both Peru and Sri Lanka have relatively high resilience scores, while Pakistan scores relatively low.

To determine which indicators might contribute to the variances in resilience scores between countries, it is possible to examine the individual scores of each PC. Sri Lanka shows a relatively low score for PC2, which is associated with high population density. Peru scores relatively low on PC5, indicating low health capacity. However, as the loading score for this PC is relatively low, it is necessary to consider other variables as well. This suggests that the state legitimacy index might explain the variance with other countries. The low resilience score of Pakistan is mainly explained by PC1, indicating possible low government effectiveness, limited internet use, and low priority to health. Upon further examination of the data, it becomes apparent that the variance might mostly be explained by the first two indicators.

Table 26: Resilience scores for each country based on the loadings of the original variables on each principal component, * indicating countries that were found to have a positive correlation between floods and vector borne disease outbreak(s).

	Category	PC1	PC2	PC3	PC4	PC5	PC6	Total Score
Australia	-	26.64	4.62	-0.76	-1.78	1.57	-0.55	29.74
China	High	9.83	-5.26	7.50	0.66	0.43	-0.35	12.81
Peru*	Low	7.05	0.67	-0.09	0.08	-0.98	1.14	7.86
Dominican Republic	Low	3.08	2.15	0.10	1.76	-2.57	-1.48	3.04
Sri Lanka*	Low	5.12	-3.48	-1.66	0.43	-0.57	1.93	1.77
Mozambique	Low	-9.89	4.37	1.38	2.74	2.21	0.80	1.60
Philippines	High	3.19	-0.10	-1.06	-0.95	-1.35	1.11	0.84
Tanzania	Low	-6.06	2.63	0.78	1.66	-0.63	-0.68	-2.30
India	High	-0.81	-1.64	-4.31	0.23	1.01	-0.75	-6.26
Pakistan*	High	-7.52	-1.42	-1.92	0.51	0.86	0.55	-8.93
Nigeria	High	-9.87	1.93	-0.06	-1.71	-1.19	-0.17	-11.07
Bangladesh	High	-3.56	-5.50	-2.92	-0.16	0.76	-1.49	-12.86
South Sudan	Low	-17.20	1.04	3.01	-3.45	0.44	-0.09	-16.24

6.2.8 Varimax Rotation

Varimax rotation was performed to enhance the interpretability of the PCA results. A summary of the results is shown in Table 27. Based on these results, a resilience score can be assigned based on the six PCs that were retained. It should be noted that the loadings of the PCs calculated by the Varimax rotation are higher than in the unrotated results, especially for the higher PCs. The variance of each PC is equal to the original variance. However, as a result of varimax rotation, the loadings of the original variables on each PC are adjusted, as varimax rotation reorients the axes in the multidimensional space. This adjustment can lead to a redistribution of variance within the PCs, causing different indicators to contribute more or less to the variance explained by each PC. The resilience score based on the retained principal components is shown in Table 28.

When comparing Peru, Sri Lanka, and Pakistan with the other countries, the following observations can be made. Firstly, Peru and Sri Lanka again score relatively high, and Pakistan relatively low. The high score of Peru is mainly determined by PC3: Social equity. No PC score indicates a large deficiency. Sri Lanka scores well on PC5: Health performance. However, the country scores relatively low on PC2: economic disparity. By analyzing the dataset, it can be seen that this is primarily caused by a high housing overcrowding rate. Pakistan scores overall relatively low, but primarily on PC2: economic disparity. In this case, both dominant indicators contribute to the low score.

Table 27: Summary of rotated results PCA

Component	Variance	plus or min	Dominant variables	Loadings
PC1: Health Capacity	46.90%	+	Health capacity in HC Supply chain of health care system	0.88 0.83
PC2: Economic disparity	13.91%	-	Housing overcrowding Under 5 underweight	-0.93 -0.86
PC3: Social equity	12.29%	-	Individual political freedom, State legitimacy Population 65+ Individual autonomy and safety	-0.93 -0.89 0.81 -0.78
PC4: Disease prevalence	6.98%	+	HIV	0.94
PC5: Health performance	5.62%	+	maternal mortality rate, DTP vaccination coverage	0.93 0.83
PC6: Financial security	4.38%	+	Health insurance	0.64

Table 28: Resilience scores for each country based on the loadings of the original variables on each rotated principal component, * indicating countries that were found to have a positive correlation between floods and vector borne disease outbreak(s)

	Category	PC1	PC2	PC3	PC4	PC5	PC6	Total Score
Australia	-	13.46	10.69	17.48	6.48	9.14	3.35	60.60
China	High	9.00	6.57	-2.45	3.38	6.00	1.92	24.42
Peru*	Low	2.85	2.78	4.02	2.42	2.87	2.51	17.45
Sri lanka*	Low	2.90	-1.12	1.43	3.36	4.06	2.42	13.05
Philippines	Low	1.00	0.39	2.18	2.43	0.76	1.68	8.45
Dominican Republic	Low	-1.51	3.55	2.37	1.16	2.03	0.14	7.74
India	High	-0.80	-4.07	1.19	0.64	1.20	-1.74	-3.59
Bangladesh	Low	-0.60	-5.53	-3.38	1.61	1.05	-2.88	-9.72
Tanzania	High	-4.80	0.03	-2.28	-3.45	-2.57	-1.03	-14.10
Pakistan*	High	-3.59	-5.08	-3.62	-2.29	-2.23	-1.23	-18.04
Nigeria	High	-5.72	-2.57	-3.88	-2.37	-6.58	-1.50	-22.62
Mozambique	High	-5.67	-1.42	-3.39	-8.14	-4.38	-0.77	-23.79
South Sudan	Low	-6.51	-4.22	-9.65	-5.23	-11.36	-2.89	-39.86

6.3 Results PCA: categories

Earlier in the process, the countries were categorized based on their exposure level to floods and diseases as the likelihood of the occurrence of a disease outbreak depends on both the exposure level and the resilience of the country. To further analyze the differences between the countries with high resilience and low resilience, another PCA was performed per category. Again, the results were considered both with and without rotation.

6.3.1 Low exposure

To perform the PCA, the steps presented in subsection 6.1 were performed again. After assessing the collinearity and removing redundant data, 15 indicators were selected to be included in the PCA (Table 29). Notably, all indicators representing the individual resilience category were excluded from the dataset to assure sufficient adequacy. This should be considered when interpreting the results. The suitability of the dataset was tested with the Bartlett's test ($p=0.00$) and the KMO ($KMO = 0.54$). Notably, this KMO value is quite low, indicating poor adequacy. The number of PCs to retain was identified to be 5 representing almost 100% of the variance within the dataset, as shown in Figure 22.

Table 29: Indicators included in PCA for Low Exposure countries

Component	Category
Number of children under 5 years old	Vulnerability: Socio-economic
Employment to population ratio, 15+	Vulnerability: Socio-economic
Population ages 65+	Vulnerability: Socio-economic
Under 5 underweight	Vulnerability: Ecological
% affected by natural disaster (last 3 years)	Vulnerability: Ecological
Educational attainment	Vulnerability: Behaviour
Individuals using the internet	Vulnerability: Behaviour
Population with access to 3 doses DTP3	Resilience: Health care system
Population covered by Health Insurance	Resilience: Health care system
Supply chain for health system and health care workers	Resilience: Health care system
Emergency Response	Resilience: Health care system
Maternal mortality rate	Resilience: Health care system
State legitimacy index	Resilience: Institutional
Order and security	Resilience: Institutional
Regulatory enforcement	Resilience: Institutional

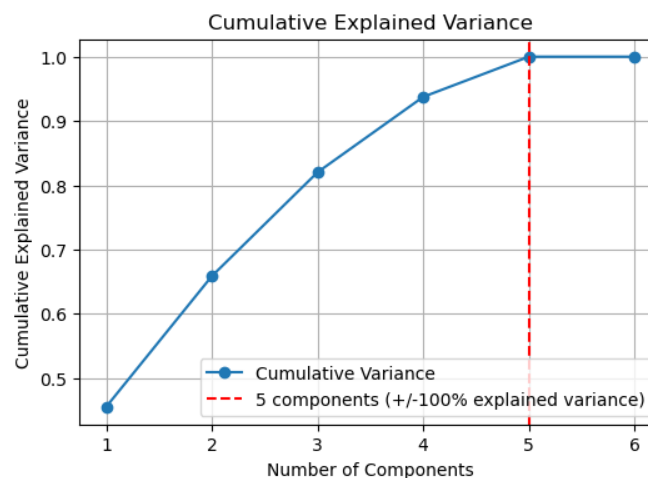


Figure 22: Cumulative explained variance for the five retained PCs of the PCA for low exposure countries

Unrotated results PCA low exposure

The summarized results of the unrotated PCA are based on the dominant variables in each PC and are presented in Table 30a. Resilience scores can again be assigned based on the loadings of the original variables on each PC, as shown in table 30b. Peru receives a high resilience score, mainly explained by its high PC1 value, characterized by good regulatory enforcement, high levels of internet usage among individuals, and high DTP coverage. However, further analysis reveals that many indicators have relatively high loadings, suggesting that this value cannot be solely explained by examining the most dominant variables. Peru scores relatively low on PC3, associated with a high number of elderly individuals in the country. Similarly, Sri Lanka also scores high on PC1 but performs poorly on both PC2 and PC3, indicating a significantly lower employment ratio and relatively low state legitimacy.

Table 30: Results (Low exposure, unrotated)

(a) Summary of results unrotated PCA, category: low exposure

PC	Variance	+/-	Dominant variables	Loadings
PC1	45.52%	-	Regulatory enforcement Individuals using the internet DTP vaccination coverage	-1 -1 -1
PC2	20.34%	-	Employment ratio State legitimacy	-0.94 -0.83
PC3	16.16%	+	Population 65+	0.93
PC4	11.73%	+	Supply chain	0.80
PC5	6.25%	-	Education Order and security	-0.5 -0.52

(b) Resilience scores based on the loadings of the original variables on each PC, * indicating countries highly impacted by floods

	PC1	PC2	PC3	PC4	PC5	Score
Peru*	8.93	0.63	-0.73	2.93	1.10	12.86
Dominican Republic	3.66	-0.49	5.33	-0.13	-0.86	7.50
Tanzania	-1.80	4.05	0.25	-2.77	1.22	0.94
Sri lanka*	7.48	-3.58	-3.05	-1.98	-0.72	-1.83
Mozambique	-6.26	4.00	-2.00	1.30	-1.40	-4.35
South Sudan	-12.01	-4.61	0.19	0.64	0.66	-15.12

Rotated results PCA low exposure

To simplify these results, the application of varimax rotation is proposed. The summary of the results of the rotated PCA is presented in Table 31a. The resilience scores, shown in 31b again show a high resilience score corresponding to Peru and Sri Lanka. For Peru, this high score is primarily influenced by high PC1 and PC4 scores, corresponding to state legitimacy, maternal mortality, U5 underweight, supply chain of the healthcare system, and government emergency response. It shows a relatively low score for PC3, representing an increased vulnerable population. Sri Lanka demonstrates a high score for PC2, primarily composed of a high score for health insurance. However, Sri Lanka also demonstrates a low score for the vulnerable population.

Table 31: PCA results (low exposure, rotated)

(a) Summary of results rotated PCA, category: low exposure

Component	Variance	+/-	Dominant variables	Loadings
PC1	45.52%	-	State legitimacy Maternal mortality U5 underweight	-1.05 -0.99 -0.93
PC2	20.34%	+	Employment ratio Health insurance Children U5	-1.03 0.93 0.86
PC3	16.16%	+	Affected by natural disaster Population 65+	1 0.91
PC4	11.73%	+	Supply chain Government emergency response	1.07 0.89
PC5	6.25%	-	Order and security Education	-1.02 -0.97

(b) Resilience scores based on the loadings of the original variables on each PC, * indicating countries highly impacted by floods

	PC1	PC2	PC3	PC4	PC5	Score
Peru*	5.38	3.04	-1.48	6.40	3.02	16.37
Dominican Republic	4.07	1.30	3.73	-0.46	3.22	11.86
Sri Lanka*	1.94	7.15	-3.54	1.99	3.32	10.87
Tanzania	0.93	-4.03	-2.17	-2.64	0.14	-7.78
Mozambique	-1.75	-4.64	-0.74	-1.44	-5.97	-14.54
South Sudan	-10.58	-2.82	4.20	-3.86	-3.73	-16.78

6.3.2 High exposure

The same method is employed to determine the relative position of the countries categorized as high exposure. Nineteen indicators were selected to perform PCA. The dataset was tested for suitability (p-value = 0.0 and KMO statistic = 0.57). Again, five PCs are retained to analyze the resilience scores.

Unrotated results PCA High exposure

The summary of the PCA results is shown in Table 32 and the resilience scores are presented in Table 32b. Compared to other highly exposed countries, Pakistan scores low on the resilience score, which is caused by low scores on several PCs, especially PC1, PC2, and PC5. These PCs primarily result from several indicators related to healthcare system performance and individual resilience. Upon further examination of the original data, this low score is primarily caused by low health capacity, low government effectiveness, low educational attainment, and a low percentage of internet usage.

Table 32: Results PCA (high exposure, unrotated)

(a) Summary of results unrotated PCA, category: high exposure

PC	Variance	+/-	Dominant variables	Loadings
PC1	41.40%	+	Physicians density	0.98
			Life expectancy	0.91
			Government effectiveness	0.90
			Health insurance	0.87
			Health capacity	0.86
PC2	23.13%	+	Individual autonomy and safety	1.07
			Education	0.84
PC3	19.29%	+	State legitimacy index	0.95
PC4	10.10%	-	Population density	-0.87
PC5	5.90%	-	Individuals using the internet	-0.57

(b) Resilience scores based on the loadings of the original variables on each PC, * indicating countries highly impacted by floods

	PC1	PC2	PC3	PC4	PC5	Score
China	17.86	-1.24	-1.30	1.48	0.24	17.05
India	-2.74	3.79	7.74	1.11	0.67	10.57
Philippines	0.67	7.86	-2.45	-2.29	-1.10	2.69
Bangladesh	-0.73	-6.32	1.15	-3.55	0.74	-8.69
Nigeria	-9.02	0.70	-5.28	1.60	1.54	-10.46
Pakistan*	-6.04	-4.80	0.14	1.65	-2.10	-11.15

Rotated results PCA High exposure

To further simplify the interpretation, the varimax rotation was applied, presented in Table 33 together with the resulting resilience score in Table 33b. Similar to the unrotated results, Pakistan scores low on the level of resilience. Although, each PC represents a relatively low value, the low score is primarily caused by low scores on PC2 and PC5 which are represented by mobile cellular subscription, education, individuals using internet and employment ratio which are all indicators related to human capital development.

Table 33: Results PCA (high exposure, rotated)

(a) Summary of results rotated PCA, high exposure

PC	Variance	+/-	Dominant variables	Loadings
PC1	41.40%	+	Life expectancy	0.99
			Health capacity	0.98
			Under 5 mortality	0.98
			% affected by disaster	0.97
			Vaccination coverage	0.95
PC2	23.13%	+	Mobile cellular subscription	1.01
			Education	0.88
PC3	19.29%	+	Constraints on government power	1.04
			State legitimacy	1.02
			Individual political freedom	0.95
PC4	10.10%	-	Population density	-1.02
PC5	5.90%	-	Individuals using internet,	-1
			Employment ratio	-0.82

(b) Resilience scores based on the loadings of the original variables on each PC, * indicating countries highly impacted by floods

	PC1	PC2	PC3	PC4	PC5	Score
China	11.82	3.23	-8.71	6.67	7.34	20.35
Philippines	-2.34	6.82	3.17	2.82	2.20	12.66
India	2.18	-2.12	8.58	-0.10	-0.72	7.81
Bangladesh	1.96	-1.42	-3.04	-5.81	-2.45	-10.77
Nigeria	-10.34	-1.96	0.81	-1.62	-0.82	-13.93
Pakistan*	-3.27	-4.55	-0.80	-1.96	-5.55	-16.13

Chapter 7

Final Framework

7 Final Framework

In the first phase of the EDA, both V-EDA and statistical testing revealed suspected correlations between floods and outbreaks of communicable diseases. Additionally, countries were categorized based on the impacts they have faced in the past. In the previous chapter, the results of the PCA were presented, comparing countries labeled as 'impacted' with those not showing evidence of being impacted. These comparisons were made using indicators that describe the vulnerability and resilience factors of healthcare systems as defined by the conceptual framework. In this chapter, the key findings from the PCA will be discussed in the context of the conceptual framework, and proposed changes to the model will be introduced to construct the final framework.

7.1 Evaluation of PCA results

7.1.1 General considerations

The PCA was performed to gain insight into the indicators contributing to variances between countries by reducing dimensionality through the composition of PCs. Based on the PCA, five or six PCs, depending on the sub-analysis, were retained to explain at least 90% of the variance between countries. Although multiple PCs are needed to identify this level of variance, PC1 consistently explains the most variance, representing over 45%. Therefore, the weight of PC1 should be considered accordingly, making indicators describing variances in this first PC relatively more important. To simplify interpretation of the results, both unrotated and rotated PCs are considered. It can be concluded that the results are easier to interpret and provide similar outcomes to the unrotated ones. However, both results are considered to cover all potential relevant findings.

7.1.2 Full sample size

Based on the PCA results, Peru and Sri Lanka exhibit relatively high overall scores, often placing them in the better-performing half of the sample. This suggests that the method, which frames resilience as a function of various determinants contributing to health including socio-economic factors, public health, behavior, health care system performance, institutional quality, and individual resilience, evaluates these countries as relatively resilient. However, this does not imply that there are no insights to be gained or areas for improvement. Upon closer examination, it is evident that Sri Lanka faces economic disparities, as indicated by issues like housing overcrowding and the prevalence of underweight children under five years old, potentially undermining its resilience. This pattern is also observed in Pakistan, which is generally considered less resilient. When comparing these three countries to the full dataset, none of them score highly on indicators of individual autonomy and safety or political freedom. Additionally, both Pakistan and Sri Lanka exhibit very high population densities and significant housing overcrowding. The impacted countries can be compared with those that score similarly on most indicators:

- Overall, Pakistan, Sri Lanka, and Peru score moderately on individual autonomy and safety and political freedom. They also have relatively high population densities.
- Pakistan and Sri Lanka vs. Bangladesh: Bangladesh has similar demographic and economic characteristics, such as population density and housing overcrowding rates. However, Bangladesh has a higher employment ratio, better government emergency response, and increased health capacity, which might suggest a better individual position and preparedness for such events.

- Peru vs. Dominican Republic and China: The Dominican Republic, which shares similar characteristics, scores better on individual autonomy and safety, political freedom, higher income, and more internet usage. Peru can also be compared to China, which has a stronger economic position and better performance across all healthcare characteristics. However, China does not score highly due to lower indicators related to individual political freedom and state legitimacy, which might prevent the country from achieving even better resilience.

It appears that multiple factors contribute to better resilience to the impacts of floods on communicable disease incidences. A country's economic position and the resources it allocates to health significantly enhance preparedness. However, when comparing countries with similar economic statuses, the standout differences are often in government emergency response and individual factors related to social cohesion, feelings of safety, and civic participation.

7.1.3 Low/High exposure

When considering the two distinctive categories, the results of the PCA corresponding to the low exposure categories reveal several findings.

- When comparing Peru and Sri Lanka with Tanzania and the Dominican Republic, Tanzania exhibits a higher employment ratio, suggesting better employment opportunities and greater trust in government institutions. In the context of emergency response, this could present challenges regarding adherence to government policies. The Dominican Republic does not have a high percentage of its population that has experienced natural disasters in the past couple of years, which could indicate the importance of recovery periods. Countries with higher frequencies of floods or other disasters might be more susceptible to their impacts. Both Tanzania and the Dominican Republic demonstrate higher state legitimacy, suggesting that citizens are more inclined to follow government directives, which is important in emergency disaster response where adherence to protocols is essential. Furthermore, Peru shows an increased elderly population, which increases vulnerability. The rotated results yield similar insights.
- High exposure: Compared to other countries in this category, Pakistan exhibits low health capacity. Contrasting with Nigeria, a country with limited financial resources, Pakistan focuses more on individual indicators such as low educational attainment levels and limited internet usage. This suggests that while Nigeria holds a vulnerable position, it mitigates some of these risks through stronger individual resilience.

The analysis reveals that countries with higher state legitimacy and better employment ratios tend to show greater resilience in disaster scenarios. Moreover, countries with frequent exposure to natural disasters may face extended recovery challenges, highlighting the importance of effective emergency response mechanisms. Strengthening healthcare systems and improving individual resilience factors, such as education and internet access, are essential for enhancing a country's overall capacity to withstand and recover from disasters.

7.1.4 Summary of key findings

Overall, the PCA results do not provide sufficient insight to establish conclusive relationships and patterns between specific health determinants and countries' resilience to the impacts of floods on communicable diseases but show several exploratory findings that could indicate a suspected impact of determinants on the level of resilience to the impacts of floods. When evaluating the combined results, it becomes apparent that besides a strong economic position—typically associated with high health expenditure and effective healthcare system performance—individual resilience within social and institutional contexts, demographic factors, and state legitimacy also influence the distinction between resilient and non-resilient countries. Indicators such as education levels, internet access, employment rates, mobile cellular subscriptions, population density, age demographics, overcrowding, and state legitimacy are important in this assessment. A summary of the findings from the PCA is shown in Table 34.

Category	Findings
Exposure	<p>Exposure to Floods</p> <ul style="list-style-type: none"> Exposure level does not correlate with countries' resilience levels. High exposure to floods positively correlates with exposure to vector-borne diseases. <p>Exposure to Vector-borne Diseases</p> <ul style="list-style-type: none"> Increase in artificial surfaces correlates with exposure to vector-borne diseases. Increase in the number of children under 5 correlates with exposure to vector-borne diseases.
Vulnerability	<p>Socio-economic:</p> <ul style="list-style-type: none"> Increased elderly population negatively affects resilience. Higher population density decreases resilience. Economic disparity impacts resilience, seen in housing overcrowding and underweight children U5. High employment ratios positively affects resilience. <p>Public health:</p> <ul style="list-style-type: none"> Public health factors indicate vulnerability but exhibit a less evident relationship when comparing resilience scores. Higher prevalence of HIV correlates with decreased resilience. Under-5 mortality shows complex correlations, potentially linked to healthcare indicators. <p>Behavior:</p> <ul style="list-style-type: none"> Internet usage correlates positively with resilience. Education plays a significant role in strengthening resilience to disasters. Mobile cellular subscriptions also positively correlate with resilience.
Coping capacity	<p>Healthcare system resilience</p> <ul style="list-style-type: none"> Health financial security is an important indicators for the level of healthcare system resilience. Health capacity and performance positively influences a country's resilience and explains significant proportion of variance between countries. <p>Institutional resilience:</p> <ul style="list-style-type: none"> High state legitimacy positively affects resilience. <p>Individual resilience</p> <ul style="list-style-type: none"> Individual political freedom, autonomy, and safety contribute to overall resilience. Variances explained between countries with low financial capacities are primarily explained by individual indicators such as education attainment levels and internet usage.

Table 34: Summary of Findings

7.2 Final framework

At the outset of this thesis, a conceptual framework (Figure 7) was presented to illustrate the interplay among three domains initially identified in the literature: exposure, vulnerability, and resilience. By collecting data, a database comprising flood data, epidemiological data, and indicators of health-care systems was constructed as a foundation to explore the relationships between these elements. Through EDA, the aim was to gain a better understanding of the relationships between the different domains and to analyze which indicators of the healthcare system might be significant when comparing resilience levels of healthcare systems. To conceptualize the main findings, as presented in Table 34, they are integrated into the final Flood-Health Risk and Response framework, presented in Figure 23. This framework identifies the most crucial vulnerability and resilience factors defining a country's healthcare system resilience against flood impacts on vector-borne diseases. Indicators that were in the PCA identified as determinant variables in explaining variances between countries are categorised into seven distinct categories. Green arrows indicate positive effects, showing measures that reduce the impact of floods on communicable diseases, while red arrows indicate negative effects, showing factors that worsen these impacts. Importantly, all domains identified in the conceptual framework: exposure, vulnerability, and resilience, are included in the final framework, highlighting the holistic approach taken. The analysis does not identify a single indicator within each domain as decisive for resilience; instead, it suggests that resilience is influenced by various factors.

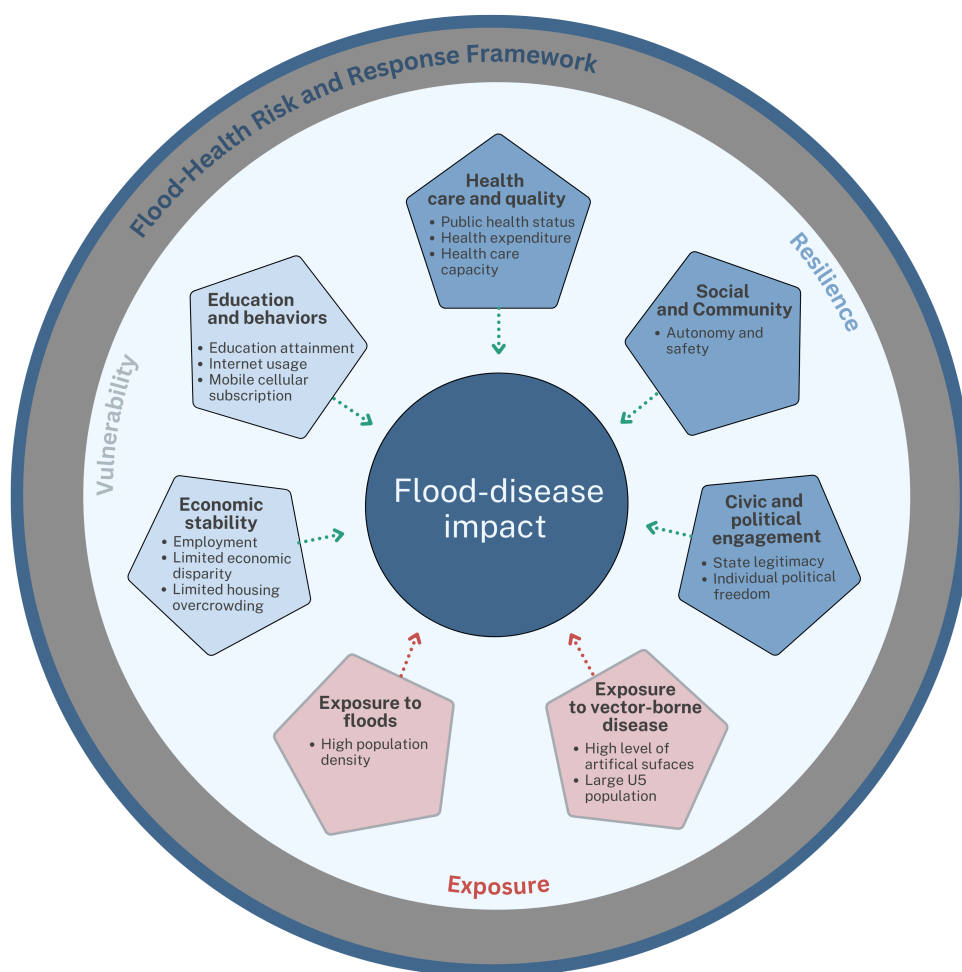


Figure 23: Final framework

7.3 Interpretation of Final Framework

The final framework (Figure 23) represents the main findings of this thesis. To understand the value of the framework, it should be understood that it does not capture the whole picture of the system describing the health effects of communicable diseases after floods but is based on the design choices made during this thesis. The final framework represent the healthcare systems' vulnerability and resilience factors that were found to be of most importance when explaining varying levels of impacts after flood disasters and thus not explain the system as a whole. It is important to realize that the interplay between floods, communicable diseases and health care systems is an interconnected system with different components leading to the risk of floods and diseases. Therefore, to understand the risk to a country, it is not enough to focus on just one of the highlighted areas; a multi-disciplinary approach should be taken. However, the results of the EDA presented several key findings that could be reflected upon.

- Economic stability provides a country with valuable resources that strengthen disaster resilience. This stability positively impacts the entire system, enhancing citizens' living conditions and improving healthcare systems. A solid financial position often correlates with higher healthcare quality and performance, as evidenced by the availability of resources and health expenditures. Additionally, strong economic health correlates with overall public health, reducing community vulnerability. Therefore, it can be concluded that LMICs are generally more susceptible to the impacts of floods due to their relative lack of economic stability.
- Additionally, recognizing individual resilience as a significant factor suggests the potential for empowerment strategies at a local level to enhance overall community resilience. Specifically, investing in education and behavioral initiatives enhances resilience by fostering greater awareness and providing channels for effective communication during disasters. Additionally, promoting autonomy and safety within communities contributes to heightened social cohesion and confidence, thereby improving the psychosocial well-being of residents. Furthermore, individual political engagement serves as a critical indicator, reflecting civic participation and involvement in resilience policies. Assessing how actively people understand and support mitigation measures by authorities gives valuable insights into the effectiveness of resilience-building efforts.
- Lastly, the positive correlation between high exposure to communicable diseases and high exposure to floods emphasizes the interconnected nature of environmental and health risks, necessitating integrated approaches to risk management. This suggests that areas prone to flooding are also more susceptible to the spread of communicable diseases, indicating shared vulnerabilities and overlapping challenges. Addressing these risks requires coordinated efforts involving environmental resilience measures, such as flood mitigation and infrastructure improvements, together with public health interventions focused on disease prevention, surveillance, and response strategies.
- The findings emphasize the importance of adopting a holistic approach to resilience-building in flood-prone areas, which include not only healthcare system capacity but also individual and institutional factors. This implies the necessity of integrated strategies that address education, access to technology, social equity, and governance effectiveness alongside traditional healthcare measures.

This thesis highlights that the relationship between floods and health impacts is complex and cannot be predicted solely by examining exposure to diseases and flood events. The resilience of countries

to the impacts of floods on vector-borne diseases was analyzed in this study, revealing that it is influenced by various factors that determine vulnerability and contribute to resilience. The key finding of this thesis is that the differing impacts observed in the past can be explained by two main categories of indicators. First, a country's economic position enhances its ability to cope with disasters such as floods. However, when comparing countries with similar economic standings, various individual factors, including education, community context, and political engagement, emerged as significant determinants. This indicates that the health effects associated with vector-borne diseases can potentially be mitigated by addressing these individual factors. Therefore, it is crucial to further explore the role of individual resilience in disaster preparedness. By improving a country's vulnerability profile and enhancing its ability to manage flood events, the severe effects of communicable diseases could be reduced in the future.

Chapter 8

Discussion

8 Discussion

8.1 Introduction

This study investigated how various factors of healthcare systems affect health outcomes for communicable diseases during flood disasters. The primary objective is to identify key indicators of health that explain the varying impacts of floods on communicable diseases, highlight gaps in current knowledge, and suggest areas for future research, aiming to improve risk assessments for vulnerable countries. To achieve this, the study addressed several sub-questions, beginning with the construction of a conceptual framework. Subsequent investigations focused on identifying patterns and relationships between floods and the incidence rates of communicable diseases following such events. Additionally, correlations between indicators of healthcare systems and the likelihood of disease outbreaks post-flood were explored. Finally, a reflective analysis of the conceptual framework synthesized the study's key findings. Among the findings, this research emphasizes the vulnerability of countries such as Peru, Sri Lanka, and Pakistan, which appear less resilient to the health effects following flood disasters. It became evident that the impacts are influenced by several interconnected domains: socio-economic vulnerability, behavioral factors, healthcare quality and performance, institutional organization, and social and community context. Notably, while economical health and healthcare system resilience were identified as important contributing factors, differences observed between countries with comparable financial and healthcare system performance highlighted the importance of individual resilience within social, community, and institutional contexts.

In this discussion chapter, a methodological reflection will be presented to analyze the challenges and constraints encountered during the research process and assess their impact on the study's outcomes. Following this, the research limitations are presented. Lastly, the contribution to the knowledge gap and future research agenda will be discussed. The thesis will be finalized with a conclusion.

8.2 Methodological reflection

8.2.1 Exploratory data analysis

EDA was chosen for this research due to its flexibility, suitability for exploratory studies, and its ability to effectively manage limited sample sizes. EDA is proficient at identifying patterns and trends within data, providing insights into complex relationships without the need for formal hypothesis testing or model building. The primary objective of EDA is to explore and generate hypotheses rather than draw definitive conclusions, particularly when dealing with small sample sizes. EDA is a suitable method in the early stages of data analysis, helping to understand data structure, identify patterns, clean data, select features, and evaluate model assumptions. Although EDA lacks strict assumptions, it follows fundamental principles such as ensuring data quality and appropriate variable normalization. Careful interpretation of visualizations is crucial to avoid biases or misinterpretations. The iterative nature of EDA allows for revisiting and refining assumptions made during early exploration. While small sample sizes can provide valuable insights, caution is necessary due to limited variability, potential sampling bias, and reduced statistical power, which complicates the detection of significant effects. Potential biases might have been introduced as the conceptual framework was constructed as an input for the EDA, which involves making subjective decisions about which variables to analyze, which visualizations to use, and how to interpret the results. The conceptual framework was useful to guide the EDA but assumptions about the relationships being explored can introduce biases into the analysis.

8.2.2 Data sources

Understanding the data sources is essential in this research, requiring a thorough consideration of potential biases, limitations, and the contextual setting in which the data was generated. Particularly relevant is the research's sample size consisting primarily of LMICs, where strong monitoring systems are often lacking. During outbreaks, intensified monitoring efforts may lead to increased reported numbers, while the emergence of the COVID-19 pandemic might have further influenced data collection practices, potentially impacting the reported figures. Given the limited data availability, diverse sources were utilized, predominantly relying on surveillance systems for diseases like Malaria and Dengue, as well as ministries of health for waterborne diseases. Surveillance systems collect data from various sources including healthcare facilities, laboratories, and community reporting, aiming to provide timely and standardized information. However, these systems may unintentionally under report certain diseases or demographic groups, particularly in resource-constrained settings where healthcare access and reporting infrastructure are deficient. Conversely, data from ministries of health, consisting of official reports, epidemiological bulletins, and statistical publications, serve as essential resources. Nonetheless, such data may be susceptible to political influence, administrative delays, and reporting biases, potentially compromising the accuracy and timeliness of the reported numbers. Collecting sufficient data proved to be a significant challenge, leading to the decision to include all available data, despite the potential compromise in data quality, to ensure sufficient input for data analysis.

8.2.3 Selection of appropriate techniques

The literature suggests various methods for data exploration; however, due to constraints in time and available data, only a limited number of methods were used. These included time series plots, histograms, scatter plots, Pearson correlation, paired t-test and PCA. This subsection will briefly reflect on the choice for these techniques.

Time series plots

Time series plots were created to visualize trends, patterns, and seasonality, examining changes in variables over time. However, these plots may not effectively capture short-term fluctuations or irregularities, especially with large time intervals between data points. They also require careful consideration of seasonality and potential confounding factors that might influence observed patterns. In this research, yearly data should be interpreted with caution, as floods often occur suddenly and within short periods.

Statistical tests

Two statistical tests were conducted to assess the statistical significance of epidemiological data during flood and non-flood periods. At first, the Pearson correlation was used to analyze yearly available data. While this method of testing correlation does not strictly require the dataset to be normally distributed, more accurate results are obtained when the data approximates a normal distribution. Although used in conjunction with visual analysis and without drawing conclusions solely from its results, it is essential to acknowledge this limitation.

For the weekly and monthly available data, the paired t-test was utilized to evaluate statistical significance across different years. Unlike the Pearson correlation test, this method accounts for the onset and duration of a flood, without considering the number of affected individuals. However, it enables better observation of trends in the weeks or months following a flood. The results of the t-test reveal

several significant correlations, both positive and negative, in roughly equal proportions. This could be attributed to either the poor quality of the data or the possibility that floods may sometimes decrease the incidence of communicable diseases afterwards. Although this potential effect of floods was not further analyzed in the thesis, it may be relevant to include in future research.

PCA

PCA was selected to reduce the dimensionality of the dataset and condense this information into a smaller set of PCs. This decision was made because the research deals with a large number of variables associated with the determinants of health. This method allows gaining insight into underlying patterns and trends, explaining the interplay between various domains, which is valuable to this research. However, there are limitations to consider. Firstly, PCA assumes that the relationships between variables are linear. If this linearity assumption is violated, PCA may not accurately capture the underlying structure of the data. In this research, the linearity of the data has not been carefully considered, which raises questions about the appropriateness of this method. PCA is sensitive to outliers, which can disproportionately influence the calculation of PCs, leading to biased results. In this study, a dilemma of balancing sample size against outliers was being faced. Consequently, retaining as much information as possible was prioritised rather than discarding outliers. Additionally, the results of PCA may vary depending on the specific sample of data used. Small alterations in the data or the inclusion/exclusion of certain observations can result in different PCs and loadings. Therefore, it is important to consider the stability of the results. As an addition to the selected methods, it is recommended to conduct sensitivity analyses to assess the robustness of the findings. While PCA is a common technique for dimensionality reduction, it is suggested to consider alternative methods such as factor analysis, independent component analysis, or nonlinear manifold learning methods like t-distributed stochastic neighbor embedding or Uniform Manifold Approximation and Projection. By comparing the results obtained through different methods, the consistency and validity of the findings can be evaluated.

8.3 Research limitations

8.3.1 Limitations in medical knowledge

The primary focus of this study is to use quantitative methods in evaluating the contribution of various factors of healthcare systems to the resilience of countries in the context of communicable diseases following flood disasters. Section 3.1.4 reviewed a selected number of past studies to understand the lag time in the context of floods. However, the understanding of transmission patterns of communicable diseases prior to this thesis was limited. Based on the literature review, assumptions have been made regarding the post-flood period necessary to detect a disease outbreak attributable to flooding. Consequently, there is a possibility that the results may not accurately reflect the aftermath of flood events. Further research is warranted to accurately track the transmission patterns of communicable diseases after floods and identify the main factors influencing disease uptake post-flooding.

8.3.2 Data

Flood data

The availability of flood data is crucial for this study, which uses quantitative methods to identify connections between floods and communicable diseases. While flood data is extensively documented in the EM-DAT database, it lacks detail. The first limitation concerns the scope of the research. Focusing on specific flood-prone areas within countries rather than entire countries could enhance the

analysis by incorporating more variables and improving the accuracy of exposure level assessments. Furthermore, the decision was made to filter flood data based on significance scores aimed at identifying the most impactful floods and reducing the sample size to manage the volume of information. This was determined by the number of affected people, total damage, and the duration of the flood. However, it should be acknowledged that these variables do not necessarily reflect floods with the greatest health impacts. For example, the total damage does not provide information about specific infrastructure damage that may not necessarily affect the risk of communicable diseases. Similarly, a long duration of a flood may complicate recovery processes but does not necessarily affect the risk of communicable diseases. Upon closer examination of the selected floods with sufficient significance, it is observed that they almost always have the highest number of affected people. Therefore, the inclusion of other factors may not significantly impact the final selection of floods to be included. Since the variables are not clearly defined, there may be debate about whether they should be considered in evaluating the significance of floods.

Health data

Regarding health data, the available morbidity and mortality data are insufficient to draw strong conclusions. Despite the EDA, including visual and statistical analyses, indicates a lack of resilience in Peru and Sri Lanka, this conclusion requires careful consideration given data limitations and methodological nuances. Insufficient yearly data directly linking floods to disease outbreaks, variations in reporting system efficacy, and the impact of the COVID-19 pandemic on surveillance systems all require attention. Moreover, external factors such as weather, political unrest, and ecological changes are not factored into the evaluation of the potential connection between floods and communicable diseases.

Data availability and quality

The EDA aimed to capture relationships between floods and communicable diseases. However, it must be acknowledged that data availability is limited and of poor quality. The methods used in this research were adapted based on data availability. With access to comprehensive data, it would be advisable to expand time-series analysis by comparing data over an increased number of years. Additionally, the thesis demonstrates that monthly and weekly data enable yearly comparisons, considering seasonal patterns and 'waiting time' (incubation time) after floods. Enhanced data availability would facilitate a more detailed examination of the relationship across various diseases. Furthermore, incorporating spatial analysis to map the distribution of floods and disease outbreaks could help identify areas most susceptible to disease transmission following flooding events.

8.3.3 Selection of indicators

The healthcare systems have been assessed by compiling a set of indicators that describe these factors in the context of communicable diseases following floods. These indicators were selected based on the framework provided by the WHO, as described below.

"The social determinants of health (SDH) are the non-medical factors that influence health outcomes. They are the conditions in which people are born, grow, work, live, and age, and the wider set of forces and systems shaping the conditions of daily life. These forces and systems include economic policies and systems, development agendas, social norms, social policies and political systems." [World Health Organization, 2024b]

Upon reflection, it becomes apparent that the determinants of health encompass a wide range of potential factors. Despite the diversity of this concept, there is a lack of defined methodology for comprehensively addressing all these factors that impact health outcomes. Similarly, there is no consensus on the method for assessing the resilience of a country, particularly in the context of communicable diseases following floods. The set of indicators used to describe resilience was composed from various sources rather than being derived from a single, focused research effort in this area. Therefore, further research is needed to precisely define resilience in the specific context of communicable diseases after floods. Informed by existing literature, review of conceptual models, and data availability, relevant indicators were incorporated into the analysis. However, it is important to acknowledge that this analysis may not capture every factor influencing the resilience to floods.

Another aspect to consider is the use of index numbers in this research. As data were not always available at the desired level, primarily in the resilience category, index numbers were used to describe performance in terms of government effectiveness, personal safety and rights, and the performance of the healthcare system. Furthermore, the MDPI and GII were included as indicators for the PCA analysis. However, these index numbers are sometimes a combination of several other indicators or are based on other types of research. These indicators should be considered with extra care as they could lead to double counting when indicators used in the index are presented separately. For example, the MDPI includes indicators such as years of schooling, child mortality, and nutrition, which are also represented by other indicators. The decision was made to include it anyway because it also represents indicators related to housing and assets for which other indicators were difficult to find. Furthermore, it has not always been considered what the source of the information is that constructed these indexes, and questions could be raised about the quality and reliability.

8.3.4 PCA

In this study, several limitations associated with the application of PCA were encountered. First of all, the research includes a limited number of countries, posing a challenge to the interpretation of the results. Increasing the sample size would enhance the credibility of the findings. Future investigations should aim to incorporate more countries, allowing for better categorization of similar countries and a more detailed examination of the differences between their healthcare systems. Secondly, several indicators used in the analysis were affected by missing data, particularly at the country-wide level. For example, factors such as the quality of a neighborhood, level of community feeling, and type of housing might influence resilience but cannot be considered on a country level. Focusing specifically on flood-prone areas may provide more informative insights, although obtaining such data can be challenging. Lastly, the presence of multicollinearity among the indicators included in the PCA raises concerns about the reliability of the results. While some degree of correlation is expected, excessively high levels of multicollinearity can distort the findings. Therefore, it is crucial to assess the extent of multicollinearity in the dataset and carefully evaluate the suitability of the dataset for PCA analysis. A methodological decision was made to remove indicators with a correlation higher than 0.85. However, correlation between 0.5 and 0.85 can still be considered a moderate correlation and thus it could be debated whether this chosen correlation level was optimal.

8.3.5 Validation

The research proposal included the plan to conduct an expert validation to assess the final framework alongside the findings from the EDA. However, as the research progressed, it became evident that

the focus was shifting from proposing a definitive framework to explaining the effects of health determinants on the flood-health risk of countries, to uncovering potential interrelations and providing guidance for further research. Consequently, expert validation no longer appeared relevant, and it was decided to exclude it from this study. Nevertheless, validation remains a crucial aspect of research, especially in exploratory studies like this one. Therefore, alternative methods of validation should be considered in future research. One approach is cross-validation, which involves using different statistical methods or datasets to confirm the robustness of the results. Additionally, qualitative research, such as conducting focus groups with individuals experienced in flooding or performing case studies, could offer deeper insights into the underlying mechanisms driving the relationships observed in the quantitative analysis.

8.4 Theoretical implication: Contribution to knowledge gap

The literature review has identified a significant knowledge gap concerning the correlation between the flood-disease risk in countries and healthcare system factors, including social, economic, institutional, and geographical characteristics. While previous research has provided evaluations and country-specific recommendations for mitigating health effects post-disaster, these recommendations often lack generalizability and fail to consider the complex interplay of social, economic, institutional, and geographical factors. By focusing on healthcare systems through the perspective of social determinants of health, this study considers the context of health risks associated with communicable diseases post-flood.

The primary objective of this study was to provide a more comprehensive understanding of the underlying mechanisms driving health outcomes in various settings. Its main contribution lies in generating knowledge and awareness that health outcomes are not solely determined by the scale of a flood or the performance of the healthcare system. Instead, by considering the system from a broader perspective, through the lens of the determinants of health, connections between underlying factors and mechanisms driving health outcomes were observed. This emphasizes the importance of careful risk assessment to identify potential areas for improvement in a country's preparedness. The study identifies potential areas for improvement in countries' preparedness and response capacities. In contrast to past studies, this research conducted a comparative analysis, providing insights into both similarities and differences in their determinants of health and identifying individual resilience factors as the main determining factors to explain variances in resilience to floods. The findings of this study suggest that by enhancing factors contributing to the individual resilience of communities, such as education, the use of communication tools, community context, and individual political engagement, the vulnerability of individuals can be reduced, and health outcomes may be mitigated. Additionally, improving the financial health of countries and the performance and quality of healthcare systems can be enhanced, increasing their capacities to cope with the health impacts of floods.

8.5 Practical implications: Remaining challenges and further research

The results of this study, which explored the relationship between floods, communicable diseases, and healthcare systems as illustrated in Figure 23, highlight the importance of several variables in evaluating healthcare system resilience in the context of floods. However, this study has also provided insights into possible directions for further research to explore this complex relationship. In this chapter, a future research agenda is proposed. In this section, the suggestions for further research are discussed and then presented in Figure 24.

Disease outbreak following floods

Past research, including this thesis, has highlighted the challenges in establishing correlations between floods and communicable diseases. These challenges arise from two primary factors. Firstly, there is a limitation in the amount and quality of available data. Additionally, numerous variables exist that could potentially influence disease transmission. To address these challenges, several research directions are suggested. Firstly, exploration of strategies to encourage governments to enhance monitoring and surveillance efforts is warranted, thereby enabling more precise measurement of incidence rates. Secondly, narrowing the research focus to flooded areas, rather than conducting country-wide analyses, could help reduce the impact of external factors. Furthermore, suspected disease outbreaks following floods should be tested alongside other variables, such as weather data and political situations like civil unrest, to ascertain a stable environment for analysis. Lastly, while this research assumes that the number of affected individuals correlates with the potential for disease transmission, future investigations should explore other variables that may influence the impact of floods on disease outbreaks. These could include factors such as damage to healthcare facilities, the number of flooded homes, and the displacement of populations.

Expand selection of diseases

The selection of diseases studied in this research was based on data availability and prior research. However, it did not account for potential gaps in reporting for different diseases. The initial search for epidemiological data, as shown in Appendix A, identified several other diseases mentioned in research related to communicable diseases after floods. These include, among others, Typhoid fever, Rotavirus, Pink eye, Conjunctivitis, and Zika virus. Moreover, diseases like COVID-19, which are relatively recent, have been monitored for only a short period. Therefore, it is advisable to critically assess whether this accurately reflects the true burden of disease following floods. Future research should broaden the selection of diseases to encompass other types of communicable diseases, thus providing a more comprehensive understanding of the actual situation.

Resilience

The analysis carried out in this research relies on indicators that describe the resilience of countries concerning communicable diseases following floods. However, while this set of indicators is informed by past research, a clear definition of resilience has not been established. To enhance the accuracy of identifying determinants of health that correlate with flood-disease risk, several areas for further research are proposed. Firstly, there is a need to establish a specific definition of resilience within the context of communicable diseases after floods. Secondly, the assessment of flood exposure is currently limited to a narrow set of indicators, not adequately capturing the full extent of exposure. This limitation arises primarily from the focus of this study on a country-level analysis, overlooking variations within countries. Thus, conducting a study that evaluates the exposure levels of flood-prone areas is recommended. Lastly, while the resilience of healthcare systems is evaluated, it does not fully

reflect their preparedness for emergencies due to data limitations. Future research should incorporate additional indicators to comprehensively evaluate resilience, particularly concerning emergency preparedness.

Domain interrelations

Key findings from the EDA indicate various factors contributing to health that potentially explain variances in the level of health impact a country faces after floods. A primary observation includes the positive linear relationship between environmental flood exposure and exposure to vector-borne diseases, but the underlying mechanisms remain inadequately understood. Secondly, concerning gaps in healthcare system resilience versus behavioral vulnerability, the results suggest that despite poor healthcare system resilience, some populations exhibit resilience due to high behavioral factor scores. This suggests that by building strong individual resilience, countries with limited financial and institutional capabilities can still improve coping capacities through behavioral interventions. For example, by implementing behavioral training programs, conducting community resilience assessments, and engaging communities in the development and implementation of resilience-building strategies. Future research should involve an extensive assessment of community resilience at the individual level and translating these research insights into actionable policies and interventions that can mitigate the health risks associated with environmental floods and enhance community resilience against future challenges. Lastly, gaps exist in the examination of state legitimacy, a broad concept assessing public confidence in state institutions and processes, and its effects on health emergency response policies and prevention awareness policies concerning floods and communicable diseases. Investigating how the level of state legitimacy might impact such policies is crucial for informing effective governance strategies in emergency contexts.

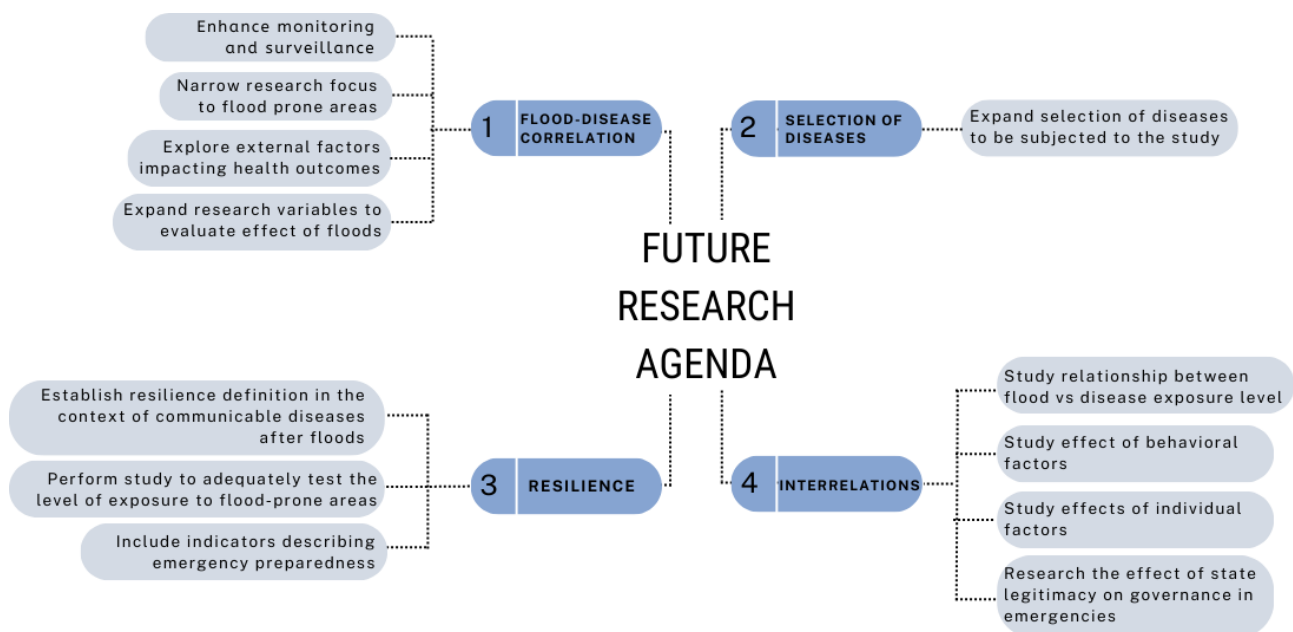


Figure 24: Future research agenda

8.6 Conclusion

This thesis explored the complex dynamics between flood disasters, healthcare systems, and communicable diseases, aiming to broaden the understanding of the relationship between flood disasters and the impacts of communicable diseases across diverse settings, investigating characteristics in 14 flood-prone countries worldwide. To answer this question, a quantitative approach was chosen to measure the extent and impact of vulnerability and resilience factors on the impacts of communicable diseases after floods. While recognizing the varied nature of resilience in the face of floods, the research highlights a crucial finding: the substantial impact of individual factors, particularly those related to human capital development, on community resilience.

Despite acknowledging the complexity of resilience, the study highlights that certain individual factors demonstrate variances between countries with similar levels of economic and healthcare system resilience. Notably, factors such as education level, individual autonomy, and access to safety and political freedom emerge as significant determinants of resilience. This suggests that by strengthening these individual factors, countries can significantly enhance their capacity to withstand and recover from the health impacts of floods. The implication of this finding is significant, especially for LMICs, which are often more vulnerable to the health risks posed by floods. While economic and healthcare system resilience remain fundamental, the research emphasizes that countries worldwide have unique opportunities to enhance their resilience by prioritizing investments in individual factors.

As climate change escalates the frequency and intensity of floods, understanding and enhancing resilience becomes essential. This deeper understanding of resilience has significant implications for risk assessment strategies. Recognizing the important role of individual factors in shaping community resilience allows countries to refine their approaches to assessing flood-related risks. Incorporating insights from this research into risk assessment methodologies can provide a more thorough view of a country's resilience capacity. This research emphasizes that strengthening healthcare systems alone is insufficient; it also highlights the importance of building individual resilience through targeted interventions. Empowering individuals with the necessary knowledge, autonomy, and resources enables countries to effectively mitigate the harmful health effects of floods and promote healthier, more resilient communities in the face of environmental uncertainties. Integrating a focus on these individual resilience factors into risk assessment processes enables policymakers and disaster management authorities to identify specific community vulnerabilities and strengths. This understanding enables the development of focused mitigation measures adapted to address the unique challenges posed by communicable diseases following floods. Such measures may include community-based education and empowerment initiatives, ensuring access to essential healthcare services, and interventions aimed at enhancing social cohesion and support networks.

In conclusion, as the frequency and intensity of floods increase, the challenges posed by flood-induced health outcomes become more evident. However, within these environmental uncertainties, this thesis highlights a promising approach to resilience. By recognizing and using the power of individual factors, particularly in LMICs, countries can effectively address the complexities of flood disasters. While floods cannot be prevented in the future, prioritizing research efforts and the development of preparedness strategies becomes important to mitigate their worst impacts. Building healthier and more resilient communities, prepared to withstand and recover from environmental challenges, remains an urgent priority.

Bibliography

- [Abaya et al., 2009] Abaya, S. W., Mandere, N., and Ewald, G. (2009). Floods and health in Gambella region, Ethiopia: A qualitative assessment of the strengths and weaknesses of coping mechanisms. *Global Health Action*, 2(1).
- [Akanda et al., 2009] Akanda, A. S., Jutla, A. S., and Islam, S. (2009). Dual peak cholera transmission in Bengal Delta: A hydroclimatological explanation. *Geophysical Research Letters*, 36(19).
- [Alcayna et al., 2022] Alcayna, T., Fletcher, I., Gibb, R., Tremblay, L., Funk, S., Rao, B., and Lowe, R. (2022). Climate-sensitive disease outbreaks in the aftermath of extreme climatic events: A scoping review.
- [Alderman et al., 2012] Alderman, K., Turner, L. R., and Tong, S. (2012). Floods and human health: A systematic review. *Environment International*, 47:37–47.
- [Ansari et al., 2003] Ansari, Z., Carson, N. J., Ackland, M. J., Vaughan, L., and Serraglio, A. (2003). A public health model of the social determinants of health. *Sozial- und Präventivmedizin*, 48(4):242–251.
- [Arnold and Wade, 2015] Arnold, R. D. and Wade, J. P. (2015). A definition of systems thinking: A systems approach. In *Procedia Computer Science*, volume 44, pages 669–678. Elsevier.
- [Asselman et al., 2022] Asselman, N., van Heeringen, K.-J., de Jong, J., and Geertsema, T. (2022). Juli 2021 overstrooming en wateroverlast in zuid-limburg: Eerste bevindingen voor valkenburg, geulmonding, roermonding en eygelshoven.
- [Basaria et al., 2023] Basaria, A. A., Ahsan, A., Nadeem, A., Tariq, R., and Raufi, N. (2023). Infectious diseases following hydrometeorological disasters: Current scenario, prevention, and control measures. *Annals of Medicine and Surgery*, 85(8):3778–3782.
- [Bolan et al., 2024] Bolan, S., Padhye, L. P., Jasemizad, T., Govarthanan, M., Karmegam, N., Wijesekara, H., Amarasiri, D., Hou, D., Zhou, P., Biswal, B. K., Balasubramanian, R., Wang, H., Siddique, K. H., Rinklebe, J., Kirkham, M. B., and Bolan, N. (2024). Impacts of climate change on the fate of contaminants through extreme weather events.
- [Braeken and Van Assen, 2017] Braeken, J. and Van Assen, M. A. (2017). An empirical Kaiser criterion. *Psychological Methods*, 22(3):450–466.
- [Brown and Murray, 2013] Brown, L. and Murray, V. (2013). Examining the relationship between infectious diseases and flooding in Europe. *Disaster Health*, 1(2):117–127.
- [Campbell and Stanley, 1966] Campbell, D. and Stanley, J. (1966). *Experimental and Quasi-experimental Designs for Research*. R. McNally.
- [Centre for Research on the Epidemiology of Disasters, 2024] Centre for Research on the Epidemiology of Disasters (2024). EM-DAT - the international disaster database. <https://www.emdat.be/>.
- [Chadsuthi et al., 2021] Chadsuthi, S., Chalvet-Monfray, K., Wiratsudakul, A., and Modchang, C. (2021). The effects of flooding and weather conditions on leptospirosis transmission in Thailand. *Scientific Reports*, 11(1).

- [Climate Council, 2022] Climate Council (2022). Everything you need to know about floods and climate change: Explainer. <https://www.climatecouncil.org.au/resources/climate-change-floods/>.
- [Coalson et al., 2021] Coalson, J. E., Anderson, E. J., Santos, E. M., Garcia, V. M., Romine, J. K., Dominguez, B., Richard, D. M., Little, A. C., Hayden, M. H., and Ernst, K. C. (2021). The complex epidemiological relationship between flooding events and human outbreaks of mosquito-borne diseases: A scoping review.
- [Cook, 2015] Cook, R. D. (2015). Linear hypothesis: Regression (graphics). In Wright, J. D., editor, *International Encyclopedia of the Social Behavioral Sciences (Second Edition)*, pages 157–161. Elsevier, Oxford, second edition edition.
- [Creswell and Creswell, 2023] Creswell, J. W. and Creswell, J. D. (2023). *Research design: Qualitative, quantitative, and mixed methods approaches*. SAGE.
- [Cutter et al., 2014] Cutter, S. L., Ash, K. D., and Emrich, C. T. (2014). The geographies of community disaster resilience. *Global Environmental Change*, 29:65–77.
- [Doocy et al., 2013] Doocy, S., Daniels, A., Murray, S., and Kirsch, T. D. (2013). The human impact of floods: A historical review of events 1980-2009 and systematic literature review. *PLoS Currents*.
- [Du et al., 2010] Du, W., FitzGerald, G. J., Clark, M., and Hou, X.-Y. (2010). Health impacts of floods. *Prehospital and disaster medicine*, 25(3):265–272.
- [Elimian et al., 2020] Elimian, K. O., Mezue, S., Musah, A., Oyebanji, O., Fall, I. S., Yennan, S., Yao, M., Abok, P. O., Williams, N., Omar, L. H., Balde, T., Ampah, K., Okudo, I., Ibrahim, L., Jinadu, A., Alemu, W., Peter, C., and Ihekweazu, C. (2020). What are the drivers of recurrent cholera transmission in Nigeria? Evidence from a scoping review. *BMC Public Health*, 20(1).
- [Endendijk et al., 2023] Endendijk, T., Botzen, W., Moel, H., Aerts, J., Duijndam, S., Slager, K., Kolen, B., and Kok, M. (2023). Experience from the 2021 floods in the netherlands: Household survey results on impacts and responses.
- [Englande, 2008] Englande, A. J. (2008). Katrina and the Thai Tsunami-Water Quality and Public Health Aspects Mitigation and Research Needs. Technical Report 5.
- [Erickson et al., 2019] Erickson, T. B., Brooks, J., Nilles, E. J., Pham, P. N., and Vinck, P. (2019). Environmental health effects attributed to toxic and infectious agents following hurricanes, cyclones, flash floods and major hydrometeorological events.
- [European Commission, 2023] European Commission (2023). Facing increasing river flood risk in europe: Adaptation measures can save lives and billions of euro. <https://bit.ly/4c5B96Y>.
- [European Commission and Joint Research Center, 2024] European Commission and Joint Research Center (2024). Home. <https://drmkc.jrc.ec.europa.eu/inform-index>.
- [Few, 2007] Few, R. (2007). Health and climatic hazards: Framing social research on vulnerability, response and adaptation. *Global Environmental Change*, 17(2):281–295.
- [Frenk and Murray, 2000] Frenk, J. and Murray, C. J. L. (2000). Article in Bulletin of the World Health Organization.

- [Frost, 1976] Frost, W. H. (1976). Some conceptions of epidemics in general1. *American Journal of Epidemiology*, 103(2):141–151.
- [Gonzalez-Daza et al., 2023] Gonzalez-Daza, W., Vivero-Gómez, R. J., Altamiranda-Saavedra, M., Muylaert, R. L., and Landeiro, V. L. (2023). Time lag effect on malaria transmission dynamics in an Amazonian Colombian municipality and importance for early warning systems. *Scientific Reports*, 13(1).
- [Hashizume et al., 2013] Hashizume, M., Chaves, L. F., Faruque, A. S., Yunus, M., Streatfield, K., and Moji, K. (2013). A Differential Effect of Indian Ocean Dipole and El Niño on Cholera Dynamics in Bangladesh. *PLoS ONE*, 8(3).
- [Heo et al., 2008] Heo, J.-H., Kim, M.-H., Koh, S.-B., Noh, S., Park, J.-H., Ahn, J.-S., Park, K.-C., Shin, J., and Min, S. (2008). A Prospective Study on Changes in Health Status Following Flood Disaster Correspondence. *Psychiatry Invest*, 5:186–192.
- [Ikotun et al., 2023] Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., and Heming, J. (2023). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, 622:178–210.
- [Jolliffe and Cadima, 2016] Jolliffe, I. T. and Cadima, J. (2016). Principal component analysis: A review and recent developments.
- [Jorgensen and Hansen, 2012] Jorgensen, K. W. and Hansen, L. K. (2012). Model selection for gaussian kernel PCA denoising. *IEEE Transactions on Neural Networks and Learning Systems*, 23(1):163–168.
- [Khan et al., 2023] Khan, M. B., Yang, Z. S., Lin, C. Y., Hsu, M. C., Urbina, A. N., Assavalapsakul, W., Wang, W. H., Chen, Y. H., and Wang, S. F. (2023). Dengue overview: An updated systemic review.
- [Khan et al., 2018] Khan, Y., O’Sullivan, T., Brown, A., Tracey, S., Gibson, J., Génèreux, M., Henry, B., and Schwartz, B. (2018). Public health emergency preparedness: A framework to promote resilience. *BMC Public Health*, 18(1).
- [Kouadio et al., 2012] Kouadio, I. K., Aljunid, S., Kamigaki, T., Hammad, K., and Oshitani, H. (2012). Infectious diseases following natural disasters: Prevention and control measures.
- [Kumar, 2010] Kumar, U. (2010). *From Surviving to Thriving: Understanding the needs of a healthy life*.
- [Lee et al., 2020] Lee, J., Perera, D., Glickman, T., and Taing, L. (2020). Water-related disasters and their health impacts: A global review.
- [Liang and Messenger, 2018] Liang, S. Y. and Messenger, N. (2018). Infectious Diseases After Hydrologic Disasters.
- [Ligon, 2006] Ligon, B. L. (2006). Infectious diseases that pose specific challenges after natural disasters: A review. *Seminars in Pediatric Infectious Diseases*, 17(1):36–45.

- [Lock et al., 2012] Lock, S., Rubin, G. J., Murray, V., Rogers, M. B., Amlôt, R., and Williams, R. (2012). Secondary stressors and extreme events and disasters: A systematic review of primary research from 2010-2011.
- [Luijken et al., 2022] Luijken, K., Dekkers, O. M., Rosendaal, F. R., and Groenwold, R. H. (2022). Exploratory analyses in aetiologic research and considerations for assessment of credibility: mini-review of literature.
- [Lutz et al., 2013] Lutz, C., Erken, M., Noorian, P., Sun, S., and McDougald, D. (2013). Environmental reservoirs and mechanisms of persistence of *Vibrio cholerae*.
- [Maharrani et al., 2024] Maharrani, R. H., Abda'u, P. D., and Faiz, M. N. (2024). Clustering method for criminal crime acts using K-means and principal component analysis. *Indonesian Journal of Electrical Engineering and Computer Science*, 34(1):224–232.
- [Mallakpour and Villarini, 2015] Mallakpour, I. and Villarini, G. (2015). The changing nature of flooding across the central United States. *Nature Climate Change*, 5(3):250–254.
- [Mark et al., 2015] Mark, O., Jørgensen, C., Hammond, M., Khan, D., Tjener, R., Erichsen, A., and Helwigh, B. (2015). A new methodology for modelling of health risk from urban flooding exemplified by cholera - case dhaka, bangladesh. *Journal of Flood Risk Management*, 11.
- [Matthieu Komorowski et al., 2016] Matthieu Komorowski, Dominic C. Marshall, Justin D. Salci-ccioli, and Yves Crutain. (2016). *Secondary Analysis of Electronic Health Records*. Springer International Publishing.
- [Mohajervatan et al., 2023] Mohajervatan, A., Tavakoli, N., Khankeh, H., Raeisi, A. R., and Atighechian, G. (2023). Identifying flood response experiences in Iranian health system: A qualitative study. *Journal of Flood Risk Management*, 16(4).
- [Morin et al., 2013] Morin, C. W., Comrie, A. C., and Ernst, K. (2013). Climate and dengue transmission: Evidence and implications.
- [National Severe Storms Laboratory, 2023] National Severe Storms Laboratory (2023). Flood types. Available at <https://www.nssl.noaa.gov/education/svrwx101/floods/types/>.
- [Ndicunguye, 2012] Ndicunguye, R. (2012). Health Risks in Flood Prone Areas of Kigali City: A case study of Nyabugogo floodplain area.
- [Nederlandse Omroep Stichting, 2021] Nederlandse Omroep Stichting (2021). Overstromingen in limburg en buurlanden op één na duurste natuurramp van 2021. <https://bit.ly/4b1CJp4>.
- [Niven et al., 2012] Niven, D. J., Berthiaume, L. R., Fick, G. H., and Laupland, K. B. (2012). Matched case-control studies: A review of reported statistical methodology.
- [Ochani et al., 2022] Ochani, S., Aaqil, S. I., Nazir, A., Athar, F. B., Ochani, K., and Ullah, K. (2022). Various health-related challenges amidst recent floods in Pakistan; strategies for future prevention and control. *Annals of Medicine and Surgery*, 82.
- [Okaka and Odhiambo, 2018] Okaka, F. O. and Odhiambo, B. D. (2018). Relationship between flooding and out break of infectious diseasesin Kenya: A review of the literature. *Journal of Environmental and Public Health*, 2018.

- [Olanrewaju et al., 2019] Olanrewaju, C. C., Chitakira, M., Olanrewaju, O. A., Louw, E., and Olanrewaju, C. (2019). Jàmbá-Journal of Disaster Risk Studies.
- [Pal et al., 2016] Pal, S., Juyal, D., Sharma, M., Kotian, S., Negi, V., and Sharma, N. (2016). An outbreak of hepatitis A virus among children in a flood rescue camp: A post-disaster catastrophe. *Indian Journal of Medical Microbiology*, 34(2):233–236.
- [Parry et al., 2007] Parry, M., Canziani, O., and Palutikof, J. (2007). Ipcc climate change. impacts, adaptation and vulnerability.
- [Paterson et al., 2018] Paterson, D. L., Wright, H., and Harris, P. N. (2018). Health risks of flood disasters.
- [Pearson, 1931] Pearson, E. S. (1931). The Analysis of Variance in Cases of Non-Normal Variation. Technical Report 1.
- [Phalkey and Louis, 2016] Phalkey, R. K. and Louis, V. R. (2016). Two hot to handle: How do we manage the simultaneous impacts of climate change and natural disasters on human health?
- [PreventionWeb, 2017] PreventionWeb (2017). In next decades, frequency of coastal flooding will double globally. <https://www.preventionweb.net/news/next-decades-frequency-coastal-flooding-will-double-globally>.
- [Ramin and McMichael, 2009] Ramin, B. M. and McMichael, A. J. (2009). Climate change and health in sub-saharan africa: A case-based perspective. *EcoHealth*, 6(1):52–57.
- [Ray-Bennett et al., 2019] Ray-Bennett, N. S., Corsel, D. M. J., Goswami, N., and Ghosh, A. (2019). Understanding reproductive health challenges during a flood: insights from Belkuchi Upazila, Bangladesh. *Gates Open Research*, 3:788.
- [Red Cross, 2023] Red Cross, R. (2023). Pakistan floods and climate change. <https://www.redcross.org.uk/stories/disasters-and-emergencies/world/climate-change-and-pakistan-flooding-affecting-millions>.
- [Rentschler et al., 2022] Rentschler, J., Salhab, M., and Jafino, B. A. (2022). Flood exposure and poverty in 188 countries. *Nature Communications*, 13(1).
- [Roland et al., 2021] Roland, H. B., Lee, D., Curtis, C. D., Malecki, K. J., Brossard, K., Block, D., Roland, H., Lee, D., Wirz, C., Curtis, K., Malecki, K., Brossard, D., Block, P., and Lois, A. N. (2021). Stakeholders' Perspectives on Flood Risk and Vulnerability in Peru Please Cite As. Technical report.
- [Salami et al., 2017] Salami, R. O., von Meding, J. K., and Giggins, H. (2017). Urban settlements' vulnerability to flood risks in african cities: A conceptual framework. *Jamba: Journal of Disaster Risk Studies*, 9(1).
- [Salas, 2023] Salas, E. B. (2023). Countries most exposed to floods worldwide 2023. <https://www.statista.com/statistics/1306264/countries-most-exposed-to-floods-by-risk-index-score/>.

- [Sarkar, 2020] Sarkar, S. (2020). Flash floods surge in himalayas due to extreme rain. <https://indiaclimatedialogue.net/2020/09/11/flash-floods-surge-in-himalayas-due-to-extreme-rain/>.
- [Savi, 2022] Savi, M. K. (2022). An Overview of Malaria Transmission Mechanisms, Control, and Modeling. *Medical Sciences*, 11(1):3.
- [Schenker and Rumrill, 2004] Schenker, J. D. and Rumrill, P. D. (2004). Causal-comparative research designs. *Journal of Vocational Rehabilitation*, 21:117–121.
- [Schober and Schwarte, 2018] Schober, P. and Schwarte, L. A. (2018). Correlation coefficients: Appropriate use and interpretation. *Anesthesia and Analgesia*, 126(5):1763–1768.
- [Shackleton et al., 2023] Shackleton, D., Memon, F. A., Nichols, G., Phalkey, R., and Chen, A. S. (2023). Mechanisms of cholera transmission via environment in India and Bangladesh: state of the science review. *Reviews on Environmental Health*.
- [Shortus et al., 2016] Shortus, M., Musto, J., Bugoro, H., Butafa, C., Sio, A., and Joshua, C. (2016). Vector-control response in a post-flood disaster setting, Honiara, Solomon Islands, 2014. *Western Pacific surveillance and response journal : WPSAR*, 7(1):38–43.
- [Stephenson et al., 2014] Stephenson, J., Vaganay, M., Cameron, R., and Joseph, P. (2014). The long-term health impacts of repeated flood events. *Flood Recovery, Innovation and Response IV: Proverbs, D., Brebbia, CA, Eds*, pages 201–215.
- [Suwanpakdee et al., 2015] Suwanpakdee, S., Kaewkungwal, J., White, L. J., Asensio, N., Ratanakorn, P., Singhasivanon, P., Day, N. P., and Pan-Ngum, W. (2015). Spatio-temporal patterns of leptospirosis in Thailand: Is flooding a risk factor? *Epidemiology and Infection*, 143(10):2106–2115.
- [Tamene, 2016] Tamene, E. H. (2016). Theorizing conceptual framework. *Asian Journal of Educational Research*, 4(2).
- [Török, 2018] Török, I. (2018). Qualitative assessment of social vulnerability to flood hazards in Romania. *Sustainability (Switzerland)*, 10(10).
- [United Nations Children’s Fund, 2022] United Nations Children’s Fund (2022). Devastating floods in pakistan. <https://www.unicef.org/emergencies/devastating-floods-pakistan-2022>.
- [United Nations Development Programme, 2023] United Nations Development Programme (2023). 2023 global multidimensional poverty index (mpi). *UNDP (United Nations Development Programme)*.
- [United Nations Development Programme, 2024] United Nations Development Programme (2024). Gender inequality index. <https://hdr.undp.org/data-center/thematic-composite-indices/gender-inequality-index#/indicies/GII>.
- [Upadhyay, 2015] Upadhyay, P. (2015). Designing theoretical and conceptual frameworks. Technical report.

- [Viana and Ignotti, 2013] Viana, D. V. and Ignotti, E. (2013). A ocorrência da dengue e variações meteorológicas no brasil: Revisão sistemática. *Revista Brasileira de Epidemiologia*, 16(2):240–256.
- [Walika et al., 2023] Walika, M., Moitinho De Almeida, M., Castro Delgado, R., and Arcos González, P. (2023). Outbreaks following natural disasters: A review of the literature. *Disaster Medicine and Public Health Preparedness*, 17:e444.
- [Willa Dhany and Izhari, 2023] Willa Dhany, H. and Izhari, F. (2023). Journal of Intelligent Decision Support System (IDSS) Exploratory Data Analysis (EDA) methods for healthcare classification. Technical Report 4.
- [World Bank, 2023] World Bank (2023). Libya storm and flooding 2023. rapid damage and needs assessment.
- [World Health Organization, 2014] World Health Organization (2014). A global brief on vector-borne diseases. Technical report.
- [World Health Organization, 2024a] World Health Organization (2024a). Floods. https://www.who.int/health-topics/floods#tab=tab_1.
- [World Health Organization, 2024b] World Health Organization (2024b). Global health observatory. <https://www.who.int/data/gho>.
- [World Health Organization, 2024c] World Health Organization (2024c). Indicators index. <https://www.who.int/data/gho/data/indicators/indicators-index>.
- [World Health Organization, 2024d] World Health Organization (2024d). “it was just the perfect storm for malaria” – pakistan responds to surge in cases following the 2022 floods. <https://bit.ly/3VpZcpL>.
- [World Health Organization, 2023] World Health Organization, W. (2023). Situation report libya flood response.
- [World Health Organization, 2024] World Health Organization, W. (2024). Social determinants of health. <https://www.who.int/health-topics/social-determinants-of-health>.
- [Wu and Huang, 2022] Wu, Y. and Huang, C. (2022). Climate Change and Vector-Borne Diseases in China: A Review of Evidence and Implications for Risk Management. *Biology*, 11(3).

Appendices

A Epidemiological data

Table A.1: Significant outbreak of water-borne communicable diseases after flood disasters

Countries	Typhoid fever	Cholera	Lepto-spirosis	Hepatitis A	Other diarrhoeal disease
Bangladesh		x			x
Vietnam					
Cambodia		x	x		x
Iraq		x			x
Pakistan					x
China					x
India		x	x	x	x
Philippines	x		x		x
Sri Lanka			x		
Tanzania		x	x		x
Nigeria		x			x
Somalia		x			
South Sudan		x			x
Mozambique		x			x
Dominican Republic		x	x		
Peru					
Guyana			x		
Guatemala					x
Russia			x		
Serbia					
Czech			x		
Australia			x		

Table A.2: Significant reported outbreak of vector-borne communicable diseases after flood disasters + other reported communicable diseases

Countries	Malaria	Dengue	West Nile Fever	Other
Bangladesh		x		Rotavirus, Respiratory infection
Vietnam		x		Pink eye, Dermatitis, Conjunctivitis
Cambodia				Ear, nose and throat infections, Dermatitis and Conjunctivitis
Iraq				
Pakistan	x			Conjunctivitis, Respiratory infection
China	x			Schistosomiasis
India				Respiratory infections, Rotavirus, Chicken pox
Philippines		x		Respiratory infection, Chicken pox, Measles
Sri Lanka		x		
Tanzania	x			
Nigeria				Diphtheria
Somalia	x			Measles
South Sudan	x			Rift Valley fever, Pneumonia, Rheumatic fever, Measles, Covid, Polio
Mozambique	x			Respiratory infection
Dominican Republic	x	x		Zika
Peru		x		
Guyana				
Guatemala		x		Respiratory infection
Russia				Chikungunya, Zika, Respiratory infection, Tularemia
Serbia			x	
Czech			x	Tahyna virus
Australia				

B Overview of Reviewed Frameworks

A Social Determinants of Health

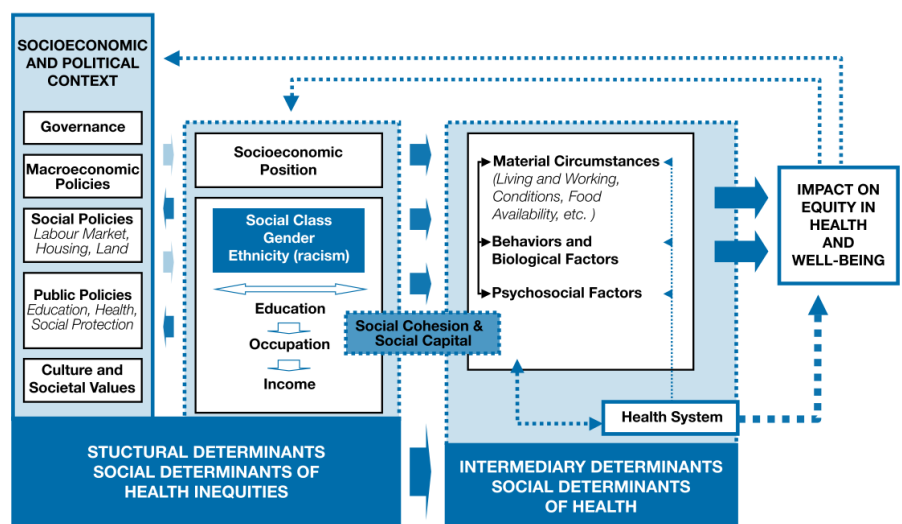


Figure B.1: Commission on Social Determinants of Health (CSDH) framework [Kumar, 2010]

B Health impact pathway for flooding

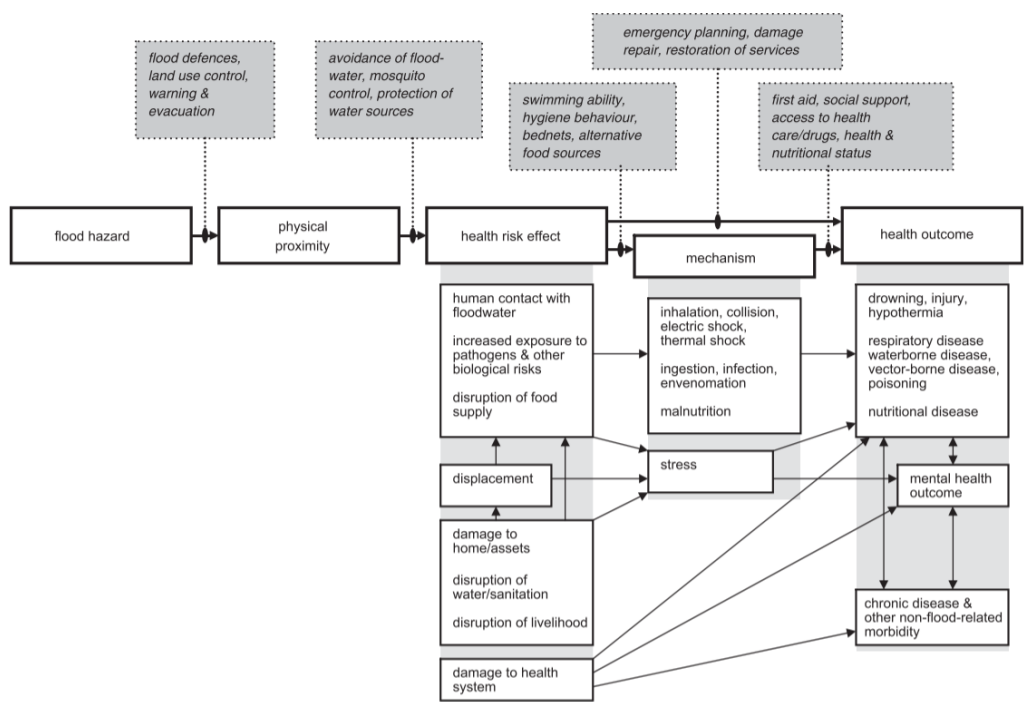


Figure B.2: Health Impact Pathway for flooding [Few, 2007]

C Flood vulnerability assessment framework

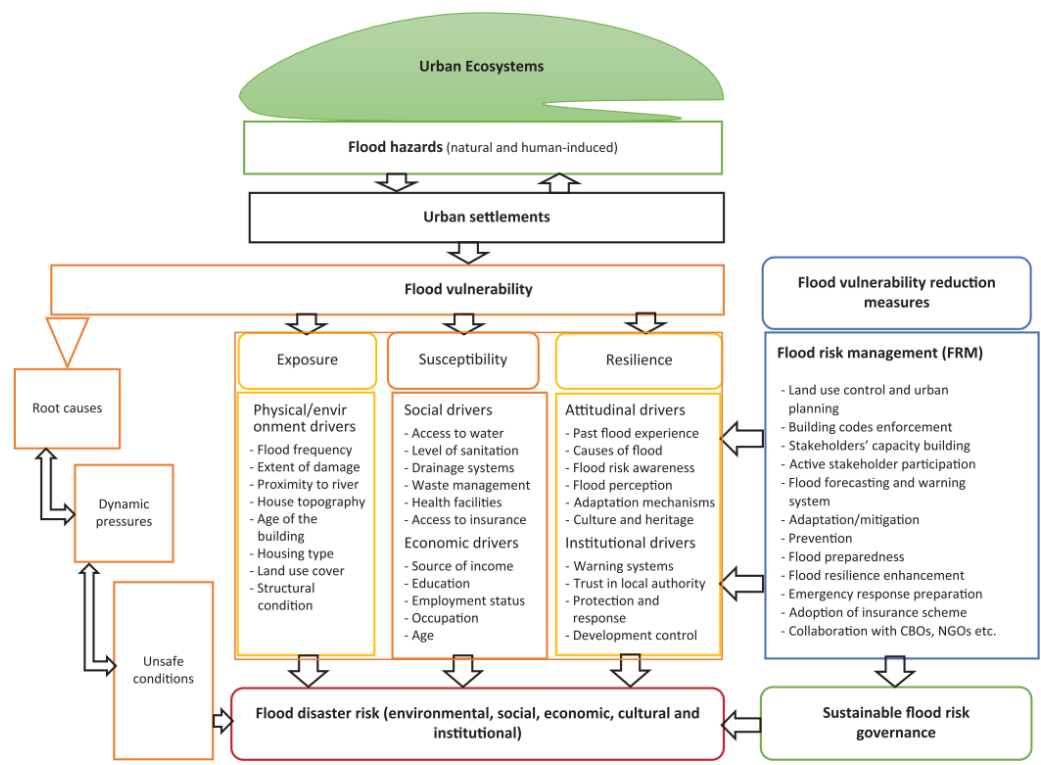


Figure B.3: Flood Vulnerability Assessment Framework [Salami et al., 2017]

D Public Health Emergency Preparedness Framework



Figure B.4: Public Health Emergency Preparedness

C Healthcare system indicators

Table C.3: Healthcare System indicators

Domain	Indicator Name	Explanation
Exposure	Annual Expected Exposed People to Coastal Floods	The projected number of people likely to be affected by coastal floods within a specified time period.
Exposure	Annual Expected Exposed People to River Floods	The projected number of people likely to be affected by river floods within a specified time period.
Exposure	Population Density	The number of people per unit area, typically measured in square kilometers.
Demographics	Population Living in Urban Areas	The proportion of the total population living in urban areas.
Exposure	Flood Frequency	The frequency of occurrence of floods within a specified area or time period.
Exposure	Epidemic Risk Index - Vector-borne	The level of risk posed by vector-borne diseases based on various factors such as climate, vector presence, and public health measures.
Exposure	Epidemic Risk Index - Waterborne	The level of risk posed by waterborne diseases based on factors such as water quality, sanitation, and healthcare infrastructure.
Material Circumstances	Quality of Water and Sanitation System	The condition and reliability of water supply and sanitation systems within a given area.
Material Circumstances	Housing Overcrowding Rate	The degree of overcrowding within housing units, typically measured by the number of people per room.
Material Circumstances	Land Use - Artificial Surfaces	The proportion of land area covered by artificial surfaces such as buildings, roads, and pavement.
Material Circumstances	People Using at Least Basic Sanitation Services (% of Population)	The percentage of the population with access to basic sanitation services, which ensure hygienic separation of human excreta from human contact.
Demographics	Household Size	The average number of people living in a household.
Economic	Current Health Expenditure per Capita	The average healthcare expenditure per person within a specific time frame.
Socio-economic	Multidimensional Poverty Index	An index measuring poverty based on various factors such as income, education, and access to basic services.
Socio-economic	Prevalence of Undernourishment	The percentage of the population experiencing undernourishment, indicating insufficient food intake to meet dietary energy requirements.
Socio-economic	Adult Literacy Rate, Population 15+ Years	The percentage of the population aged 15 years and older who are literate.
Socio-economic	GDP per Capita	The gross domestic product (GDP) per person, indicating the economic output per individual within a country.
Health System Resilience	Physicians Density	The number of physicians per unit area, typically measured in square kilometers.
Health System Resilience	Access to Healthcare Index	An index measuring the overall accessibility of healthcare services within a region.
Institutional Resilience	Government Effectiveness	The effectiveness of government institutions and their ability to implement policies and provide services to the population.

Table C.3: Continued

Domain	Indicator Name	Explanation
Institutional Resilience	Corruption Perception Index	The perceived level of corruption within government institutions and society as a whole.
Institutional Resilience	State Legitimacy Index	The perceived legitimacy and authority of the state government among the population.
Institutional Resilience	Public Services Index	The quality and accessibility of public services provided by the government.
Individual Resilience	Individuals Using the Internet (% of Population)	The percentage of the population with access to and usage of the internet.
Individual Resilience	Happiness Score	A measure of subjective well-being and life satisfaction within a population.
Demographics	Urban Population	The proportion of the total population living in urban areas.
Demographics	Median Age	The age that separates the younger half of the population from the older half.
Demographics	Gender	The distribution of individuals based on their gender within a population.
Socio-economic	Mean Income	The average income earned by individuals within a specified population or geographic area.
Socio-economic	Educational Attainment, at Least Completed Primary, Population 25+ Years	The percentage of the population aged 25 years and older who have completed at least primary education.
Socio-economic	Literacy Rate	The percentage of the population that can read and write at a specified age.
Socio-economic	Employment to Population Ratio, 15+	The proportion of the working-age population (15 years and older) that is employed.
Socio-economic	Prevalence of Undernourishment	The percentage of the population experiencing undernourishment, indicating insufficient food intake to meet dietary energy requirements.
Socio-economic	Prevalence of Moderate or Severe Food Insecurity in the Population	The percentage of the population experiencing moderate to severe levels of food insecurity, indicating inadequate access to food due to financial constraints or other factors.
Socio-economic	Crime Index	A measure of the level of crime and safety within a given area, typically based on reported crime rates and perceptions of safety.
Health System Resilience	Medical Doctors	The number of medical doctors available per unit area or population size.
Health System Resilience	Hospital Beds	The number of beds available in hospitals for patient care.
Health System Resilience	Healthcare Access and Quality Index	An index measuring the overall accessibility and quality of healthcare services within a region.
Health System Resilience	Percentage of Population Covered by Insurance	The proportion of the population covered by health insurance plans.
Health System Resilience	Out-of-Pocket Expenditure per Capita	The average healthcare expenses paid directly by individuals at the point of service per person within a specified time frame.
Individual Resilience	Individuals Using the Internet (% of Population)	The percentage of the population with access to and usage of the internet.

Table C.3: Continued

Domain	Indicator Name	Explanation
Health System Resilience	Health Capacity in Clinics, Hospitals, and Community Care Centers	The capacity of healthcare facilities, including clinics, hospitals, and community care centers, to provide medical services and support to the population.
Health System Resilience	Supply Chain for Health System and Healthcare Workers	The efficiency and effectiveness of the supply chain management system for healthcare-related goods and services.
Health System Resilience	Healthcare Access	The ease of access to healthcare services, including barriers such as geographical distance, financial constraints, and availability of services.
Health System Resilience	Prevention	Measures taken to prevent the occurrence or spread of diseases and health-related problems within a population.
Health System Resilience	Detection and Reporting	Systems and processes in place to detect, monitor, and report health-related events and diseases within a population.
Health System Resilience	Rapid Response	The ability of the healthcare system to respond promptly and effectively to health emergencies and crises.
Health System Resilience	Risk Environment	The conditions and factors within the environment that contribute to health risks and vulnerabilities within a population.
Economic	Health Spending US\$ per Capita	Total health spending per capita
Individual Resilience	Political Rights	The extent to which individuals within a society enjoy political rights such as freedom of speech, assembly, and participation in political processes.
Individual Resilience	Freedom of Peaceful Assembly	The degree to which individuals are free to assemble peacefully without interference or repression from authorities.
Individual Resilience	Freedom Over Life Choices	The extent to which individuals have autonomy and freedom in making life choices such as marriage, employment, and religion.
Individual Resilience	Count on Help	The level of trust and confidence individuals have in receiving help and support from others when needed.
Individual Resilience	Discrimination and Violence Against Minorities	The prevalence of discrimination and violence directed towards minority groups within society.
Individual Resilience	Equal Access Index	An index measuring the level of equality and fairness in access to opportunities and resources within society.
Individual Resilience	Interpersonal Violence	The incidence and prevalence of violence occurring between individuals within society.
Institutional Resilience	Constraints on Government Power	The extent to which there are checks and balances on government authority and power.
Institutional Resilience	Open Government	The degree of transparency, accountability, and openness of government institutions and processes.
Institutional Resilience	Fundamental Rights	The extent to which fundamental rights and freedoms are protected and respected within society.
Institutional Resilience	Order and Security	The level of public order and security within society, including crime rates and law enforcement effectiveness.
Institutional Resilience	Regulatory Enforcement	The effectiveness of regulatory agencies in enforcing laws and regulations within society.
Institutional Resilience	State Legitimacy	The perceived legitimacy and authority of the state government among the population.

D Formulas Statistical tests

A Paired t-test

The t-statistics and p-values for the paired t-test are calculated using formula 2.

$$t = \frac{\sum d}{\sqrt{\frac{\sum d^2 - \left(\frac{\sum d}{n}\right)^2}{n-1}}} \quad (2)$$

where:

- t is the calculated t-statistic,
- $\sum d$ is the sum of the differences between paired observations,
- $\sum d^2$ is the sum of the squared differences between paired observations, and
- n is the number of paired observations.

B Pearson correlation test

The pearson correlation is calculated by the following formula:

$$r = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}} \quad (3)$$

where:

- n = the number of data points, i.e., (x, y) pairs, in the data set.
- $\sum xy$ = the sum of the product of the x-value and y-value for each point in the data set.
- $\sum x$ = the sum of the x-values in the data set.
- $\sum y$ = the sum of the y-values in the data set.
- $\sum x^2$ = the sum of the squares of the x-values in the data set.
- $\sum y^2$ = the sum of the squares of the y-values in the data set.

C Linear Regression test

The coefficients for the linear regression model are calculated using formula 4.

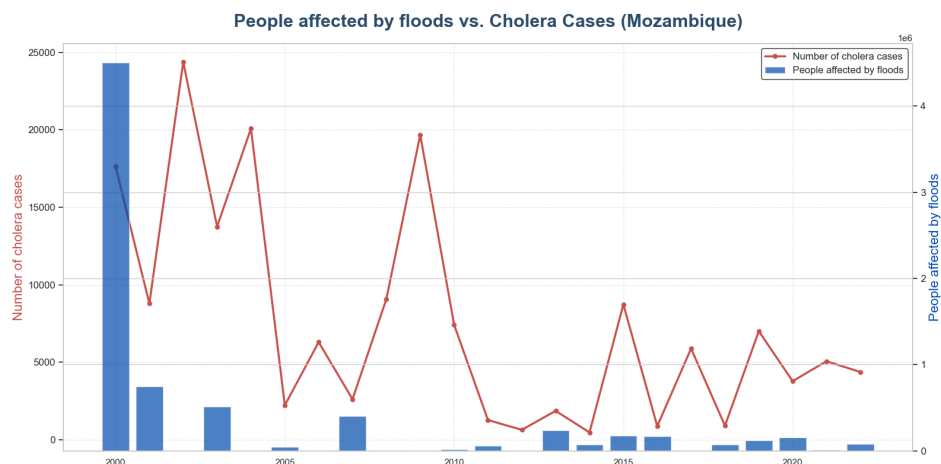
$$\mathbf{b} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (4)$$

where:

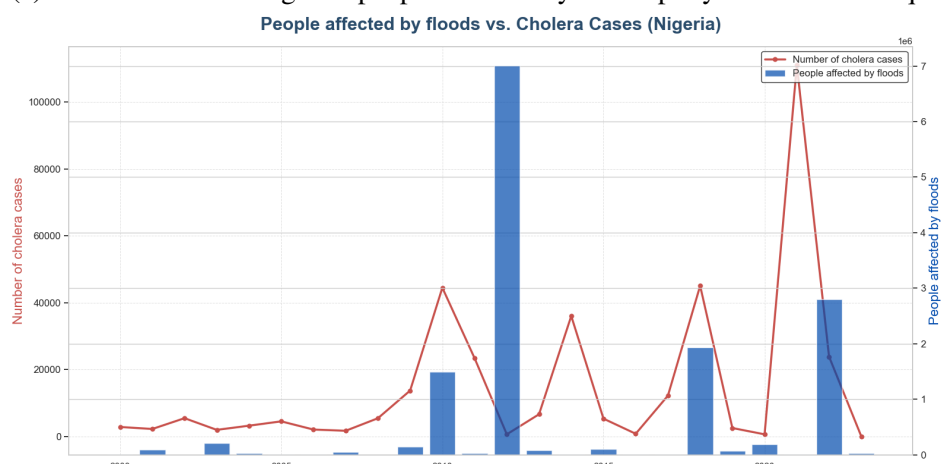
- \mathbf{b} is the vector of estimated coefficients,
- \mathbf{X} is the matrix of input features,
- \mathbf{X}^T is the transpose of the matrix of input features,
- $(\mathbf{X}^T \mathbf{X})^{-1}$ is the inverse of the product of \mathbf{X}^T and \mathbf{X} , and
- \mathbf{y} is the vector of observed outputs.

E Visualisation results

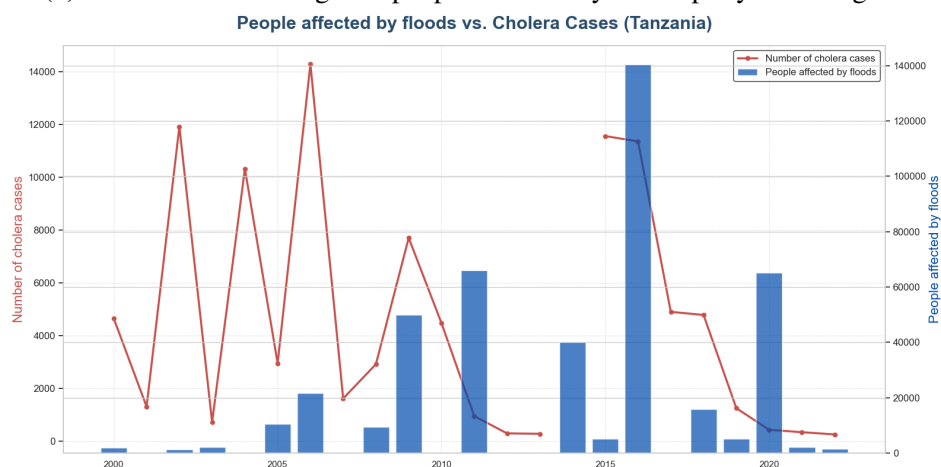
A Yearly data - Cholera



(a) Cholera incidence against people affected by floods per year in Mozambique



(b) Cholera incidence against people affected by floods per year in Nigeria



(c) Cholera incidence against people affected by floods per year in Tanzania

Figure E.5: Caption for all subfigures

B Yearly data - Malaria

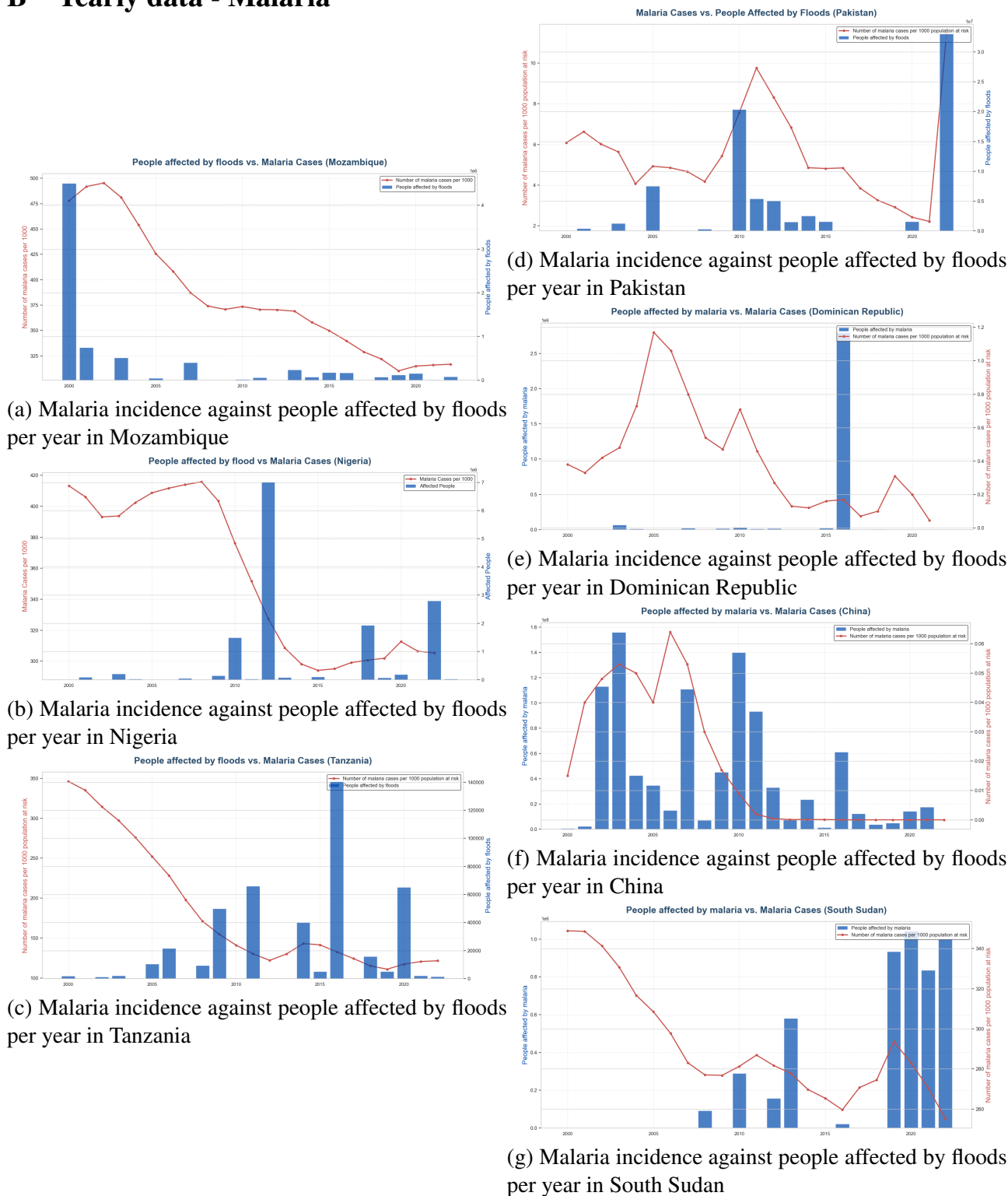
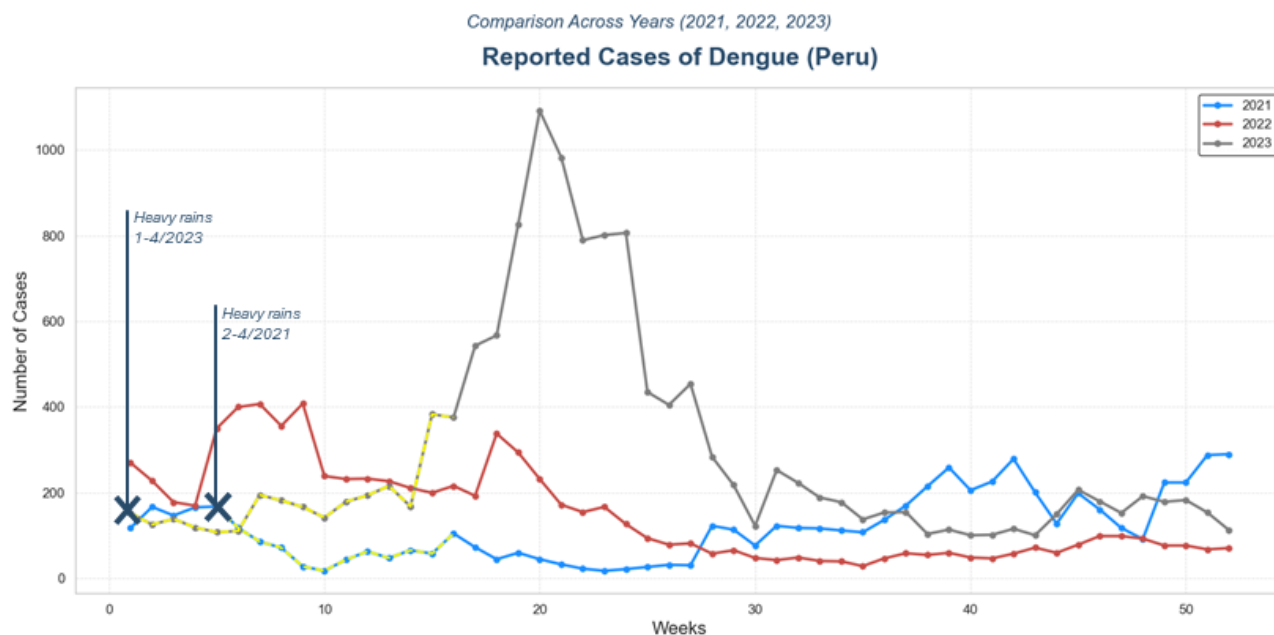
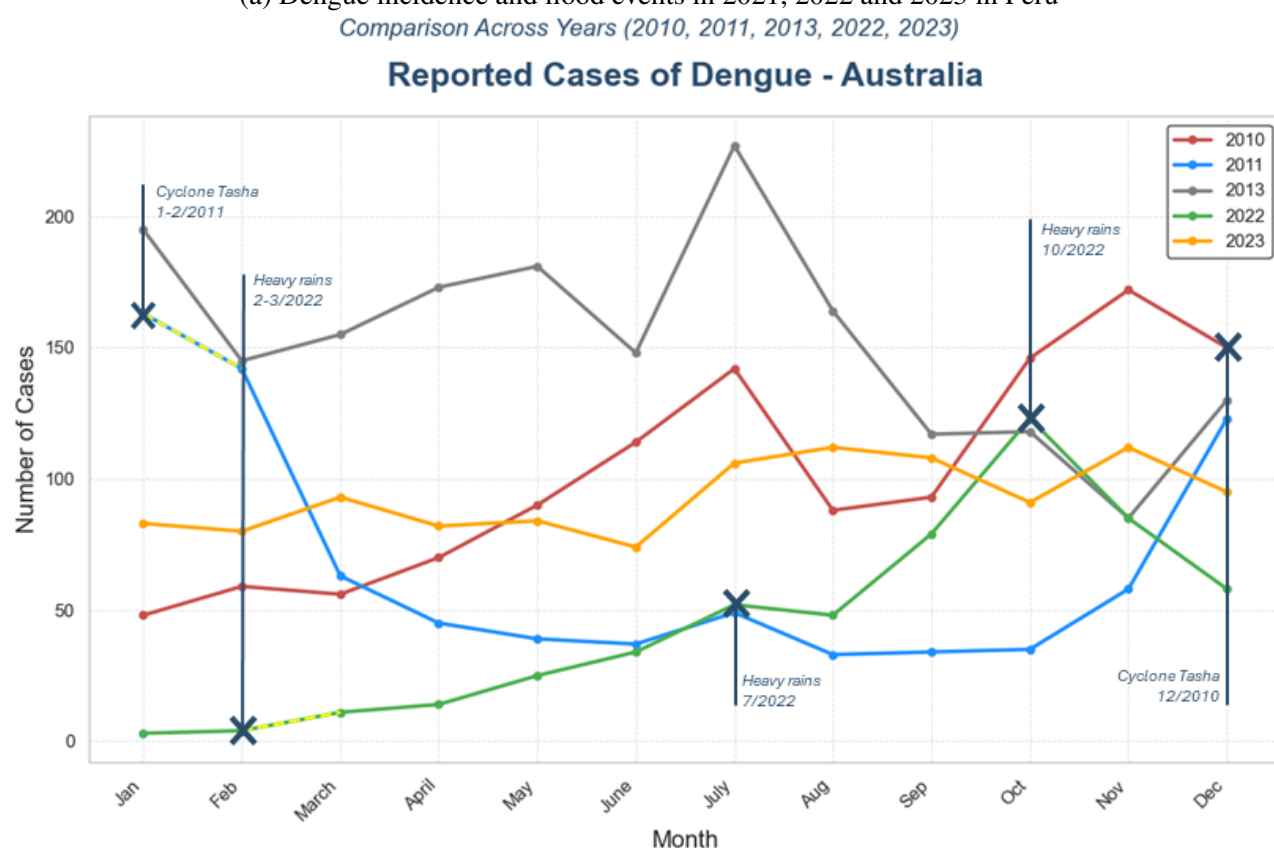


Figure E.6: Reported number of malaria cases /1000 population at risk versus number of affected people by floods, 2000 - 2022 (various countries)

C Monthly and weekly data - Dengue

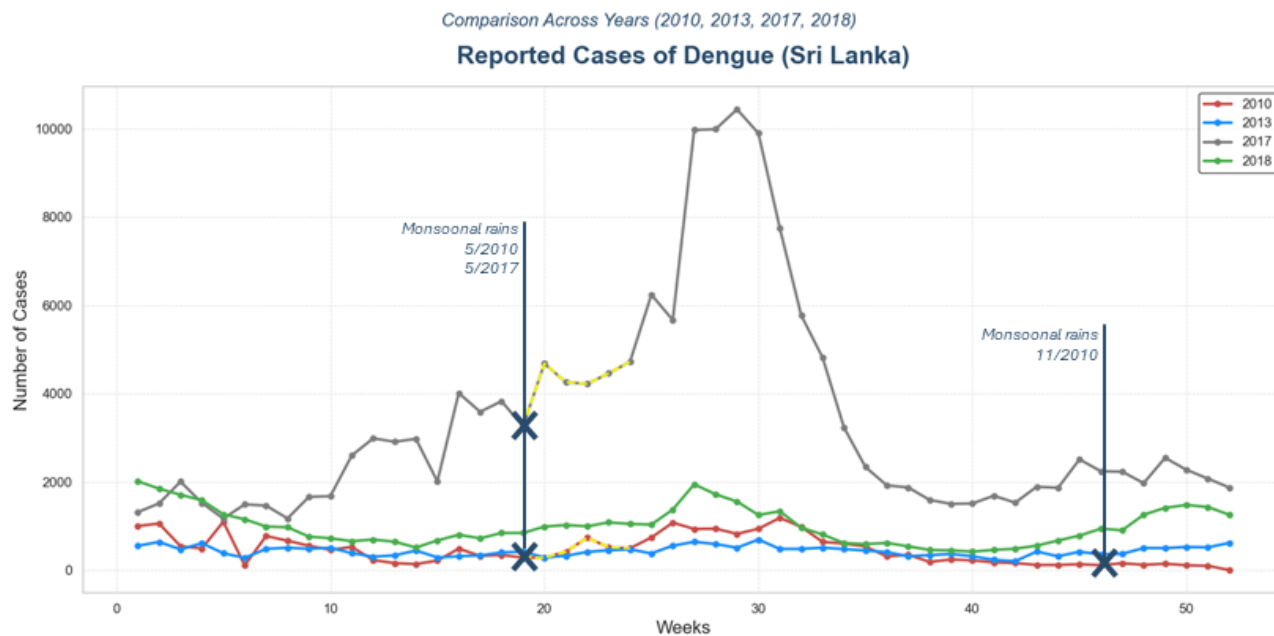


(a) Dengue incidence and flood events in 2021, 2022 and 2023 in Peru

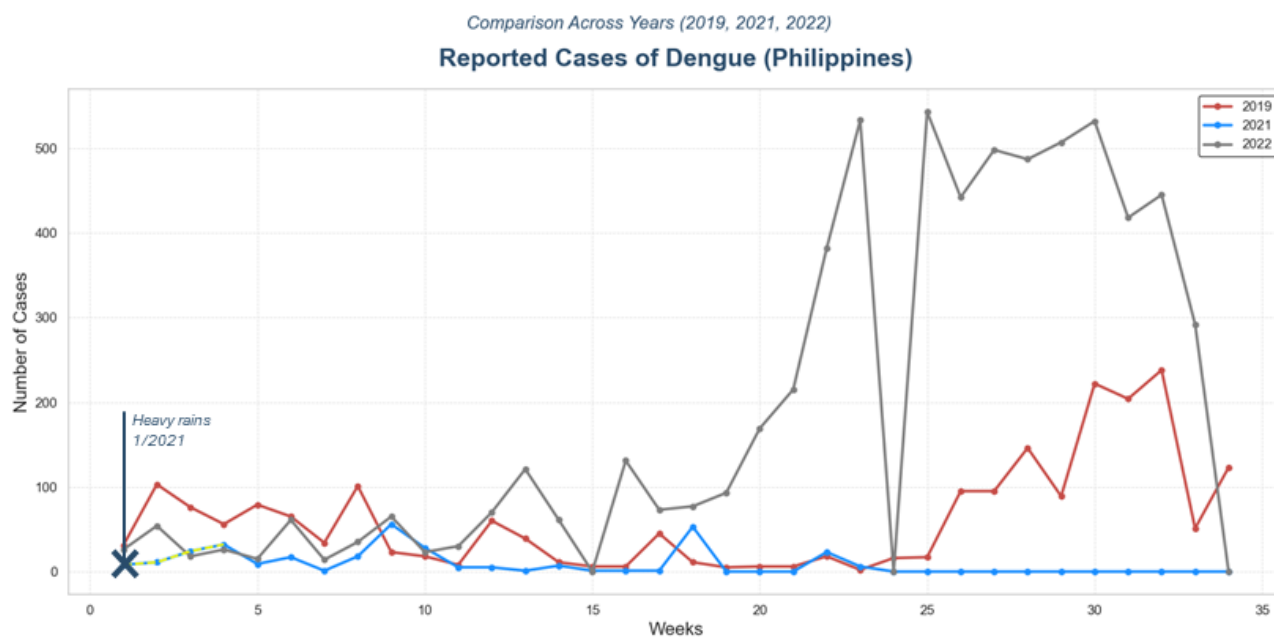


(b) Dengue incidence and flood events in 2010, 2011, 2013, 2022, 2023 in Australia

Figure E.7: Monthly and weekly data - Dengue (Peru and Australia)



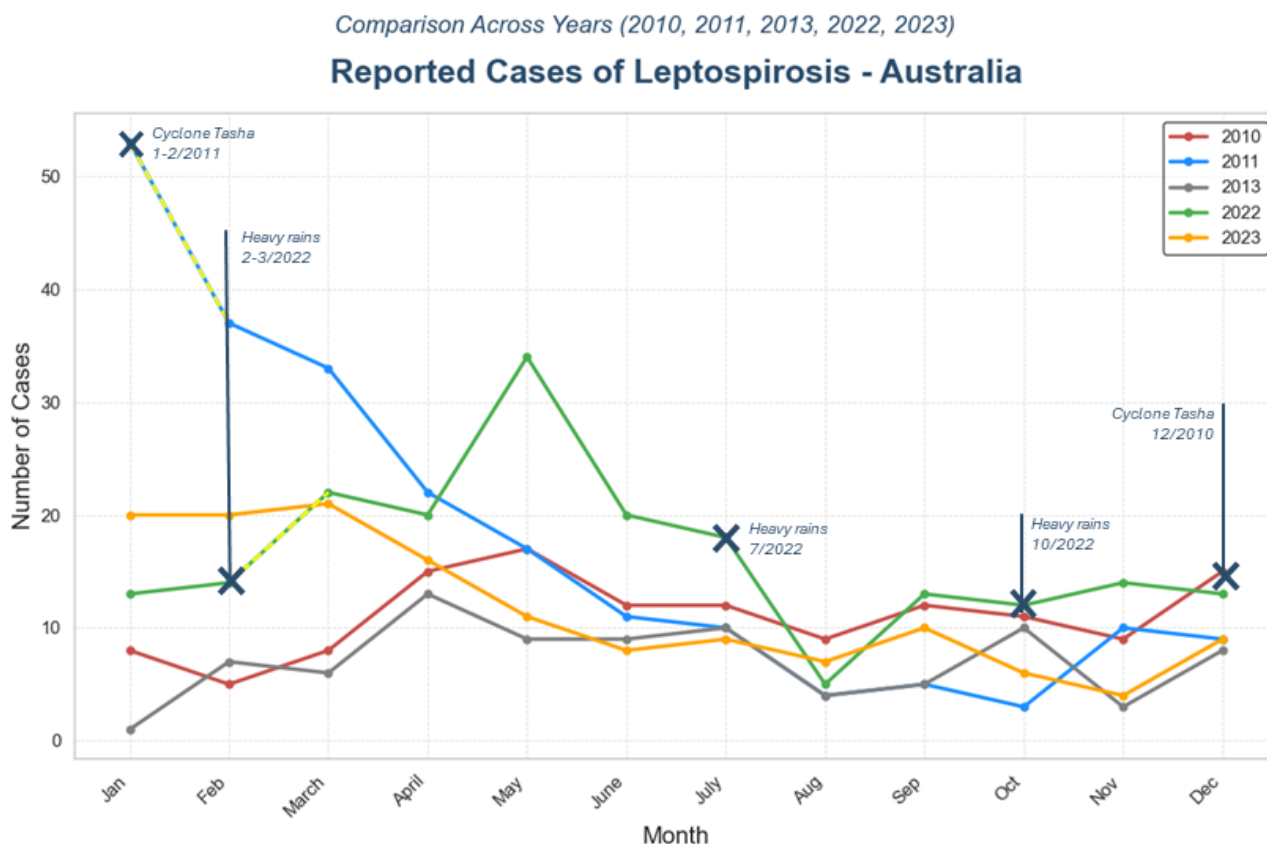
(c) Dengue incidence and flood events in 2010, 2013, 2017 and 2018 in Sri Lanka



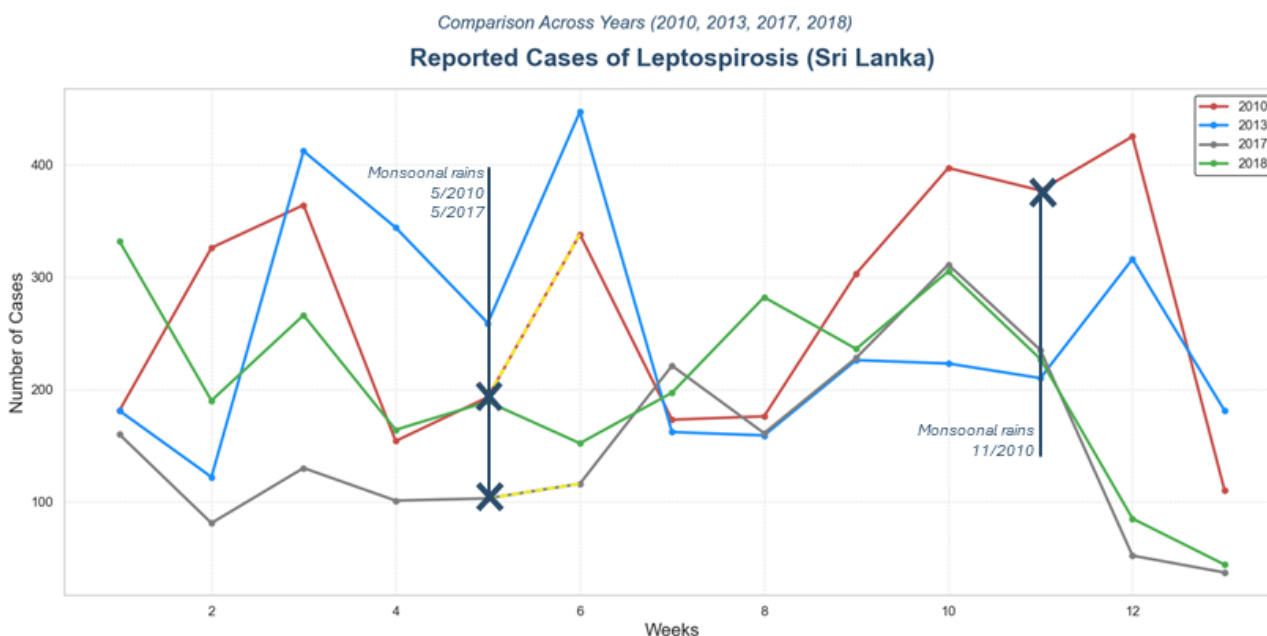
(d) Dengue incidence and flood events in 2019, 2021 and 2022 in the Philippines

Figure E.7: Monthly and weekly data - Dengue (Sri Lanka and Philippines)

D Monthly and weekly data - Leptospirosis and ADD

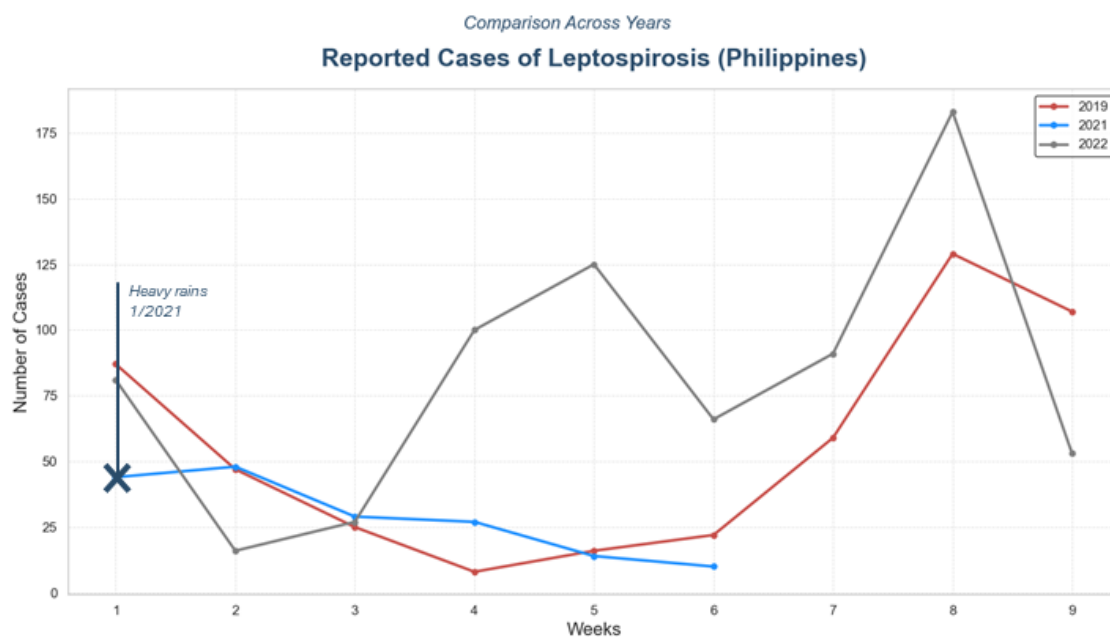


(a) Leptospirosis incidence and flood events in 2010, 2011, 2013, 2022, 2023 in Australia

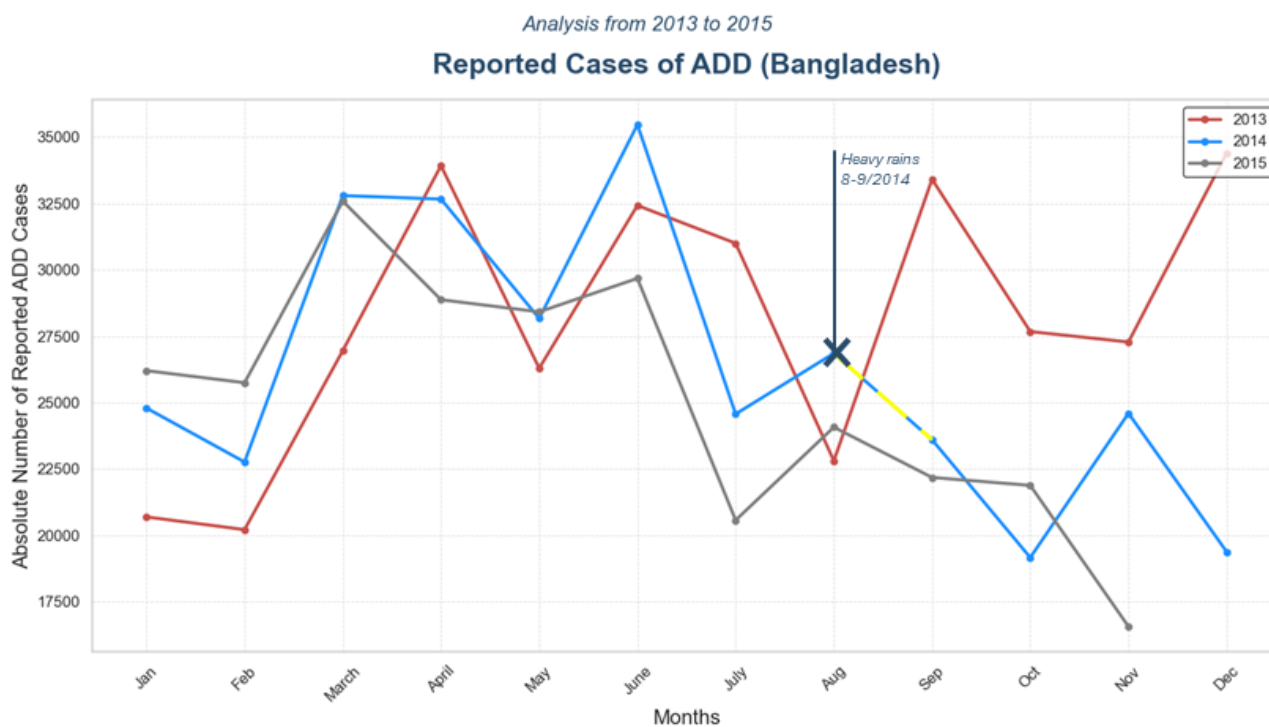


(b) Leptospirosis incidence and flood events in 2010, 2013, 2017 and 2018 in Sri Lanka

Figure E.8: Monthly and weekly data - Leptospirosis (Australia and Sri Lanka)



(c) Leptospirosis incidence and flood events in 2019, 2021 and 2022 in Philippines



(d) ADD incidence and flood events in 2019, 2021 and 2022 in Bangladesh

Figure E.8: Monthly and weekly data - Leptospirosis and ADD (Philippines, Bangladesh)

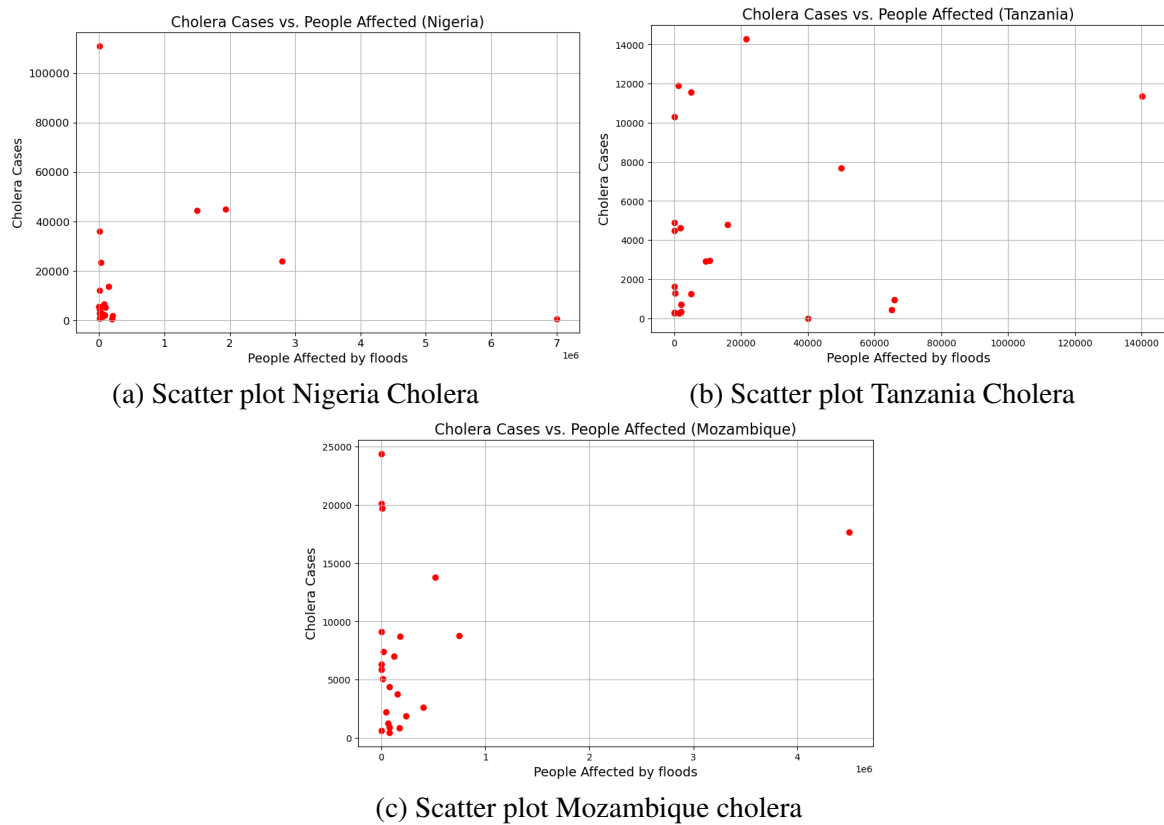


Figure F.10: Scatter plots of Waterborne disease - Cholera, for Nigeria, Tanzania and Mozambique

F Scatter plots

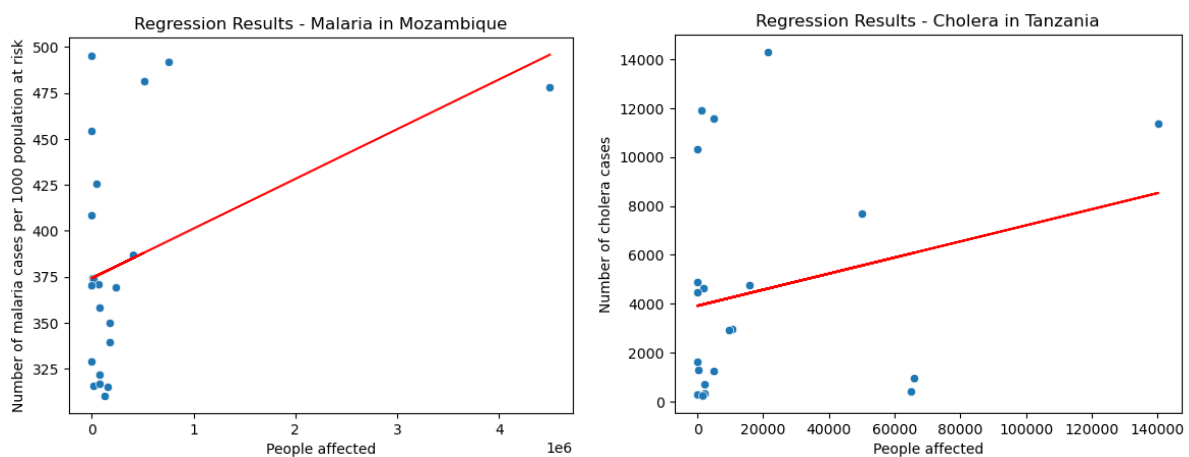


Figure F.9: Example of scatter plots with suspected linear relationship (a) and without suspected linear relationship (b)

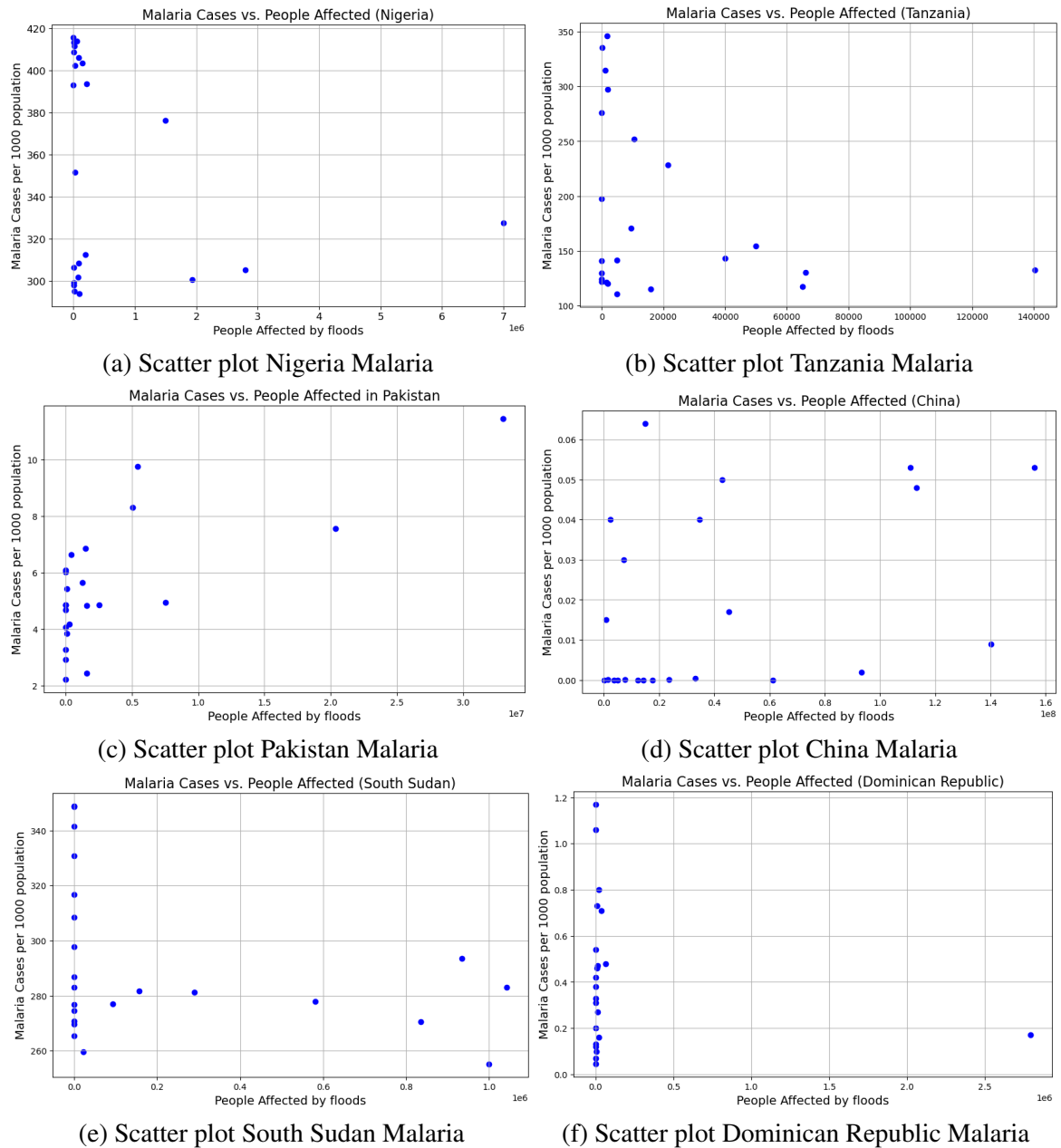


Figure F.11: Scatter plots of Vector-borne disease - Malaria, for Nigeria, Tanzania and Mozambique, Pakistan, China, South Sudan and Dominican Republic

G Descriptive statistics - histograms

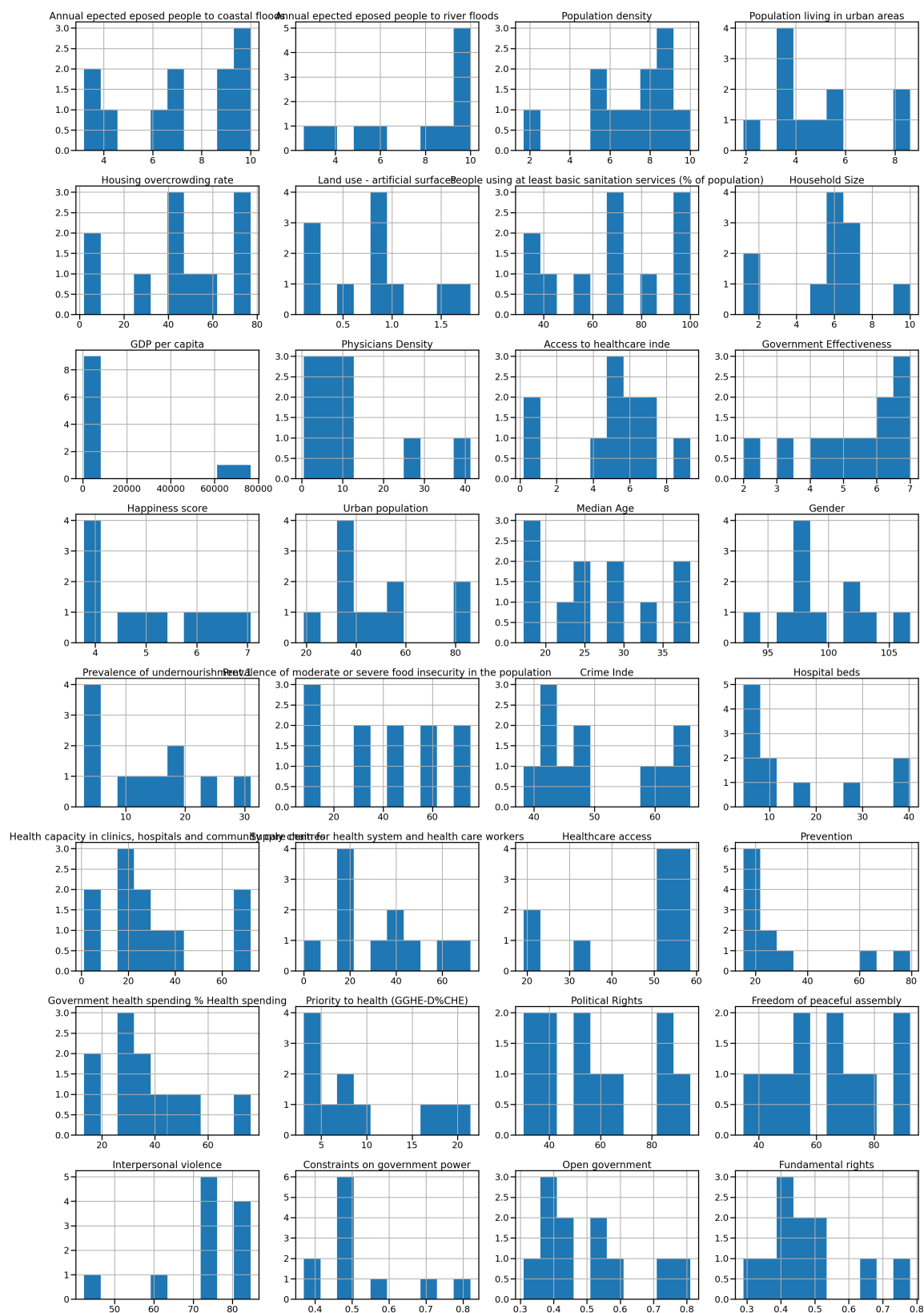


Figure G.12: Histograms of indicator data

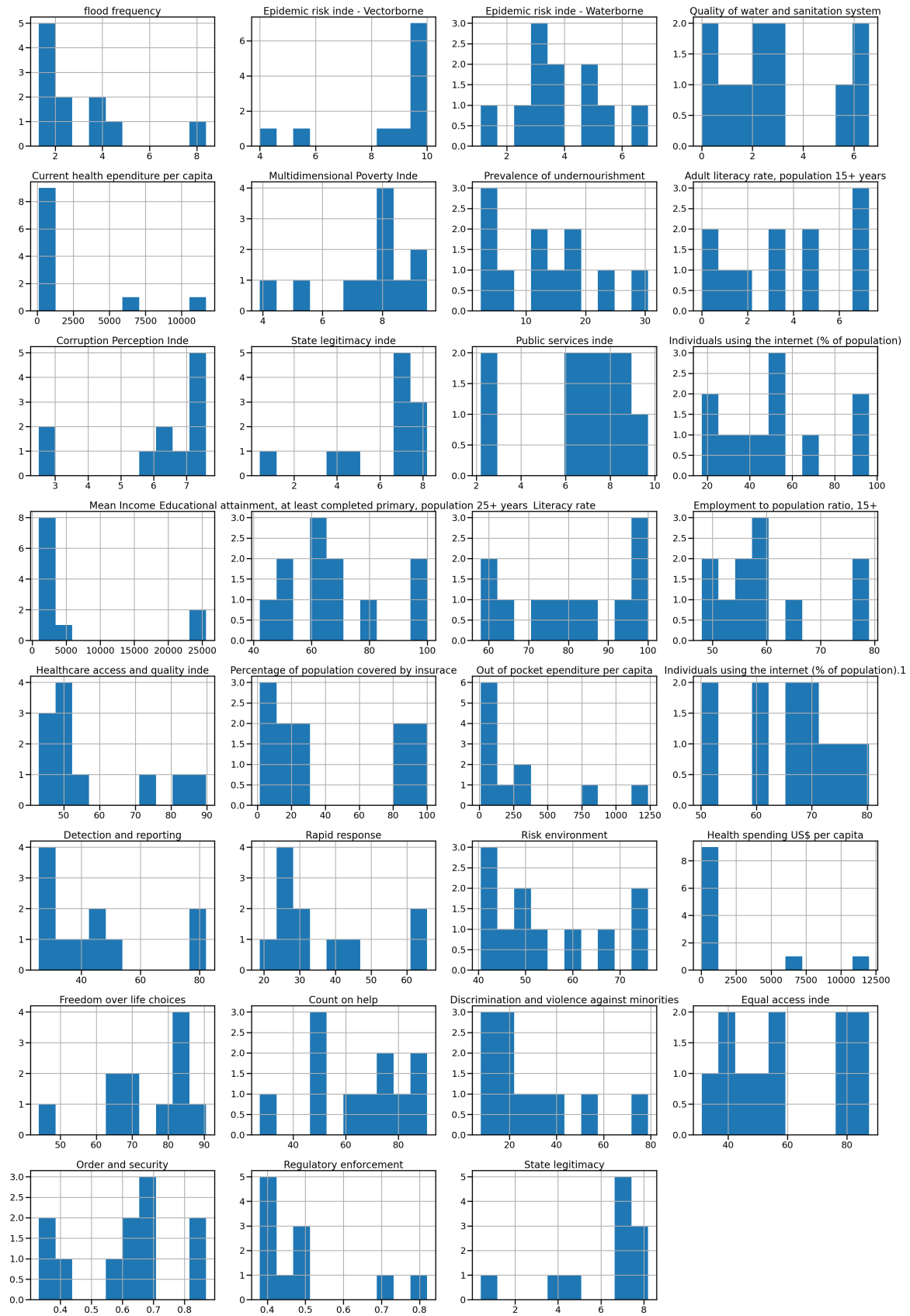


Figure H.14: Continued

Table G.6: Summary of descriptive statistics for key indicators

Indicator	mean	std	min	max
Annual expected exposed people to coastal floods	7.1	2.5	3.2	10.0
Annual expected exposed people to river floods	7.5	2.8	2.6	10.0
Population density	7.1	2.3	1.7	10.0
Population living in urban areas	4.8	2.0	1.9	8.6
flood frequency	3.0	2.1	1.3	8.4
Epidemic risk index - Vector-borne	8.7	2.0	4.0	10.0
Epidemic risk index - Waterborne	3.9	1.5	1.1	6.9
Quality of water and sanitation system	2.9	2.4	0.0	6.6
Housing overcrowding rate	44.7	25.8	2.0	77.0
Land use - artificial surfaces	0.8	0.6	0.1	1.8
People using at least basic sanitation services (% of population)	68.1	24.5	31.8	100.0
Household Size	5.7	2.5	1.2	10.0
Current health expenditure per capita	1818.7	3696.4	99.1	11702.4
Multidimensional Poverty Index	7.5	1.7	3.9	9.5
Prevalence of undernourishment	13.2	9.0	2.4	30.5
Adult literacy rate, population 15+ years	3.6	2.7	0.0	7.3
GDP per capita	14891.6	27626.0	541.5	76398.6
Physicians Density	11.8	12.0	0.5	41.3
Access to healthcare index	5.0	2.7	0.2	9.3
Government Effectiveness	5.3	1.6	2.0	7.0
Corruption Perception Inde	6.2	1.8	2.5	7.6
State legitimacy inde	6.2	2.3	0.4	8.2
Public services inde	6.7	2.4	2.2	9.7
Individuals using the internet (% of population)	51.7	25.6	17.4	96.2
Happiness score	5.1	1.2	3.8	7.1
Urban population	48.4	20.2	19.0	86.0
Median Age	27.0	7.5	17.2	38.5
Gender	99.5	3.7	93.0	106.7
Mean Income	6109.7	9368.5	947.0	25583.0
Educational attainment, at least completed primary, population 25+ years	67.9	18.6	42.1	99.9
Literacy rate	80.5	15.3	58.0	100.0
Employment to population ratio, 15+	60.1	9.8	48.0	79.0
Prevalence of undernourishment.1	13.4	9.1	3.0	31.0
Prevalence of moderate or severe food insecurity in the population	40.2	24.1	7.8	75.4
Crime Inde	50.8	10.2	38.3	65.8
Hospital beds	15.3	13.6	4.4	40.2
Healthcare access and quality index	57.8	16.1	43.0	89.8
Percentage of population covered by insurance	41.8	40.7	1.4	100.0
Out of pocket expenditure per capita	296.6	388.1	10.5	1235.3
Individuals using the internet (% of population).1	66.0	9.4	50.1	80.4
Health capacity in clinics, hospitals and community care centres	29.3	23.7	1.1	72.2
Supply chain for health system and health care workers	32.3	21.6	0.0	72.2
Healthcare access	46.7	14.8	19.2	58.5
Prevention	30.5	21.4	15.4	79.4
Detection and reporting	44.5	19.8	25.6	82.2
Rapid response	35.8	15.2	18.8	65.7
Risk environment	54.0	13.0	40.5	76.0
Health spending US\$ per capita	1853.2	3957.3	37.0	12012.0
Government health spending % Health spending	36.4	17.8	13.3	76.0
Priority to health (GGHE-D%CHE)	9.5	6.7	3.1	21.4
Political Rights	56.4	22.0	30.0	95.0
Freedom of peaceful assembly	63.4	18.6	34.8	92.4
Freedom over life choices	74.3	13.0	44.0	90.5
Count on help	64.5	20.0	27.3	90.9
Discrimination and violence against minorities	29.3	21.5	7.8	78.9
Equal access inde	59.4	20.0	31.1	87.5
Interpersonal violence	72.7	11.8	42.3	84.6
Constraints on government power	0.5	0.1	0.4	0.8
Open government	0.5	0.2	0.3	0.8
Fundamental rights	0.5	0.1	0.3	0.8
Order and security	0.6	0.2	0.3	0.9
Regulatory enforcement	0.5	0.1	0.4	0.8
State legitimacy	6.2	2.3	0.4	8.2

H Correlation matrix of exposure indicators

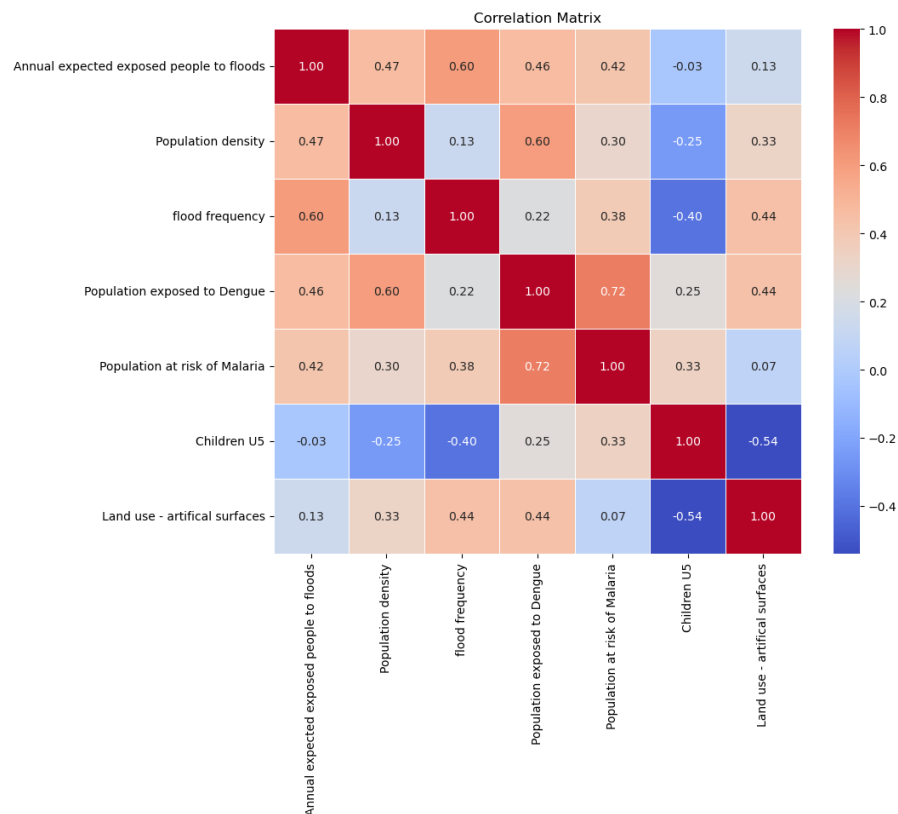


Figure H.14: Correlation matrix of indicators describing level of exposure to floods and disease

I Shapiro Wilk

Table I.7: Results of Shapiro-Wilk test for all indicators included in the PCA, indicating rejecting of H0

Indicator	Shapiro-Wilk Test Statistic	p-value	Normally distributed? (fails to reject H0)
Education attainment	0.935	0.4941	Yes
Literacy rate	0.906	0.2540	Yes
Prevalence of HIV	0.668	0.0003	No
Under 5 mortality	0.9094	0.2767	Yes
Urban population	0.8692	0.0979	Yes
Children under 5	0.8697	0.0991	Yes
Access to healthcare index	0.9017	0.2290	Yes
Hospital beds	0.7438	0.0030	No
Population covered by insurance	0.8463	0.0524	Yes
Health capacity in clinics, hospitals and community care centers	0.8984	0.2102	Yes
Supply chain for health system and health care workers	0.9241	0.3928	Yes
Individual Political Freedom	0.9127	0.2998	Yes
Individual autonomy and safety	0.9281	0.4298	Yes
Regulatory enforcement	0.8734	0.1095	Yes
State legitimacy	0.6928	0.0007	No